3rd Edition

Certified Cloud Security Professional

An (ISC)² Certification

The Official (ISC)²® CCSP® CBK® Reference

Leslie Fife
Aaron Kraus
Bryan Lewis

SYBEX®
A Wiley Brand

The Official (ISC)²® CCSP® CBK® Reference

Third Edition

CCSP®: Certified Cloud Security Professional

The Official (ISC)²® CCSP® CBK® Reference

Third Edition

LESLIE FIFE
AARON KRAUS
BRYAN LEWIS

SYBEX®
A Wiley Brand

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

ISBN: 978-1-119-60343-6
ISBN: 978-1-119-60345-0 (ebk.)
ISBN: 978-1-119-60346-7 (ebk.)

Library of Congress Control Number: 2021934228

Cover Design: Wiley and (ISC)²

Acknowledgments

First and foremost, we offer our deepest appreciation to our spouses, children, and families. Their support and understanding during the long hours of writing and review gave us the time necessary to create this book. This book would not have been possible without our wonderful families.

We would also like to express our appreciation to (ISC)^2 for providing the CCSP certification and these certification preparation materials. We are excited to be part of this transformative growth and development of secure cloud computing in the world today.

We would also like to thank John Wiley & Sons, and associate publisher Jim Minatel for entrusting us with the role of creating this study guide. We wish to thank Aaron Kraus for his review and input on the work of other sections, and our technical editor Raven Sims, whose attention to detail made this book so much better. Thanks also goes to project editor Kelly Talbot, content refinement specialist Saravanan Dakshinamurthy, copy editor Kim Wimpsett, and the entire team at Wiley for their guidance and assistance in making this book. We'd also like to thank all of our colleagues and experts who consulted with us while writing this book. You are too many to name here, but we are grateful for your suggestions and contributions.

More than anyone else, we would like to thank our readers. We are grateful for the trust you have placed in us to help you study for the exam.

—The Authors

About the Authors

Leslie D. Fife, CISSP-ISSMP, CCSP, C|CISO, CISA, CISM, CRISC, GDAT, GCED, CBCP, CIPM (and more than 20 other certifications), has more than 40 years of experience in information technology, cybersecurity, and risk management. He is currently an information security risk manager for the Church of Jesus Christ of Latter-day Saints, an assistant professor of practice at Southern Illinois University Carbondale, and an adjunct at the University of Utah. He is also a commissioner for the Computing Accreditation Commission of ABET. His career includes the U.S. Navy submarine service, software development in the defense industry and the oil and gas field service industry, incident response and business continuity in the financial services sector, as well as 22 years as a professor of computer science. He has a PhD in computer science from the University of Oklahoma.

Aaron Kraus, CCSP, CISSP, is an information security professional with more than 15 years of experience in security risk management, auditing, and teaching information security topics. He has worked in security and compliance roles across industries including U.S. federal government civilian agencies, financial services, and technology startups, and he is currently the security engagement manager at Coalition, Inc., a cyber risk insurtech company. His experience includes creating alignment between security teams and the organizations they support, by evaluating the unique threat landscape facing each organization and the unique objectives each organization is pursuing to deliver a balanced, risk-based security control program. As a consultant to a financial services firm he designed, executed, and matured the third-party vendor audit programs to provide oversight of key compliance initiatives, and he led the global audit teams to perform reviews covering physical security, logical security, and regulatory compliance. Aaron is a course author, instructor, and cybersecurity curriculum dean with more than 13 years of experience at Learning Tree International, and he most recently taught the Official (ISC)^2 CISSP CBK Review Seminar. He has served as a technical editor for numerous Wiley publications including *(ISC)^2 CCSP Certified Cloud Security Professional Official Study Guide, 2nd Edition*; *CCSP Official (ISC)^2 Practice Tests, 1st Edition*; *The Official (ISC)^2 Guide to the CISSP CBK Reference, 5th Edition*; and *(ISC)^2 CISSP Certified Information Systems Security Professional Official Practice Tests, 2nd Edition*.

Bryan Lewis, EdD, currently serves as an assistant dean and IT area lecturer for the McIntire School of Commerce at the University of Virginia. Certified as both a CISSP and CCSP, he has extensive experience with cybersecurity operations, research, and instruction in both the public and private sectors. Prior to joining the McIntire School, Dr. Lewis served as a company officer and principal for an audio visual and telecommunications design, engineering, and manufacturing company. His past experience includes large-scale network infrastructure and secure system design, deployments, and migrations, including secure distance-based learning and collaborative space design. He currently serves as a lecturer on network, data, and cloud security with a focus on defensive technologies, secure communications, and the business impacts of information security in the graduate and undergraduate curricula. His primary consulting interests focus on distance learning design, large-scale visualization, information security in the public sector, and collaborative space design projects.

About the Technical Editor

Raven Sims, CISSP, CCSP, SSCP, is a space systems senior principal cyber architect in the Strategic Deterrent division of a notable defense contractor. In this role, Sims has responsibility for the division's cyber architecture within the weapon system command-and-control business portfolio, including full-spectrum cyber, cloud computing, as well as mission-enabling cyber solutions supporting domestic and international customers. Most recently, Sims was a cyber architect of the Department of Justice (DoJ) Cybersecurity Services (CSS) team in providing cloud security guidance to all 14+ DoJ components. She was responsible for designing, deploying, and maintaining enterprise-class security, network, and systems management applications within an Amazon Web Services (AWS) and Azure environment. Within this role, she led incident response guidance for the DoJ as it pertained to securing the cloud and how to proactively respond to events within their cloud infrastructure. Sims has held business development, functional, and program positions of increasing responsibility in multiple sectors of the company. Her program experience includes government and international partnerships. Sims earned a bachelor's degree in computer science from Old Dominion University in Norfolk, Virginia, and a master's degree in technology management from Georgetown University in Washington, D.C. She is now pursuing a doctoral degree from Dakota State University in cyber operations. She serves on the board of directors of FeedTheStreetsRVA (FTSRVA); is a member of Society of Women Engineers (SWE) and Zeta Phi Beta Sorority, Inc.; and is the owner of Sims Designs. Sims is nationally recognized for her advancements in cyber and mission solutions as an awardee of the 2019 Black Engineer of the Year (BEYA): Modern Day Technology Award, and UK Cybercenturion awards.

Contents at a Glance

Contents

Foreword to the Third Edition

EARNING THE GLOBALLY RECOGNIZED CCSP® cloud security certification is a proven way to build your career and better secure critical assets in the cloud. Whether you are picking up this book to supplement your preparation to sit for the exam or you are an existing CCSP using this as a desk reference, you'll find the *Official (ISC)² Guide to the CCSP CBK* to be the perfect primer on the cloud security topics covered in the CCSP CBK.

Cloud computing security is one of the most in-demand skillsets in IT today. The designation of CCSP instantly communicates to everyone within our industry that you have the advanced technical skills and knowledge to design, manage, and secure data, applications, and infrastructure in the cloud using best practices, policies, and procedures established by the cybersecurity experts at (ISC)².

The recognized leader in the field of information security education and certification, (ISC)² promotes the development of information security professionals throughout the world. As a CCSP with all the benefits of (ISC)² membership, you are part of a global network of more than 157,000 certified professionals who are working to inspire a safe and secure cyber world.

Drawing from a comprehensive, up-to-date global body of knowledge, the *CCSP CBK* provides you with valuable insights on how to implement cloud security across different digital platforms that your organization may be using.

If you are an experienced CCSP, you will find this edition of the *CCSP CBK* to be an indispensable reference on best practices. If you are still gaining the experience and knowledge you need to join the ranks of CCSPs, the *CCSP CBK* is a deep dive that can be used to supplement your studies.

As the largest nonprofit membership body of certified information security professionals worldwide, (ISC)² recognizes the need to identify and validate not only information security competency, but also the ability to connect knowledge of several cloud security domains when managing or migrating data to and from the cloud. The CCSP represents advanced knowledge and competency in cloud security architecture, design, operations, and service orchestration.

The opportunity has never been greater for dedicated professionals to carve out a meaningful career and make a difference in their organizations. The *CCSP CBK* will be your

constant companion in protecting and securing the critical data assets of your organization that will serve you for years to come.

Sincerely,

Clar Rosso
CEO, (ISC)^2

Introduction

THE CERTIFIED CLOUD SECURITY Professional (CCSP) denotes a professional with demonstrated ability across important aspects of architecture, data security, and risk management in cloud computing. The exam covers knowledge and skills across six domains of practice related to cloud security, codified in the (ISC)^2 CCSP Common Body of Knowledge (CBK):

- Domain 1: Cloud Concepts, Architecture, and Design
- Domain 2: Cloud Data Security
- Domain 3: Cloud Platform and Infrastructure Security
- Domain 4: Cloud Application Security
- Domain 5: Cloud Security Operations
- Domain 6: Legal, Risk, and Compliance

Passing the exam is one condition of certification, and to qualify for the certification, a professional must have five years of experience in information technology, of which three years must be in a security-specific capacity and at least one year dedicated to one or more of the six CCSP domains.

Professionals take many paths into information security, and there are variations in acceptable practices across different industries and regions. The CCSP CBK represents a baseline standard of security knowledge relevant to cloud security and management, though the rapid pace of change in cloud computing means a professional must continuously maintain their knowledge to stay current. As you read this guide, consider not only the scenarios or circumstances presented to highlight the CBK topics, but also connect it to common practices and norms in your organization, region, and culture. Once you achieve CCSP certification, you will be asked to maintain your knowledge with continuing education, so keep topics of interest in mind for further study once you have passed the exam.

Domain 1: Cloud Concepts, Architecture, and Design

Understanding cloud computing begins with the building blocks of cloud services, and the Cloud Concepts, Architecture, and Design domain introduces these foundational concepts. This includes two vital participants: cloud service providers and cloud consumers, as well as reference architectures used to deliver cloud services like infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). These relatively new methods of accessing IT resources offer interesting business benefits like shifting spending from capital expenditure (CapEx) to operating expenditure (OpEx). This changes the way organizations budget and pay for the IT resources needed to run their business, so it is not uncommon to see financial leaders driving adoption of cloud services. New IT service models bring with them new forms of information security risks, however, which must be assessed and weighed so the organization achieves an optimal balance of cost (in the form of risk) with benefits (in the form of reduced IT spending). This will drive decisions on which cloud deployment model to adopt, like public or private cloud, as well as key internal governance initiatives when migrating to and managing cloud computing.

Domain 2: Cloud Data Security

Information security is fundamentally concerned with preserving the confidentiality, integrity, and availability of data. Although cloud computing upends many legacy IT models and practices, security risks to information systems remain. The Cloud Data Security domain does introduce new concepts like the cloud data lifecycle, as well as cloud-specific considerations like data dispersion and loss of physical control over storage media that requires unique approaches to data disposal. Cloud security practitioners must understand how to implement controls for audit and accountability of data stored or processed in the cloud, as well as crucial oversight tasks like data discovery to create an inventory. This domain introduces proactive safeguards intended to manage sensitive data stored in the cloud, like masking, tokenization, data loss prevention (DLP), and classification of data. Cloud-specific considerations and adaptations of traditional controls are a primary concern, since cloud services remove traditional capabilities like physical destruction of disk drives, while adding new capabilities like instantaneous global data replication.

Domain 3: Cloud Platform and Infrastructure Security

There are two perspectives treated in the Cloud Platform and Infrastructure Security domain. Cloud providers require skilled security practitioners to design, deploy, and maintain both physically and logically secure environments. This includes buildings, facilities, and utilities needed to provide the cloud service offering, as well as

configuration and management of software systems like hypervisors, storage area networks (SANs), and software-defined networking (SDN) infrastructure. A key concern is the security of data stored by the cloud consumers, particularly properly isolating tenant data to avoid leakage between cloud tenants. From the perspective of the cloud consumer, traditional security controls will require adaptation for cloud environments, such as the use of virtualized hardware security modules (HSM) to generate and manage cryptographic keys, and additional layers of encryption required to reduce the risk associated with giving up physical control of storage media. Audit mechanisms like log collection are traditionally present in cloud environments, but abilities like packet capture and analysis may not be available due to multitenant data concerns. Disaster recovery and business continuity are also presented in this domain; while the inherent high availability nature of many cloud services is beneficial for organizations, proper configuration to take advantage of these features is required.

Domain 4: Cloud Application Security

Security practitioners working in cloud computing environments face the challenge of more rapid deployment, coupled with the relative ease with which more users can develop sophisticated cloud applications. Again, these are advantages to the business at the possible expense of security, so the Cloud Application Security domain presents key requirements for recognizing the benefits offered by cloud applications without introducing unacceptable risks. These begin with a focus on the importance of fostering awareness throughout the organization of common cloud security basics, as well as specific training for cloud app developers on vulnerabilities, pitfalls, and strategies to avoid them. Modifications to the software development lifecycle (SDLC) are presented to help accommodate changes introduced by cloud-specific risks, such as architectures designed to avoid vendor lock-in and threat modeling specific to the broadly accessible nature of cloud platforms. Since many cloud computing services are delivered by third parties, this domain introduces assurance, validation, and testing methods tailored to address the lack of direct control over acquired IT services and applications. It also introduces common application security controls and specifics of their implementation for cloud environments, like web application firewalls (WAF), sandboxing, and Extensible Markup Language (XML) gateways. Many cloud services rely heavily on functionality offered via application programming interfaces (APIs), so it is crucial that security practitioners understand how data is exchanged, processed, and protected by APIs.

Domain 5: Cloud Security Operations

The Cloud Security Operations domain is a companion to many of the concepts introduced in the Cloud Platform and Infrastructure Security domain. It deals with issues of

implementing, building, operating, and managing the physical and logical infrastructure needed for a cloud environment. There is a heavy focus on the cloud service provider's perspective, so concepts in this domain may be unfamiliar to some security practitioners who have only worked to secure cloud services as a consumer. The concepts are largely similar to legacy or on-premises security, such as the secure configuration of BIOS and use of Trusted Platform Module (TPM) for hardware security, deployment of virtualization management tools, and configuring remote maintenance capabilities to allow remote administrative tasks. Considerations unique to cloud environments include the additional rigor required in the configuration of isolation features, which prevent data access across tenants, as well as the much larger demands of managing capacity, availability, and monitoring of vast, multicountry data centers. Traditional security operations (SecOps) are also of critical concern for security practitioners in a cloud environment, such as the management of vulnerability and patch management programs, network access and security controls, as well as configuration and change management programs. Additional SecOps activities covered in this domain include supporting incident response and digital forensics when security incidents occur, as well as traditional security operations center (SOC) oversight and monitoring functions for network security, log capture and analysis, and service incident management. These tasks are also covered from the cloud consumer's perspective, as many cloud services and security tools provide log data that must be analyzed to support policy enforcement and incident detection.

Domain 6: Legal, Risk, and Compliance

Legal and regulatory requirements are a significant driver of the work many information security professionals perform, and cloud computing makes this increasingly more complex due to its inherently global nature. The Legal, Risk, and Compliance domain details the conflicting international laws and regulations that organizations will encounter when using cloud services. These present financial risks, additional compliance obligations and risk, as well as technical challenges like verifying that cloud applications and services are configured in accordance with compliance requirements. One particularly important area of focus is privacy legislation; with many countries and localities introducing strict requirements to safeguard privacy data, organizations using the cloud must weigh any financial benefits of a cloud migration against potential fines if they violate these laws. New challenges are also emerging around jurisdiction over multinational cloud services: how do you determine jurisdiction for a U.S. based company operating a cloud data center in Singapore processing data belonging to a Swiss citizen? Three different laws potentially overlap in this scenario. Processes for audits, assurance, and reporting are also covered, as security practitioners must understand and be able to implement both internal oversight mechanisms like gap analysis and audit planning, while also selecting and supporting external auditors for standards like Service Organization Control

(SOC) audit reports. Some organizations may even find themselves in such heavily regulated industries, like healthcare or national defense, that the potential risks of cloud computing outweigh any cost savings. These types of decisions must be driven by solid risk management principles, which require adequate assessment and mitigation techniques. Since cloud service providers are third parties not directly under the control of the organization, vendor risk management practices like contract design and service level agreements (SLAs) must be utilized to execute the chosen risk management strategy.

HOW TO CONTACT THE PUBLISHER

If you believe you've found a mistake in this book, please bring it to our attention. At John Wiley & Sons, we understand how important it is to provide our customers with accurate content, but even with our best efforts, an error may occur.

To submit your possible errata, please email it to our Customer Service Team at `wileysupport@wiley.com` with the subject line "Possible Book Errata Submission."

DOMAIN 1

CCSP®

Cloud Concepts, Architecture, and Design

FOUNDATIONAL TO THE UNDERSTANDING and use of the cloud and cloud computing is the information found in Domain 1. This information is fundamental for all other topics in cloud computing. A set of common definitions, architectural standards, and design patterns will put everyone on the same level when discussing these ideas and using the cloud effectively and efficiently.

UNDERSTAND CLOUD COMPUTING CONCEPTS

The first task is to define common concepts. In the following sections, we will provide common definitions for cloud computing terms and will discuss the various participants in the cloud computing ecosystem. We will also discuss the characteristics of cloud computing, answering the question "What is cloud computing?" We will also examine the technologies that make cloud computing possible.

Cloud Computing Definitions

The basic concepts of cloud computing, service models, and deployment models form the foundation of cloud computing practice. It is essential to understand each of them.

Cloud Computing

In NIST SP 800-145, cloud computing is defined as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources. . . . that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Cloud computing is more than distributed computing or parallel computing even when done over a network (local area network or Internet). It is a philosophy that creates access to computing resources in a simple, self-driven way. If an individual has to call up the vendor and negotiate a contract for a fixed service, it is probably not cloud computing. Similarly, a company may negotiate rates and services in a cloud environment. But, the provisioning of services must not require ongoing involvement by the vendor.

Cloud computing requires a network in order to provide broad access to infrastructure, development tools, and software solutions. It requires some form of self-service to allow users to reserve and access these resources at times and in ways that are convenient to the user.

The provisioning of resources needs to be automated so that human involvement is limited. Any user should be able to access their account and procure additional resources or reduce current resource levels by themselves.

An example is Dropbox, a cloud-based file storage system. An individual creates an account, chooses the level of service they want or need, and provides payment information, and then the service and storage are immediately available. A company might negotiate contract rates more favorable than are available to the average consumer. But, once the contract is in place, the employees access this resource in much the same way as an individual user of this service.

Service Models

There are three service models: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). These models determine the type of user the cloud service is designed for: end users, developers, or system administrators.

The different service models also dictate the level of control over software applications, operating systems, networking, and other components. The least control for the end user exists in the SaaS model, with only basic configuration controls available, if any. The most control for the end user is the IaaS model where operating system selection and configuration, patching, and software tools and applications are under the control of the end user.

When the service is provided to a company, the distinction can be less clear. While a PaaS may be intended for use by developers, there may be some administration of the

service by the company as well. In fact, the lines often blur when a corporation enters into a business relationship with a cloud provider but does much of the provisioning and administrative work in house.

For example, Office 365 can be considered a SaaS solution, and to the individual consumer there is little or no administrative overhead. But, if a company contracts for Office 365, they may in fact administer the system, overseeing account provisioning, system monitoring, and other tasks that would be the domain of developers and administrators.

Deployment Models

There are four deployment models: public, private, community, and hybrid clouds. These define who owns and controls the underlying infrastructure of a cloud service and who can access a specific cloud service.

A public cloud deployment makes resources available for anyone who chooses to create an account and purchase access to the service. A service like Dropbox is available to the public SaaS deployment. Accounts on various cloud service providers such as Amazon Web Services (AWS), Google, and IBM Cloud are also public deployments of services.

A private cloud deployment consists of a set of cloud resources for a single organization (business, non-profit, etc.). The cloud may be located on-premise in the organization's data center or may be in a single tenant cloud environment provided by a CSP. The services (SaaS, PaaS, or IaaS) are available solely to that organization. You get many of the advantages of a cloud such as the on-demand resources and minimal management effort. However, the company still owns the infrastructure. This can provide the benefits of cloud computing for files and data that are too sensitive to put on a public cloud.

A community cloud is most similar to a public cloud. It is a cloud deployment for a related group of companies or individuals such as a consortium of universities or a group of local or state governments. The cloud may be implemented in one of the organizations, with services provided to all members. Or, it can be implemented in an infrastructure like AWS or Google. However, access to the cloud resources is available only to the members of the group.

A hybrid cloud is any combination of these. A company may have a private cloud that accesses public cloud resources for some of its functions. The hybrid cloud allows the organization of cloud resources in whatever way makes the most sense to the organization. Private individuals are not usually involved in a hybrid cloud. This is because few individuals have their own private cloud or belong to a community cloud as individuals.

These concepts will be discussed further in the "Cloud Deployment Models" section later in this chapter.

Cloud Computing Roles

There are a number of roles in cloud computing, and understanding each role allows clearer understanding of each of the cloud service models, deployment models, security responsibilities, and other aspects of cloud computing.

Cloud Service Customer

The cloud service customer (CSC) is the company or person purchasing the cloud service, or in the case of an internal customer, the employee using the cloud service. For example, a SaaS CSC would be any individual or organization that subscribes to a cloud-based email service. A PaaS CSC would be an individual or organization subscribing to a PaaS resource. A PaaS resource could be a development platform. With an IaaS solution, the customer is a system administrator who needs infrastructure to support their enterprise. In a very real sense, the customer is the individual the particular service model was created to support.

Cloud Service Provider

The cloud service provider (CSP) is the company or other entity offering cloud services. A CSP may offer SaaS, PaaS, or IaaS services in any combination. For example, major CSPs such as AWS, Microsoft Azure, and Google Cloud offer both PaaS and IaaS services.

Depending on the service provided (SaaS, PaaS, or IaaS), the responsibilities of the CSP vary considerably. In all cases, security in the cloud is a shared responsibility between the CSP and the customer. This shared responsibility is a continuum, with the customer taking a larger security role in an IaaS service model and the CSP taking a larger role in the security in a SaaS service model. The responsibilities of a PaaS fall somewhere in between. But even when a CSP has most of the responsibility in a SaaS solution, the customer is ultimately responsible for the data and processes they put into the cloud.

The basic infrastructure is the responsibility of the CSP, including the overall security of the cloud environment and the infrastructure components provided. This would include responsibilities such as physical security of data centers. For example, AWS is always responsible for securing the AWS Cloud environment. The customer is responsible for the security of what they do in the cloud. The customer has ultimate responsibility for the security of their customer and other sensitive data and how they use the cloud and cloud components. The CSP may provide many security services, but the customer may choose not to use some or all of those services.

As the cloud environment becomes more complicated, with hybrid clouds and community clouds that federate across multiple cloud environments, the responsibility

for security becomes ever more complex. As the customer owns their data and processes, they have a responsibility to review the security policies and procedures of the CSP, and the federated responsibilities that may exist between multiple CSPs and data centers.

Cloud Service Partner

A cloud service partner is a third party offering a variety of cloud-based services (infrastructure, storage and application services, and platform services) using the associated CSP. An AWS cloud service partner uses AWS to provide their services. The cloud service partner can provide customized interfaces, load balancing, and a variety of services. It may be an easier entrance to cloud computing, as an existing customer vendor may already be a cloud service partner. The partner has experience with the underlying CSP and can introduce a customer to the cloud more easily.

The cloud partner network is also a way to extend the reach of a CSP. The cloud service partner will brand its association with the CSP. Some partners align with multiple CSPs, giving the customer a great deal of flexibility.

Some partners provide their own or most of their own infrastructure and extend the service areas they can reach through the use of partnerships. For example, Dropbox extends its reach to service areas where it does not have infrastructure through a continued partnership with AWS. This also allows Dropbox to expand beyond what its own infrastructure will currently handle.

Cloud Service Broker

A cloud service broker is similar to a broker in any industry. Companies use a broker to find solutions to their cloud computing needs. The broker will package services in a manner that benefits the customer. This may involve the services of multiple CSPs. A broker is a value-add service and can be an easy way for a company to begin a move into the cloud. A broker adds value through aggregation of services from multiple parties, integration of services with a company's existing infrastructure, and customization of services that a CSP cannot or will not make.

In both cases, or with one of the dozens of other CSBs, it is important to thoroughly vet the CSB as you would any new vendor. Each serves a specific market, utilizing different cloud technologies. It is important that the CSBs selected are a good fit for the customer organization and its cloud strategy.

Key Cloud Computing Characteristics

The NIST definition of cloud computing describes certain characteristics that clouds share. Not every third-party solution is a cloud solution. Understanding the key

characteristics of cloud computing will allow you to distinguish between cloud solutions and noncloud solutions. This is important as these characteristics result in certain security challenges that may not be shared by noncloud solutions.

On-Demand Self-Service

The NIST definition of cloud computing identifies an on-demand service as one "that can be rapidly provisioned and released with minimal management effort or service provider interaction." This means the user must be able to provision these services simply and easily when they are needed. If you need a Dropbox account, you simply set up an account and pay for the amount of storage you want, and you have that storage capacity nearly immediately. If you already have an account, you can expand the space you need by simply paying for more space. The access to storage space is on demand. Neither creating an account nor expanding the amount of storage available requires the involvement of people other than the customer. This access is automated and provided via a dashboard or other simple interface.

This can facilitate the poor practice often labeled as *shadow IT*. The ease with which a service can be provisioned makes it easy for an individual, team, or department to bypass company policies and procedures that handle the provisioning and control of IT services. A team that wants to collaborate may choose OneDrive, Dropbox, SharePoint, or another service to facilitate collaboration. This can lead to sensitive data being stored in locations that do not adhere to required corporate controls and places the data in locations the larger business is unaware of and cannot adequately protect.

The pricing of these services may fall below corporate spending limits that would otherwise trigger involvement of the vendor management office (VMO) and information security and may simply be placed on a purchase card rather than through an invoice and vendor contract. Without VMO involvement, the corporate master services agreement will not be in effect.

If this behavior is allowed to proliferate, the organization can lose control of its sensitive data and processes. For example, the actuary department at an insurance company may decide to create a file-sharing account on one of several available services. As information security was not involved, company policies, procedures, risk management, and controls programs are not followed. As this is not monitored by the security operations center (SOC), a data breach may go unnoticed, and the data that gives the company a competitive advantage could be stolen, altered, or deleted.

Broad Network Access

Cloud services assume the presence of a network. For public and community clouds, this is the Internet. For a private cloud, it could be the corporate network—generally an IP-based network. In either case, cloud services are not local solutions stored on

your individual computer. They are solutions that exist on a network—in the cloud. Without broad and ubiquitous network access, the cloud becomes inaccessible and is no longer useful.

Not all protocols and services on IP-based networks are secure. Part of the strategy to implementing a secure cloud solution is to choose secure protocols and services. For example, Hypertext Transfer Protocol (HTTP) and File Transfer Protocol (FTP) should not be used to move data to and from cloud services as they pass the data in the clear. HTTP Secure (HTTPS), Secure FTP (SFTP), and other encryption-based transmission should be used so that data in motion may be intercepted but not read.

If you are able to access the cloud service and obtain access to your data anywhere in the world, so can others. The requirement for identification and authentication becomes more important in this public-facing environment. The security of accessing your cloud services over the Internet can be improved in a number of ways including improved passwords, multifactor authentication (MFA), virtual private networks (VPNs), etc. The increased security needs of a system available over the network where security is shared between the CSP and customer makes these additional steps more important.

Multitenancy

One way to get the improved efficiencies of cloud computing is through the sharing of infrastructure. A server may have more than one company purchasing access to its resources. These resources are shared by the tenants. Like an apartment building, these tenants share resources and services but have their own dedicated space. Virtualization allows the appearance of single tenancy in a multitenancy situation. Each tenant's data remains private and secure in the same way that your belongings (data) in an apartment building remain secure and isolated from the belongings (data) of your neighbor.

However, as the building is shared, it is still the responsibility of each tenant to exercise care to maintain the integrity and confidentiality of their own data. If the door is left unsecured, a neighbor could easily enter and take your things. It is also necessary to consider the availability of the data as the actions of another tenant could make your data inaccessible for a time due to no fault of your own. In our example, if another tenant is involved in illegal activity, the entire building could be shut down. Or, if another tenant damaged the building, your access might be reduced or eliminated. A multitenancy environment increases the importance of disaster recovery (DR) and business continuity (BC) planning.

Rapid Elasticity and Scalability

In a traditional computing model, a company would need to buy the infrastructure needed for any future, potential, or anticipated growth. If they estimate poorly, they either will have a lot of excess capacity or will run out of room. Neither situation is optimal. In a

cloud solution, the space needed grows and shrinks as necessary to support the customer. If there is a peak in usage or resource needs, the service grows with the needs. When the needs are gone, the resources used decrease. This supports a pay-as-you-go model, where a customer pays only for the resources needed and used.

For the CSP, this presents a challenge. The CSP must have the excess capacity to serve all their customers without having to incur the cost of the total possible resource usage. They must, in effect, estimate how much excess capacity they must have to serve all of their customers. If they estimate poorly, the customer will suffer and the CSP's customer base could decrease.

However, there is a cost to maintaining this excess capacity. The cost must be built into the cost model. In this way, all customers share in the cost of the CSP, maintaining some level of excess capacity. In the banking world, a bank must keep cash reserves of a certain percentage so that they can meet the withdrawal needs of their customers. But if every customer wanted all of their money at the same time, the bank would run out of cash on hand. In the same way, if every customer's potential peak usage occurred at the same time, the CSP would run out of resources, and the customers would be constrained (and unhappy).

The customer must also take care in setting internal limits on resource use. The ease of expanding resource use can make it easy to consume more resources than are truly necessary. Rather than cleaning up and returning resources no longer needed, it is easy to just spin up more resources. If care is not taken to set limits, a customer can find themselves with a large and unnecessary bill for resources "used."

Resource Pooling

In many ways, this is the core of cloud computing. Multiple customers share a set of resources including servers, storage, application services, etc. They do not each have to buy the infrastructure necessary to provide their IT needs. Instead, they share these resources with each other through the orchestration of the CSP. Everyone pays for what they need and use. The goal is that resources are used efficiently by the group of customers.

This resource pooling presents some challenges for the cybersecurity professional. When resources are pooled, it can lead to multitenancy. A competitor or a rival can be sharing the same physical hardware. If the system, especially the hypervisor, is compromised, sensitive data could be exposed.

Resource pooling also implies that resources are allocated and deallocated as needed. The inability to ensure data erasure can mean that remnants of sensitive files could exist on storage allocated to another user. This increases the importance of data encryption and key management.

Measured Service

Metering service usage allows a CSP to charge for the resources used. In a private cloud, this can allow an organization to charge each department based on their usage of the cloud. For a public cloud, it allows each customer to pay for the resources used or consumed. With a measured service, everyone pays their share of the costs.

The cloud is especially advantageous for organizations with peaks in their resource needs or cycles of usage. For example, a tax preparer uses more resources in the United States in the beginning of the year, peaking on April 15. Many industries have sales dates: Memorial Day, President's Day, Black Friday, Cyber Monday, Arbor Day, etc. Okay, maybe not Arbor Day. Resource needs peak at these times. A company can pay for the metered service for these peak times rather than maintaining the maximum resource level throughout the year. Maintaining the maximum resources in-house would be expensive and a waste of resources.

Building Block Technologies

These technologies are the elements that make cloud computing possible. Without virtualization, there would be no resource pooling. Advances in networking allow for ubiquitous access. Improvements in storage and databases allow remote virtual storage in a shared resource pool. Orchestration puts all the pieces together. The combination of these technologies allows better resource utilization and improves the cost structure of technology. Providing the same resources on-premise can also be accomplished by these technologies, but with lower resource utilization and at a higher cost in many situations. Where costs are not decreased by cloud computing, a case for on-premise resources can be made.

Virtualization

Virtualization allows the sharing of servers. Virtualization is not unique to cloud computing and can be used to share corporate resources among multiple process and services. For example, a service can have VMware installed and run a mail server on one virtual machine (VM) and a web server on another VM, both using the same physical hardware. This is resource sharing.

Cloud computing takes this idea and expands it beyond what most companies are capable of doing. The CSP shares resources among a large number of services and customers (also called *tenants*). Each tenant has full use of their environment without knowledge of the other tenants. This increases the efficient use of the resources significantly.

In addition, a CSP may have multiple locations. This allows services and data to move seamlessly between locations, improving resource use by the CSP. Services and

data can easily be in multiple locations, improving business continuity and fault tolerance. The CSP can use the ease with which virtualization allows the movement of data and services to take advantage of available space and excess capacity, wherever it may be located.

This can create some security and compliance concerns, when data cannot move freely across borders or jurisdictional issues exist. These issues are best handled during contract negotiation. Another concern is if the hypervisor is compromised, as it controls all VMs on a machine. If the hypervisor is compromised, all data can be compromised. The security of the hypervisor is the responsibility of the CSP.

Storage

A variety of storage solutions allow cloud computing to work. Two of these are storage area networks (SANs) and network-attached storage (NAS). These and other advances in storage allow a CSP to offer flexible and scalable storage capabilities.

A SAN provides secure storage among multiple computers within a specific customer's domain. A SAN appears like a single disk to the customer, while the storage is spread across multiple locations. This is one type of shared storage that works across a network.

Another type of networked storage is the NAS. This network storage solution uses TCP/IP and allows file-level access. A NAS appears to the customer as a single file system. This is a solution that works well in a cloud computing environment.

The responsibility for choosing the storage technology lies with the CSP and will change over time as new technologies are introduced. These changes should be transparent to the customer. The CSP is responsible for the security of the shared storage resource.

Shared storage can create security challenges if file fragments remain on a disk after it has been deallocated from one customer and allocated to another. A customer has no way to securely wipe the drives in use, as the customer does not control the physical hardware. However, the use of crypto-shredding can make these fragments unusable if recovered.

Networking

As all resources in a cloud environment are accessed through the network, a robust, available network is an essential element. The Internet is the network used by public and community clouds, as well as many private clouds. This network has proven to be widely available with broad capabilities. The Internet has become ubiquitous in society, allowing for the expansion of cloud-based services.

An IP-based network is only part of what is needed for cloud computing. Low latency, high bandwidth, and relatively error-free transmissions make cloud computing possible. The use of public networks also creates some security concerns. If access to cloud

resources is via a public network, like the Internet, the traffic can be intercepted. If transmitted in the clear, the data can be read. The use of encryption and secure transport keeps the data in motion secure and cloud computing safer.

Databases

Databases allow for the organization of customer data. By using a database in a cloud environment, the administration of the underlying database becomes the responsibility of the CSP. They become responsible for patching, tuning, and other database administrator services. The exception is IaaS, where the user is responsible for whatever database they install.

The other advantage of databases offered through a cloud service is the number of different database types and options that can be used together. While traditional relational databases are available, so are other types. By using traditional databases and other data storage tools as well as large amounts of data resources, data warehouses, data lakes, and other data storage strategies can be implemented.

Orchestration

Cloud orchestration is the use of technology to manage the cloud infrastructure. In a modern organization, there is a great deal of complexity. This has been called the multicloud. An organization may contract through the VMO with multiple SaaS services. In addition, they may have accounts with multiple CSPs, such as AWS, IBM Cloud Foundry, and Microsoft Azure. In addition, they may be using public, private, and community clouds.

This complexity could lead to data being out of sync, processes being broken, and the workforce unable to keep track of all the part. Like the conductor of an orchestra, cloud orchestration partners keep all of these pieces working together including data, processes, and application services. Orchestration is the glue that ties all of the pieces together through programming and automation. Orchestration is valuable whether an organization runs a single cloud environment or a multicloud environment.

This is more than simply automating a task here and a task there. However, automation is used by the cloud orchestration service to create one seemingly seamless organizational cloud environment. In addition to hiding much of the complexity of an organization's cloud environment, cloud orchestration can reduce costs, improve efficiency, and support the overall workforce.

The major CSPs provide orchestration tools. These include IBM Cloud Orchestrator, Microsoft's OMS Management Suite, Oracle Cloud Management Solutions, and AWS Cloud Formation. Like all such offerings, they vary considerably in the tools provided and the integration with other vendors' cloud offerings.

DESCRIBE CLOUD REFERENCE ARCHITECTURE

The purpose of a reference architecture (RA) is to allow a wide variety of cloud vendors and services to be interoperable. An RA creates a framework or mapping of cloud computing activities and cloud capabilities to allow the services of different vendors to be mapped and potentially work together more seamlessly. An example of this approach is the seven-layer Open Systems Interconnection (OSI) model of networking, which is used to discuss many networking protocols. As companies are engaging in a wide variety of cloud solutions from multiple vendors, interoperability is becoming more important, and the reference architecture helps make that more easily occur.

The National Institute of Standards and Technology (NIST) provides a cloud computing reference architecture in SP 500-292 as do other organizations. Some models, such as NIST are role based. Other RAs, such as the IBM conceptual reference model, are layer based. The NIST RA is intended to be vendor neutral and defines five roles: cloud consumer, cloud provider, cloud auditor, cloud broker, and cloud carrier.

Cloud Computing Activities

Cloud computing activities in an RA depend on whether the RA is role based or layer based. As an example, the role-based NIST RA will be used to describe cloud computing activities. A similar description could be made for a layer-based model. In a role-based RA, cloud computing activities are the activities of each of the roles. The NIST model includes five roles, with the following types of activities:

- **Cloud consumer**: The procurement and use of cloud services. This involves reviewing available services, requesting services, setting up accounts and executing contracts, and using the service. What the activities consist of depends on the cloud service model. For a SaaS consumer, the activities are typical end-user activities such as email, social networks, and collaboration tools. The activities with a PaaS customer center around development activities, business intelligence, and application deployment. IaaS customers focus on activities such as business continuity and disaster recovery, storage, and compute.
- **Cloud provider**: The entity that makes a service available. These activities include service deployment, orchestration, and management as well as security and privacy.
- **Cloud auditor**: An entity capable of independent examination and evaluation of cloud service controls. These activities are especially important for entities with contractual or regulatory compliance obligations. Audits are usually focused on compliance, security, or privacy.

- **Cloud broker**: This entity is involved in three primary activities: aggregation of services from one or several CSPs, integration with existing infrastructure (cloud and noncloud), and customization of services.
- **Cloud carrier**: The entity that provides the network or telecommunication connectivity that permits the delivery and use of cloud services.

Cloud Service Capabilities

Capability types are another way to look at cloud service models. In this view, we look at the capabilities provided by each model. Our three service models are SaaS, PaaS, and IaaS. Each provides a different level and type of service to the customer. The shared security responsibilities differ for each type as well.

Application Capability Types

Application capabilities include the ability to access an application over the network from multiple devices and from multiple locations. Application access may be made through a web interface, through a thin client, or in some other manner. As the application and data are stored in the cloud, the same data is available to a user from whichever device they connect from. Depending on the end user, the look of the interface may be different.

Users do not have the capability to control or modify the underlying cloud infrastructure, although they may be able to customize their interface of the cloud solution. What the user gets is a positive experience when working on a laptop or phone. The organization does not have to be concerned with the different types of endpoints in use in their organization (as it relates to cloud service access). Supporting all of the different types of devices is the responsibility of the application service provider.

Platform Capability Types

A platform has the capability of developing and deploying solutions through the cloud. These solutions may be developed with available tools, they may be acquired solutions that are delivered through the cloud, or they may be solutions that are acquired and customized prior to delivery. The user of a platform service may modify the solutions they deploy, particularly the ones they develop and customize. However, the user has no capability to modify the underlying infrastructure.

What the user gets in a platform service are tools that are specifically tailored to the cloud environment. In addition, the user can experiment with a variety of platform tools, methods, and approaches to determine what is best for a particular organization or development environment without the expense of acquiring all those tool and the

underlying infrastructure costs. It provides a development sandbox at a lower cost than doing it all in house.

Infrastructure Capability Types

An infrastructure customer cannot control the underlying hardware but has control over the operating system, installed tools, solutions installed, and provisioning of infrastructure compute, storage, and network and other computing resources.

This capability provides the customer with the ability to spin up an environment quickly. The environment may be needed for only hours or days. The parent organization does not have to purchase the hardware or physical space for this infrastructure or pay for its setup and continuing maintenance for usage spikes, temporary needs, or even regular cycles of use.

Cloud Service Categories

There are three primary cloud service categories: SaaS, PaaS, and IaaS. In addition, other service categories are sometimes suggested, such as storage as a service (STaaS), database as a service (DBaaS), and even everything as a service (XaaS). However, these can be described in terms of the three basic types and have not caught on in common usage. They are most often used in marketing.

Security of systems and data is a shared responsibility between the customer and service provider. The point at which responsibilities of the service provider end and the responsibilities of the customer begin depends on the service category.

When talking about SaaS, PaaS, or IaaS solutions, we must know which service model is being discussed. Each is discussed in some detail next. Which model you are referring to is in part determined by where in the process you are.

If you are an end user, you are likely using a SaaS solution. If you are a developer, you may be offering a SaaS solution you developed in-house or through the use of a PaaS development environment. It is possible that the cloud service you provide is a development environment, so you offer a PaaS service you built on an IaaS service. Some customers work at all three levels. They use an IaaS service to build a development environment to create a SaaS solution. In each case, the security responsibilities are shared, as described elsewhere, by the customer and the CSP. However, that shared responsibility can become rather complex if the customer uses multiple services at differing service levels.

Software as a Service

SaaS is the most common cloud service that most people have experience with. This is where we find the end user, which at times is each of us. If you have shared a file through Google Docs, stored a file on Dropbox, signed a document using DocuSign, or created a

document with Office 365, you have used a SaaS solution. They are usually subscription-based services and are easy to set up and use. Corporations often negotiate and purchase a site license. The amount of control over security will vary by the CSP and the size of the contract.

Platform as a Service

PaaS is the domain of developers. With a PaaS solution, the service provider is responsible for infrastructure, networking, virtualization, compute, storage, and operating systems. Everything built on top of that is the responsibility of the developer and their organization. Many PaaS service providers offer tools that may be used by the developers to create their own applications. How these tools are used and configured are the responsibility of the developers and their organizations.

With a PaaS solution, a developer can work from any location with an Internet connection. The developer's organization no longer has to provide the servers and other costly infrastructure needed. This can be especially useful when testing new solutions and developing experimental ideas. In addition, the CSP provides patching and updates for all services provided. Major CSPs offer PaaS solutions.

Infrastructure as a Service

IaaS is where we find the system administrators (SysAdmins). In a typical IaaS offering, the IaaS service provider is responsible for the provisioning of the hardware, networking, and storage, as well as any virtualization necessary to create the IaaS environment. The SysAdmin is responsible for everything built on top of that, including the operating system, developer tools, and end-user applications as needed.

The IaaS service may be created to handle resource surge needs, to create a development environment for a distributed DevOps team, or even to develop and offer SaaS products.

Cloud Deployment Models

There are three cloud deployment models and one hybrid model. The hybrid model is a combination of any two or more other deployment models. Each deployment model has advantages and disadvantages. A cloud deployment model tells you who owns the cloud and who can access the cloud—or at least, who controls access to the cloud. The deployment model may also tell you something about the size of the cloud.

Public Cloud

In a public cloud, anyone with access to the Internet may access the resources provided, usually through a subscription-based service. The resources and application services are provided by third-party service providers, and the systems and data reside on third-party

servers. For example, Dropbox provides a file storage product to end users. The details of how Dropbox provides this service are for the business to determine. For the customer, it is simply a publicly available cloud service.

There are concerns with privacy and security in a public cloud. And, while that may have been the case in the past, public clouds have made great strides in both privacy and security. The responsibility for both—data privacy and security—remains with the data owner (customer). Concerns about reliability can sometimes be handled contractually through the use of an service-level agreement (SLA). However, for many public cloud services, the contractual terms are fixed for both individual or corporate accounts.

Concerns also exist for vendor lock-in and access to data if the service provider goes out of business or is breached. The biggest drawback may be in customization. A public cloud provides those services and tools it determines will be profitable, and the customer often must choose from among the options provided. Each cloud service provider has a varied set of tools.

Private Cloud

A private cloud is built in the same manner as a public cloud, architecturally. The difference is in ownership. A private cloud belongs to a single company and contains data and services for use by that company. There is not a subscription service for the general public. In this case, the infrastructure may be built internally or hosted on third-party servers.

A private cloud is usually more customizable, and the company controls access, security, and privacy. A private cloud is also generally more expensive. There are no other customers to share the infrastructure costs. With no other customers, the cost of providing excess capacity is not shared.

A private cloud may not save on infrastructure costs, but it provides cloud services to the company's employees in a more controlled and secure fashion. The major cloud vendors provide both a public cloud and the ability for an organization to build a private cloud environment.

The primary advantage to a private cloud is security. With more control over the environment and only one customer, it is easier to avoid the security issues of multitenancy. And when the cloud is internal to the organization, a secure wipe of hardware becomes a possibility.

Community Cloud

A community cloud falls somewhere between public and private clouds. The cloud is built for the needs of multiple organizations, all in the same industry. These common

industries might be banks; governments such as a group of states; or resources shared between local, county (or parish), and state governments. Universities often set up consortiums for research, and this can be facilitated through a community cloud. Structured like public and private clouds, the infrastructure may be hosted by one of the community partners or by a third-party. Access is restricted to members of the community and may be subscription based.

While a community cloud can facilitate data sharing among similar entities, each remains independent and is responsible for what it shares with others. As in any other model, the owner of the data remains responsible for its privacy and security, sharing only what is appropriate, when it is appropriate.

Hybrid Cloud

A hybrid cloud can be a combination of any of the other cloud deployment models but is usually a combination of the private and public cloud deployment models and can be used in ways that enhance security when necessary and allows scalability and flexibility.

When an organization has highly sensitive information, the additional cost of a private cloud is warranted. The private cloud provides the access, resource pooling, and other benefits of a cloud deployment in a more secure fashion.

However, an organization will also have less sensitive information (e.g., email, memos, and reports). In most cases, the amount of this data is much larger. A public cloud can provide the benefits of cloud computing in a cost-effective manner for this less sensitive data. As most of an organization's data is usually of the less sensitive type, the cost savings of a public cloud realized can be substantial, while protecting the more sensitive data in the private cloud. The overall cost savings remains, and the benefits of cloud computing are realized.

In a hybrid model, the disadvantages and benefits of each type of cloud deployment remains for the portion of the cloud using that deployment model. Cloud orchestration can be used to keep this hybrid cloud manageable for the workforce to use.

Cloud Shared Considerations

All cloud customers and CSPs share a set of concerns or considerations. It is no longer the case that all companies use a single CSP or SaaS vendor. In fact, larger companies may use multiple vendors and two or more CSPs in their delivery of services. The business choice is to use the best service for a particular use (best being defined by the customer based on features, cost, or availability). The sections that follow discuss some major considerations that allow the use of multiple CSPs and vendors, in support of the complex cloud environment that exists.

Interoperability

With the concern over vendor lock-in, interoperability is a primary consideration. Interoperability creates the ability to communicate with and share data across multiple platforms and between traditional and cloud services provided by different vendors. Avoiding vendor lock-in allows the customer to make decisions based on the cost, feature set, or availability of a particular service regardless of the vendor providing the service. Interoperability leads to a richer set of alternatives and more choices in pricing.

Portability

Portability may refer to data portability or architecture portability. Data portability is focused on the ability to move data between traditional and cloud services or between different cloud services without having to port the data under challenging and lossy methods or significant changes to either service or the loss of metadata.

Data portability matters to an organization that uses a multicloud approach, as data moves between vendors. Each move cannot create a data porting exercise, or it is not seamless or useful. It is also important in a loud bursting scenario, where peak usage expands into a cloud environment and then shrinks back to its original noncloud size. This must be seamless to make the strategy useful. Data backups are increasingly to the cloud, and a restore to in-house servers must be handled easily.

Architecture portability is concerned with the ability to access and run a cloud service from a wide variety of devices, running different operating systems. This allows users on a Windows laptop and a MacBook Pro to use the same application services, share the same data, and collaborate easily.

Reversibility

Reversibility is a measure of the extent your cloud services can be moved from one cloud environment to another. This includes moving between a cloud environment and an on-premise traditional environment. The movement between environments must be simple and automatic. Companies now move to and from the cloud and between clouds in a multicloud environment and when cloud bursting.

The movement between environments needs to be secure or the movement is not simple nor low cost. Reversibility also decreases vendor lock-in as solutions need to be able to move between CSPs and to and from the cloud. It will become important as application software and data will eventually reside in different locations and the mature cloud environment will not care.

Availability

Availability has two components. The first is one leg of the CIA triad. Within the constraints of the agreed-upon SLA, the purchased services and company or individual data must be made available to the customer by the CSP. If the SLA is not met, the contract will spell out the penalties or recourses available. In this example, if a customer has paid for Dropbox, but when they try to access the service, it is not available, the service availability fails. If this failure is not within the requirements of the SLA, the customer has a claim against the service provider.

The second component of availability is concerned with the elasticity and scalability of the cloud service. If the CSP has not properly planned for expansion, a customer may need to grow their use of the contracted service, and the resources may not be available. Consider a service like Dropbox. If the customer pays for 2TB of storage and it is not available, when they need it, the service fails in terms of availability, even if access to files already stored with the service continues to be provided.

Security

Cloud security is a challenging endeavor. It is true that the larger CSPs spend resources and focus on creating a secure environment. It is equally true that a large CSP is a large target, and there are aspects of cloud computing, such as multitenancy, that create new complexities to security.

One issue that is part of various national laws such as the European Union's General Data Protection Regulation is the restriction on cross-border transfers of data. In an environment where the actual hardware could be anywhere, it is an important consideration to know where your data resides. When there are law enforcement issues, location of the data may also be a jurisdictional challenge.

The owner of data remains ultimately responsible for the security of the data, regardless of what cloud or noncloud services are used. Cloud security involves more than protection of the data but includes the applications and infrastructure.

Privacy

The involvement of third-party providers, in an off-premises situation, creates challenges to data protection and privacy. The end user cannot always determine what controls are in place to protect the privacy of their data and must rely on privacy practice documents and other reports to determine if they trust the third party to protect their data privacy.

Privacy concerns include access to data both during a contract and at the end of a contract as well as the erasure or destruction of data when requested or as required within the contract. Regulatory and contractual requirements such as HIPAA and PCI are also key concerns. Monitoring and logging of data access and modification, and the location of data storage, are additional privacy concerns.

Resiliency

Resilience is the ability to continue operating under adverse or unexpected conditions. This involves both business continuity and disaster recovery planning and implementation. Business continuity might dictate that a customer stores their data in multiple regions so that a service interruption in one region does not prevent continued operations.

The cloud also provides resiliency when a customer suffers a severe incident such as weather, facilities damage, terrorism, civil unrest, or similar events. A cloud strategy allows the company to continue to operate during and after these incidents. The plan may require movement of personnel or contracting personnel at a new location. The cloud strategy handles the data and processes as these remain available anywhere network connectivity exists.

Major CSPs use multiple regions and redundancy to increase the ability of a recovery. Many organizations plan a resilient strategy that includes internal resources and the capabilities of the cloud.

Performance

Performance is measured through an SLA. Performance of a cloud service is generally quite high as major CSPs build redundancy into their systems. The major performance concerns are network availability and bandwidth. A network is a hard requirement of a cloud service, and if the network is down, the service is unavailable. In addition, if you are in an area of limited bandwidth, performance will be impacted.

Governance

Cloud governance uses the same mechanisms as governance of your on-premises IT solutions. This includes policies, procedures, and controls. Controls include encryption, access control lists (ACLs), and identity and access management. As many organizations have cloud services from multiple vendors, a cloud governance framework and application can make the maintenance and automation of cloud governance manageable. This may be another cloud solution.

A variety of governance solutions, some cloud based, exist to support this need. Without governance, cloud solutions can easily grow beyond what can be easily managed. For example, a company may want to govern the number of CSP accounts, the number of server instances, the amount of storage utilized, the size of databases, and other storage tools. Each of these add to the cost of cloud computing. A tool that tracks usage and associated costs will help an organization use the cloud efficiently and keep its use under budget.

Maintenance and Versioning

Maintenance and versioning in a cloud environment have some advantages and disadvantages. Each party is responsible for the maintenance and versioning of their portion of the cloud stack. In a SaaS solution, the maintenance and versioning of all parts is the responsibility of the CSP, from the hardware to the SaaS solution. In a PaaS solution, the customer is responsible for the maintenance and versioning of the applications they acquire and develop. The platform and tools provided by the platforms, as well as the underlying infrastructure, are the responsibility of the CSP. In an IaaS solution, the CSP is responsible for maintenance and versioning of hardware, network and storage, and the virtualization software. The remainder of the maintenance and versioning is the responsibility of the customer.

What this means in practical terms is that updates and patches in a SaaS or PaaS environment may occur without the knowledge of the customer. If properly tested before being deployed, it will also be unnoticed by the customer. There remains the potential for something to break when an update or patch occurs, as it is impossible to test every possible variation that may exist in the cloud environment of the customers. This is true in a traditional on-premise environment as well. In an IaaS environment, the customer has much more control over patch and update testing and deployment.

On the positive side, there will not be the endpoints that exist in every organization that never get updated and have older, insecure versions of potentially unlicensed software. When connecting to the cloud service, the customer will always be using the newest, most secure version of the solution in a SaaS solution.

In a PaaS or IaaS, the customer is responsible for some of the maintenance and versioning. However, each customer that connects to the PaaS and IaaS environment will be accessing the most current version provided. The maintenance and versioning are simplified by restricting the maintenance and versioning to the cloud environment. It is not necessary to update each endpoint running a particular piece of software. Everyone connecting to the cloud is running the same version, even if it is old and has not been updated.

Service Levels and Service Level Agreements

Contractually, an SLA specifies the required performance parameters of a solution. This negotiation will impact the price, as more stringent requirements can be more expensive. For example, if you need 24-hour support, this will be less expensive than 4-hour support.

Some CSPs will provide a predefined set of SLAs, and customers choose the level of service they need. The customer can be an individual or an organization. For the customer contracting with a CSP, this is a straightforward approach. The CSP publishes their performance options and the price of each, and the customer selects the one that best suits their needs and resources.

In other cases, a customer specifies their requirements, and the CSP will provide the price. If the CSP cannot deliver services at the level specified or if the price is more than the customer is willing to pay, the negotiation continues. Once agreed upon, the SLA becomes part of the contract. This is generally true only for large customers. The cost of negotiating and customizing an SLA and the associated environment is not generally cost effective for smaller contracts and individuals.

Auditability

A cloud solution needs to be auditable. This is an independent examination of the cloud services controls, with the expression of an opinion on their function with respect to their purpose. Are the controls properly implemented? Are the controls functioning and achieving their goal? These are the questions of an auditor.

A CSP will rarely allow a customer to perform on audit on their controls. Instead, independent third parties will perform assessments that are provided to the customer. Some assessments require a nondisclosure agreement (NDA), and others are publicly available. These include SOC reports, vulnerability scans, and penetration tests.

Regulatory

Proper oversight and auditing of a CSP makes regulatory compliance more manageable. A regulatory environment is one where a principle or rule controls or manages an organization. Governance of the regulatory environment is the implementation of policies, procedures, and controls that assist an organization in meeting regulatory requirements.

One form of regulations are those governmental requirements that have the force of law. The Health Insurance Portability and Accountability Act (HIPAA), Gramm-Leach-Bliley Act (GLBA), Sarbanes-Oxley Act (SOX) in the United States, and GDPR in the European Union are examples of laws that are implemented through regulations and have the force of law. If any of these apply to an organization, governance will put a framework in place to ensure compliance with these regulations.

Another form of regulations is those put in place through contractual requirements. An SLA takes the form of a contractual obligation as do the rules associated with credit and debit cards through the Payment Card Industry Data Security Standard (PCI DSS). Enforcement of contractual rules can be through the civil courts governing contracts. Governance must again put in place the framework to ensure compliance.

A third form of regulations is found through standards bodies like International Organization for Standardization (ISO) and NIST as well as nongovernmental groups such as the Cloud Security Alliance and the Center for Internet Security. These organizations make recommendations and provide best practices in the governance of security and risk. These support improved security and risk management. While this form of regulation does not usually have the force of law, an organization or industry may voluntarily choose to be regulated by a specific set of guidelines. For example, U.S. federal agencies are required to follow NIST requirements. If an organization or industry chooses to follow a set of guidelines under ISO, NIST, or other group, they must put the governance framework in place to ensure compliance. While often voluntary, once an organization chooses to follow these guidelines, the governance process ensures the organization complies with these regulations.

Impact of Related Technologies

The technologies in this section may be termed *transformative technologies*. Without them, the cloud computing still works and retains its benefits. These transformative technologies either improves your capabilities in the cloud or expands the capabilities and benefits of cloud computing. In the following sections, the specific use cases for the technology will be described.

Machine Learning

Machine learning (ML) is a key component of artificial intelligence (AI) and is becoming more widely used in the cloud. Machine learning creates the ability for a solution to learn and improve without the use of additional programming. Many of the CSPs provide ML tools. There is some concern and regulatory movement when ML makes decisions about individuals without the involvement of a person in the process.

The availability of large amounts of inexpensive data storage coupled with vast amounts of computing power increases the effectiveness of ML. A data warehouse, or even a data lake, can hold amounts of data that could not be easily approached before. ML tools can mine this data for answers to questions that could not be asked before because of the computing power required. This capability has the potential to transform how we use data and the answers we can extract from our data.

The security concern has to do with both the data and the processing. If all of your data is available in one large data lake, access to the data must be tightly controlled. If your data store is breached, all of your data is at risk. Controls to protect the data at rest and access to this data are crucial to make this capability safe for use.

The other concern is with how the data is used. More specifically, how will it impact the privacy of the individuals whose data is in the data store? Will questions be asked where the answers can be used to discriminate against groups of people with costly characteristics? Might insurance companies refuse to cover individuals when the health history of their entire family tree suggests they are an even greater risk than would be traditionally believed?

Governmental bodies and Non-Governmental Organizations (NGOs) are addressing these concerns to some degree. For example, Article 22 of the EU GDPR has a prohibition on automated decision-making, which often involves ML, when that decision is made without human intervention if the decision has a significant impact on the individual. For example, a decision on a mortgage loan could involve ML. The final loan decision cannot be made by the ML solution. A human must review the information and make the final decision.

Artificial Intelligence

Machine learning is not the only AI technology. The goal of AI is to create a machine that has the capabilities of a human and cannot be distinguished from a human. It is possible that AI could create intelligent agents online that are indistinguishable to human agents. This has the potential to impact the workforce, particularly in the lower skill areas. There is also concern about how agents could be manipulated to affect consumer behavior and choices. An unethical individual could use these tools to impact humanity. Safeguards in the technology and legal protections will need to be in place to protect the customers.

With the vast amount of data in the cloud, the use of AI is a security and privacy concern beyond the data mining and decision-making of ML. This greater ability to aggregate and manipulate data through the tools created through AI research creates growing concerns over security and privacy of that data and the uses that will be devised for this data.

These concerns and trends will continue to be important over the next several years.

Blockchain

Blockchain is similar to cloud computing, with some significant differences. A blockchain is an open distributed ledger of transactions, often financial, between two parties. This transaction is recorded in a permanent and verifiable manner. The records, or *blocks*, are linked cryptographically and are distributed across a set of computers, owned by a variety of entities.

Blockchain provides a secure way to perform anonymous transactions that also maintain nonrepudiation. The ability to securely store a set of records across multiple servers, perhaps in different CSPs or on-premise, could lead to new and powerful storage approaches. Any data transaction would be committed to the chain and could be verifiable and secure. Blockchain technology pushes the boundaries of cryptographic research in ways that support secure distributed computing.

In cloud computing, the data may be owned by a single entity. But, the ability to securely store this data across CSPs would open new storage methods and would lead to less vendor lock-in. Each data node could be in any location, on any server, within any CSP or on-premise, where each node in the data chain is not important. While not every record in the cloud is the result of a financial transaction, all data records are the result of some transaction.

Other improvements in the use of cryptography to link records in an immutable manner or improvements in the techniques used to distribute records across multiple servers would benefit both blockchain and cloud computing.

Internet of Things

With the growth of the Internet of Things (IoT), a great deal of data is being generated and stored. The cloud is a natural way to store this data. Particularly for large organizations, with IoT devices such as thermostats, cameras, irrigation controllers, and similar devices, the ability to store, aggregate, and mine this data in the cloud from any location with a network connection is beneficial.

The manufacturers of many IoT devices do not even consider the cybersecurity aspects of these devices. To an HVAC company, a smart thermostat may simply be a thermostat. These devices can be in service for many years and never have a firmware update. Patches and security updates are simply not installed, and these devices remain vulnerable.

It is not the data on the device that is always the target. The device may become part of a botnet and used in a DDoS attack. Cameras and microphones can be used to surveil individuals. Processes controlled by IoT devices can be interrupted in ways that damage equipment (e.g., Stuxnet) or reputations.

Few organizations are sufficiently mature to really protect IoT devices. This makes these devices more dangerous because they are rarely monitored. The cloud provides the ability to monitor and control a large population of devices from a central location. For some devices, such as a thermostat, this may be a small and acceptable risk. However, audio and visual feeds raise privacy, security, and safety concerns that must be addressed.

Containers

Virtualization is a core technology in cloud computing. It allows resource pooling, multitenancy, and other important characteristics. Containers are one approach to the

virtualization. In a traditional virtualization environment, the hypervisor sits atop the host OS. The VM sits atop the hypervisor. The VM contains the guest OS and all files and applications needed in that VM. A machine can have multiple VMs, each running a different machine.

In containerization, there is no hypervisor and no guest OS. A container runtime sits above the host OS, and then each container uses the container runtime to access needed system resources. The container contains the files and data necessary to run, but no guest OS. The virtualization occurs higher in the stack and is generally smaller and can start up more quickly. It also uses fewer resources by not needing an additional OS in the virtual space. The smaller size of the container image and the low overhead are the primary advantages of containers over traditional virtualization.

Containers make a predictable environment for developers and can be deployed anywhere the container runtime is available. Similar to the Java Virtual Machine, a runtime is available for common operating systems and environments. Containers can be widely deployed. This improves portability by allowing the movement of containers from one CSP to another. Versioning and maintenance of the underlying infrastructure do not impact the containers as long as the container runtime is kept current.

The container itself is treated like a privileged user, which creates security concerns that must be addressed. Techniques and servers exist to address each of these security concerns such as a Cloud Access Security Broker (CASB). Security concerns exist and must be carefully managed. All major CSPs support some form of containerization.

Quantum Computing

Quantum computers use quantum physics to build extremely powerful computers. When these are linked to the cloud, it becomes quantum cloud computing. IBM, AWS, and Azure all provide a quantum computing service to select customers. The increased power of quantum computers and the use of the cloud may make AI and ML more powerful and will allow modeling of complex systems available on a scale never seen before. Quantum cloud computing has the ability to transform medical research, AI, and communication technologies.

A concern for quantum computing is that traditional methods for encryption/decryption could become obsolete as the vast power of the cloud coupled with quantum computing makes the search space more manageable. This would effectively break current cryptographic methods. New quantum methods of encryption would be necessary or methods not susceptible to quantum computing.

UNDERSTAND SECURITY CONCEPTS RELEVANT TO CLOUD COMPUTING

Security concepts for cloud computing mirror the same concepts in on-premises security, with some differences. Most of these differences are related to the customer not having access to the physical hardware and storage media. These concepts and concerns will be discussed in the following sections.

Cryptography and Key Management

Cryptography is essential in the cloud to support security and privacy. With multitenancy and the inability to securely wipe the physical drive used in a CSP's data center, information security and data privacy are more challenging, and the primary solution is cryptography.

Data at rest and data in motion must be securely encrypted. A customer will need to be able to determine whether a VM or container has been unaltered after deployment, requiring cryptographic tools. Secure communications are essential when moving data and processes between CSPs as well as to and from on-premise users. Again, cryptography is the solution.

One of the challenges with cryptography has always been key management. With many organizations using a multicloud strategy, key management becomes even more challenging. The questions to answer are

- Where are the keys stored?
- Who manages the keys (customer or CSP)?
- Should a key management service be used?

In a multicloud environment, there are additional concerns:

- How is key management automated?
- How is key management audited and monitored?
- How is key management policy enforced?

The power of a key management service (KMS) is that many of these questions are answered.

The KMS stores keys separately from the data. One benefit of encrypting data at rest is that many data breach laws provide an exemption if the data is encrypted securely. This benefit disappears if the encryption/decryption keys are stored with the data. So, if keys

are to be stored in the cloud, they must be stored separately from the data. Outsourcing this has the benefit of bringing that expertise to the organization. However, like any outsourcing arrangement, you cannot turn it over to the KMS and forget about it. Someone still needs to oversee the KMS.

Using a KMS does not mean that you turn over the keys to another organization any more than using a cloud file repository gives away your data to the service storing your files. You choose the level of service provided by the KMS to fit your organization and needs.

The last three questions—automation, monitoring and auditing, and policy enforcement—are the questions to keep in mind when reviewing the different KMSs available. Like any other service, the features and prices vary, and each organization will have to choose the best service for their situation. A number of CSPs offer cryptographic KMSs. This KMS makes a multicloud environment scalable.

Access Control

There are three types of access control. These are physical access control, technical access control, and administrative access control. In a shared security model, the CSP and the customer have different responsibilities.

Physical access control refers to actual physical access to the servers and data centers where the data and processes of the cloud customer are stored. Physical access is entirely the responsibility of the CSP. The CSP owns the physical infrastructure and the facilities that house the infrastructure. Only they can provide physical security.

Administrative access control refers to the policies and procedures a company uses to regulate and monitor access. These policies include who can authorize access to a system, how system access is logged and monitored, and how frequently access is reviewed. The customer is responsible for determining policies and enforcing those policies as related to procedures for provisioning/deprovisioning user access and reviewing access approvals.

Technical access control is the primary area of shared responsibility. While the CSP is responsible for protecting the physical environment and the company is responsible for the creation and enforcement of policies, both the customer and the CSP share responsibilities for technical access controls.

For example, a CSP may be willing to federate with an organization's identity and access management (IAM) system. The CSP is then responsible for the integration of the IAM system, while the customer is responsible for the maintenance of the system. If a cloud IAM system is used (provided by the CSP or a third party), the customer is responsible for the provisioning and deprovisioning of users in the system and determining

access levels and system authorizations while the CSP or third-party maintains the IAM system.

Logging system access and reviewing the logs for unusual activity can also be a shared responsibility, with the CSP or third-party IAM provider logging access and the customer reviewing the logs or with the CSP providing both services. Either choice requires coordination between the customer and the CSP. Access attempts can come from a variety of devices and locations throughout the world, making IAM an essential function.

Data and Media Sanitization

Internally, it is possible to sanitize storage media as you have physical access to the media. You determine the manner of sanitization to include physical destruction of the storage media. You also determine the schedule for data deletion and media sanitization.

In the cloud this becomes more challenging. The data storage is shared and distributed, and access to the physical media is not provided. The CSP will not allow you access to the physical disks and will certainly not allow their destruction. In addition, data in the cloud is regularly moved and backed up. It may be impossible to determine if all copies of a data item have been deleted. This is a security and privacy concern. The customer will never have the level of control for data and media sanitization that they had when they had physical access and ownership of the storage hardware.

While some CSPs provide access to wipeable volumes, there is no guarantee that the wipe will be done to the level possible with physical access. Encrypted storage of data and crypto-shredding are discussed in the following sections. While not the same as physical access and secure wipe, they provide a reasonable level of security. If, after review, this level of security is not adequate for an organization's most sensitive data, this data should be retained on-premise in customer data centers or on storage media under the direct physical control of the customer.

Overwriting

Overwriting of deleted data occurs in cloud storage over time. Deleted data areas are marked for reuse, and eventually this area will be allocated to and used by the same or another customer, overwriting the data that is there. There is no specific timetable for overwriting, and the data or fragments may continue to exist for some time. Encryption is key in keeping your data secure and the information private. Encrypting all data stored in the cloud works only if the cryptographic keys are inaccessible or securely deleted.

Cryptographic Erase

Cryptographic erasure is an additional way to prevent the disclosure of data. In this process, the cryptographic keys are destroyed (crypto-shredding), eliminating the key

necessary for decryption of the data. Like data and media sanitization and overwriting, encryption is an essential step in keeping your data private and secure. Secure deletion of cryptographic keys makes data retrieval nearly impossible.

Network Security

Broad network access is a key component of cloud computing. However, if you have access to cloud resources over the network, bad actors can also have access. Bad actors threaten the security of the cloud service you are using and can threaten the privacy and security of your data.

There are a number of ways to provide network security. This list is not exhaustive, and the concepts are not mutually exclusive. Network security starts with controlling access to cloud resources through IAM, discussed previously. By controlling access to the cloud resources, we limit their exposure. We may also limit their exposure to the public Internet through VPNs and cloud gateways. The use of VPNs for Internet security is common. Cloud gateways, ingress and egress monitoring, network security groups, and contextual-based security are discussed next. These are major topics within cloud network security, but are not exhaustive in their coverage. New methods are regularly developed to improve network security as vulnerabilities and threats are constantly changing.

Network Security Groups

Security remains an important concern in cloud computing. A network security group (NSG) is one way of protecting a group of cloud resources. The NSG provides a set of security rules or virtual firewall for those resources. The NSG can apply to an individual VM, a network interface card (NIC) for that VM, or even a subnet. The NSG is essentially a layer around the VM, subnet, or other cloud resource, as part of a layered defense strategy. This gives the customer some additional control over security.

Cloud Gateways

A cloud gateway provides a level of security by keeping communication between the customer and the CSP off the public Internet. AWS regions can be connected and the traffic can be routed to any region while staying within the CSP environment.

Contextual-Based Security

Contextual-based security uses context to help secure the enterprise and, in the case of cloud computing, the cloud resources. Context includes things such as identity, determined through the IAM system, location, time of days, or endpoint type. This is more than the heuristics used to determine if unusual behavior is occurring. The context can determine the level of access and what resources may be accessed. For example,

connecting from the corporate network, through a VPN or from public WiFi may provided different levels of access. If a user attempts to access with an endpoint device that is not registered to that use, access may be blocked entirely.

Ingress and Egress Monitoring

Cloud ingress and egress must be carefully monitored. Security is provided by limiting the number of ingress/egress points available to access resources and then monitoring them. This is similar to a castle with a single entrance. It is easier to control access and prevent access by bad actors when the way in and out is carefully defined and controlled.

Ingress controls can block all or some external access attempts from the public Internet. Inbound connections can be limited to those that are in response to a request initiated from within the cloud resource. This limits connections to the Internet to only those requests initiated in the cloud environment or wanted by the cloud environment.

Egress controls are a way to prevent internal resources from connecting to unapproved and potentially dangerous locations on the Internet. If infected, egress monitoring may prevent malware for contacting their command and control locations. Monitoring what data leaves the environment can assist only in data loss prevention.

Virtualization Security

Virtualization is an important technology in cloud computing. It allows for resource sharing and multitenancy. With these benefits come security concerns. Security of the virtualization method is crucial. The two primary methods of virtualization are VMs created and managed through a hypervisor and virtualization through containers.

Hypervisor Security

A hypervisor, such as Hyper-V or vSphere, packages resources into a VM. Creating and managing the VM are both done through the hypervisor. For this reason, it is important that the hypervisor be secure. Hypervisors such as Hyper-V, VMware EXSi, or Citrix XenServer are type I hypervisors or native hypervisors that run on the host's hardware.

A type I hypervisor is faster and more secure but is more difficult to set up than type II hypervisors, such as VMware or VirtualBox, which sit on top of the operating system. These are easier to set up but less secure.

A hypervisor is a natural target of malicious users as they control all the resources used by each VM. If a hacker compromises another tenant on the server you are on and can compromise the hypervisor, they may be able to attack other customers through the hypervisor. Hypervisor vendors are continually working to make their products more secure.

For the customer, security is enhanced by controlling admin access to the virtualization solution, designing security into your virtualization solution, and securing the hypervisor. All access to the hypervisor should be logged and audited. Access to the network should be limited for the hypervisor to only the necessary access. This traffic should be logged and audited. Finally, the hypervisor must remain current, with all security patches and updates applied as soon as is reasonable. More detailed security recommendations are published in NIST SP 800-125A Rev 1 and by hypervisor vendors.

Container Security

Containerization, such as through Docker or LXC, has many benefits and some vulnerabilities. These include resource efficiency, portability, easier scaling, and agile development. Containerization also improves security by isolating the cloud solution and the host system. Security risks occur through inadequate identity and access management and through misconfigured containers. Software bugs in the container software can also be an issue. The isolation of the container from the host system does not mean that security of the host system can be ignored.

The security issues of containerization must first be addressed through education and training. Traditional DevOps practices and methodologies do not always translate to secure containerization. The use of specialized container operating systems is also beneficial as it limits the capabilities of the underlying OS to those functions a container may need. Much like disabling network ports that are unused, limiting OS functionality decreases the attack surface. Finally, all management and security tools used must be designed for containers. A number of cloud-based security services are available.

There are many containerization solutions provided by major CSPs. One can easily find articles that extoll the virtues of one solution over another. As with other areas of technology, which is best is often a matter of who you ask. Determining which solution is best for your organization requires comparing costs and features.

Common Threats

Previous sections dealt with threats that are related to the specific technologies that are key parts of cloud computing, such as virtualization, media sanitization, and network security. However, all other threats that may attack traditional services are also of concern. Controls that are used to protect access to software solutions, data transfer and storage, and identity and access control in a traditional environment must be considered in a cloud environment as well.

UNDERSTAND DESIGN PRINCIPLES OF SECURE CLOUD COMPUTING

As processes and data move to the cloud, it is only right to consider the security implications of that business decision. Cloud computing is as secure as it is configured to be. With careful review of CSPs and cloud services, as well as fulfilling the customer's shared responsibilities for cloud security, the benefits of the cloud can be obtained securely. The following sections discuss methods and requirements that help the customer work securely in the cloud environment.

Cloud Secure Data Lifecycle

As with all development efforts, the best security is the security that is designed into a system. The cloud secure data lifecycle can be broken down into six steps or phases.

- **Create**: This is the creation of new content or the modification of existing content.
- **Store**: This generally happens at creation time. This involves storing the new content in some data repository, such as a database or file system.
- **Use**: This includes all the typical data activities such as viewing, processing, and changing.
- **Share**: This is the exchange of data between two entities or systems.
- **Archive**: Data is no longer used but is being stored.
- **Destroy**: Data has reached the end of its life, as defined in a data retention policy or similar guidance. It is permanently destroyed.

At each of these steps in the data's lifecycle, there is the possibility of a data breach or data leakage. The general tools for preventing these are encryption and the use of data loss prevention (DLP) tools.

Cloud-Based Disaster Recovery and Business Continuity Planning

A business continuity plan (BCP) is focused on keeping a business running following a disaster such as weather, civil unrest, terrorism, fire, etc. The BCP may focus on critical business processes necessary to keep the business going while disaster recovery takes place. A disaster recovery plan (DRP) is focused on returning to normal business operations. This can be a lengthy process. The two plans work together.

In a BCP, business operations must continue, but they often continue from an alternate location. So, the needs of BCP include space, personnel, technology, process, and data. The cloud can support the organization with many of those needs. A cloud solution provides the technology infrastructure, processes, and data to keep the business going.

Availability zones in a region are independent data centers that protect the customer from data center failures. Larger CSPs like AWS, Azure, and Google define regions. Within a region, latency is low. However, a major disaster could impact all the data centers in a region and eliminate all availability zones in that region. A customer can set up their plan to include redundancy across a single region using multiple availability zones, or redundancy across multiple regions to provide the greatest possible availability of your necessary technology, processes, and data.

One drawback of multiregion plans is that the cost grows quickly. For this reason, many organizations only put their most critical data—the core systems that they cannot operate the business without—across two or more regions, but less critical processes and data may be stored in a single region. Functions and data that are on-premise may also utilize cloud backups. But they may not be up and running as quickly as the cloud-based solutions. The business keeps operating, although not all business processes may be enabled.

DRPs rely heavily on data backups. A DRP is about returning to normal operations. And returning the data to the on-premise environment is part of that. After the on-premise infrastructure has been rebuilt, reconfigured, or restored, the data must be returned.

One failure of many DRPs is the lack of an offsite backup or the ability to quickly access that data backup. In the cloud, a data backup exists in the locations (regions or availability zones) you specify and is available from anywhere network access is available. A cloud-based backup works only if you have network access and sufficient bandwidth to access that data. That must be part of the DRP, along with an offsite data backup. A physical, local backup can also be beneficial. Not every disaster destroys the workplace.

Cost-Benefit Analysis

Cloud computing is not always the correct solution. Which is the correct solution is a business decision guided by a cost-benefit analysis. Cloud computing benefits include reduced capital costs as the individual customers no longer have to buy the hardware and system software that the CSP provides. The lowered capital expenses are offset by higher operating costs, as the customers must pay for the services used.

In many countries, capital expenses and operational expenses are treated very differently for tax purposes. For example, capital expenses may be written off or depreciated over a number of years. To write off the entire business of expenses of new infrastructure,

purchased and installed on-premise could take many years. Operational expenses, such as the cost of cloud computing, can usually be written off as a business expense in the year the expense is incurred.

The business must understand the cost and tax implications of moving to the cloud or investing in on-premise infrastructure to make the choice most beneficial to the business. A move to the cloud may be as much (or more) for financial reasons than technical ones. This move should be considered only if the benefits justify the cost or if the benefits lower the costs.

Functional Security Requirements

Functional security requirements can make the move to cloud computing or the governance of cloud computing safer for a customer's information and processes. However, there remains some challenges with cloud computing, including portability, interoperability, and vendor lock-in.

These challenges can be lessened through the use of a vendor management process to ensure standard capabilities, clearly identifying the responsibilities or each party and the development of SLAs as appropriate. For complex or expensive systems, the RFP process can be utilized to clearly state customer requirements. The security requirements should be part of the requirements specified in the RFP and can be part of the process of choosing a vendor. A vendor that cannot meet the customer's security needs can be eliminated early on.

Portability

One-time movement is when a customer moves to a cloud platform, with no intention of moving it again. This is not common. In a modern environment, movement to and from the cloud as well as between cloud services and CSPs is much more common. These movements are not simply a forklift operation where you pick up some on-premises solution and data and drop it into a cloud account. Each CSP uses different tools and templates. So, a move from one CSP to another requires mapping to the other with the associated data cleanup. Moving from your own infrastructure to a CSP has the same challenge.

Frequent movement between CSPs and between a CSP and your own infrastructure is significantly more difficult, and data can be lost or modified in the process, violating availability and integrity rules. Portability means that the movement between environments is possible. Portable movement will move services and data seamlessly and may be automated.

The movement of data between software products is not a new issue. It can be complicated in the cloud by the need to continue paying for the old service while porting to the new one. This puts time pressure on the porting.

Interoperability

With customers using a variety of cloud services, often from different vendors, interoperability is an important consideration. In addition, some situations may require a private cloud sharing data with an on-premises solution. The ability to share data between tools and cloud environments and between clouds and corporate infrastructure is important. One issue is that the security tools and control sets differ between CSPs. A gap in security may result. Careful planning is essential, and the services of a cloud broker may also be warranted.

One way to improve the situation is through application programming interfaces (APIs). If properly designed, the API can bridge the security gap between services and allow the sharing of data and processes across multiple platforms. For example, if a SaaS tool is used to build a data inventory and supports the corporate data/system classification scheme, an API could be built to securely share that information with the governance, risk management, and compliance (GRC) or system inventory tool. This creates a single source for data, sharing it with relevant systems, and removes the potential of multiple data classification in different systems.

Vendor Lock-in

Solving the interoperability and portability challenges will go a long way toward ending vendor lock-in. This occurs when a customer is tied to a specific CSP and moving would incur significant costs including financial, technical, and legal. Vendor lock-in remains a significant concern with cloud computing. Continued advances in virtualization, improvements in portability and interoperability, and a careful design within a reference architecture have decreased this issue.

An additional concern is the use of CSP-specific services. If these are used to build capabilities, moving to a new CSP also impacts this additional capability. This is similar to using nonstandard features of a compiler in the development process. It locks you into that development environment.

One example is the use of AWS CloudTrail. CloudTrail allows auditing of your AWS account in support of governance, risk management, and compliance. If the decision is made to move away from AWS, the GRC functionality will have to be rebuilt with new services, either with the new CSP or with another vendor.

With additional improvements and careful architecture, vendor lock-in should become an issue in the past. Until then, the security challenges of cloud computing in general, and portability and interoperability in particular, remain.

Security Considerations for Different Cloud Categories

In a cloud environment, security responsibilities are shared between the service provider and the customer. In the SaaS model, the customer has the least responsibility, and in

the IaaS model, the customer has the most responsibility. In a PaaS, the responsibility is shared more equally.

The Shared Responsibility Model for cloud services is commonly presented by the major vendors, which are all similar. There is an architecture stack. Some items in the stack are the responsibility of the CSP, and some are the responsibility of the customer. In between, there is an area of varied responsibility. At times, this middle area is the responsibility of the CSP and sometimes of the customer and sometimes both. It is important for the customer to know their responsibilities, especially in this middle region.

A typical architecture stack looks like this:

- Data
- APIs
- Applications/solutions
- Middleware
- Operating systems
- Virtualization (VMs, virtual local area networks)
- Hypervisors
- Compute and memory
- Data storage
- Networks
- Physical facilities/data centers

It is generally understood that the CSP is responsible for the last five items on the list in all delivery models. However, where the line between customer and CSP exists varies beyond that.

The exact split and layer names vary by vendor, but the general principle remains the same. Both the CSP and the customer have some individual security responsibilities, and along the line where these meet, each may have some security responsibilities. The line for each delivery model is explained in the following sections.

Software as a Service

From a security standpoint, you have limited security options with a SaaS solution. Most of the security options are provided by the SaaS provider. The SaaS provider is responsible for the security of the infrastructure, operating system, application, networking, and storage of the information on their service.

In the Shared Responsibility Model, the customer is responsible for their data and may have some responsibility for the APIs. All other layers are the responsibility of the CSP.

The user of a SaaS solution has responsibilities as well. When a service is subscribed to by an organization or an individual, it is important to understand the security policies and procedures of the SaaS provider to the extent possible. In addition, the user determines how information is transferred to the SaaS provider and can do so securely through end-to-end encryption. The SaaS user is responsible for determining how the data is shared. Finally, the user can provide access security through proper use of login credentials, secure passwords, and multifactor authentication when available.

Platform as a Service

In a PaaS solution, security of the underlying infrastructure, including the servers, operating systems, virtualization, storage, and networking, remain the responsibility of the PaaS service provider. The developer is responsible for the security of any solutions developed, and the data used by their application, as well as the user responsibilities of a SaaS application regarding user access and use of the solutions developed.

In the Shared Responsibility Model, this means the customer is responsible for the data, APIs, and applications, with potentially some middleware responsibility.

Infrastructure as a Service

IaaS security leaves most of the responsibility of security with the customer. IaaS service providers secure the portions they are responsible for. These areas include the servers, virtualization, storage, and networking. The IaaS customer is responsible for the security of the operating system and everything built on top of it, including the responsibilities of a PaaS and a SaaS implementation.

In the Shared Responsibility Model, the customer is responsible for everything above the hypervisor. As in the other delivery models, the exact responsibility along this line can vary between the CSP and customer and must be clearly understood in each case.

EVALUATE CLOUD SERVICE PROVIDERS

Evaluation of CSPs is done through objective criteria. This becomes simpler if those criteria are a known standard. Standards are voluntary for some and required for others. However, the use of a standard makes comparisons between products and services more straightforward.

For example, FIPS 140-2, Federal Information Security Management Act (FISMA), and NIST standards are required for those working with the U.S. federal government. PCC DSS is contractually required by those accepting credit card payments.

Federal Information Processing Standards (FIPS), FISMA, and NIST may have been chosen as the standard in some industries but are suggestions and guidelines for everyone else. Internationally, Common Criteria and ISO standards have been chosen as required by some organizations, industries, and countries and serve as recommendations and guidelines for everyone else.

Verification against Criteria

Difference organizations have published compliance criterion. For cloud computing, these are currently regulatory or voluntary standards. The International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) standard is voluntary but may be necessary to work in some parts of the world and may prove advantageous even when not required. PCI DSS is a contractual requirement. The Payment Card Industry (PCI) Security Standards Council publishes the criteria that are required if you are a vendor that wants to accept credit cards as payment.

International Organization for Standardization/International Electrotechnical Commission

ISO/IEC 27017 and 27018 provide guidance for the implementation of cloud security and the protection of personally identifiable information (PII). 27017 added 35 supplemental controls and extended seven existing controls to the original ISO documents. Most CSPs were already compliant with these additional controls or could easily add them. Becoming compliant with this new standard is straightforward

ISO/IEC 27018 serves as a supplement to ISO 27002 and is specifically geared toward PII processors. Like 27017, these principles are recommendations and not requirements. 27018 added 14 supplementary controls and extended 25 other controls. As an international standard, adherence to this standard will help an organization address a wide and ever-changing data protection and privacy environment stretching from GDPR in the EU to standards in Russia, Brazil, the Philippines, and elsewhere around the globe.

While these are recommendations and not requirements, many international corporations strive to be ISO-compliant. In that case, the criteria provided by ISO/IEC become the governing principles of the organization, including the reference framework, cloud service models (of which there are seven instead of just SaaS, PaaS, and IaaS), and the implementation of controls from the approved control set. Auditing the controls and conducting a risk assessment should help identify which controls best address identified risk.

The ISO standard is important for companies in the international marketplace. These standards have wide acceptance throughout the world. These standards also provide an excellent framework for developing cloud services. Cloud services, because of their broad

network access, are more international than many traditional IT services. An international standard is an important consideration.

Payment Card Industry Data Security Standard

The Payment Card Industry Data Security Standard released version 3.2.1 of PCI DSS in 2020. PCI is contractual compliance between the major credit card companies and the vendor. All cloud customers that accept credit cards must comply with all 12 requirements.

In the 12 requirements, the cloud is referenced in only one place and refers to the appendix for shared hosting requirements. These requirements can be summarized as follows:

- Ensure that a customer's processes can only access their data environment.
- Restrict customer access and privileges to their data environment.
- Enable logging and audit trails that are unique to each environment, consistent with requirement 10.
- Provide processes to support forensic investigations.

In addition to these requirements, the general auditability of the cloud environment would be beneficial in assuring compliance with PCI DSS 3.2.1.

System/Subsystem Product Certifications

The following are system/subsystem product certifications.

Common Criteria

Common Criteria (CC) is an international set of guidelines and specifications to evaluate information security products. There are two parts to CC:

- **Protection profile**: Defines a standard set of security requirements for a specific product type, such as a network firewall. This creates a consistent set of standards for comparing like products.
- **Evaluation assurance level**: Scored from level 1 to 7, with 7 being the highest. This measures the amount of testing conducted on a product. It should be noted that a level 7 product is not automatically more secure than a level 5 product. It has simply undergone more testing. The customer must still decide what level of testing is sufficient. One reason to not subject every product to level 7 is the cost involved.

The testing is performed by an independent lab from an approved list. Successful completion of this certification allows sale of the product to government agencies and

may improve competitiveness outside the government market as CC becomes better known. The goal is for products to improve through testing. It also allows a customer to consider two versions of a security product.

FIPS 140-2

CC does not include a cryptographic implementation standard or test. CC is an international standard, and cryptographic standards are country specific. CC leaves cryptography to each country and organization.

For the U.S. federal government, the cryptographic standard is FIPS 140-2. Organizations wanting to do business with the U.S. government must meet the FIPS criteria. Organizations in regulated industries and nonfederal government organizations are increasingly looking to FIPS certification as their standard. As FIPS use increases, additional industries are expected to use FIPS as their cryptographic standard.

Cybersecurity companies are increasingly seeking FIPS certification to increase their market potential and maximize the value of their services.

FIPS requires that encryption (both symmetric and asymmetric), hashing, and message authentication use algorithms from an approved list. This list is in FIPS 140-2. For example, message authentication can use Triple-DES, AES, or HMAC. There are more algorithms out there than are allowed in FIPS.

Being considered FIPS-validated requires testing by one of a few specified labs through four levels of testing. Sometimes a product is referred to as FIPS-compliant, which is a much lower bar, indicating some components of the product have been tested, but perhaps not the entire product. It is important to read the fine print. *Validated* and *compliant* are not the same thing. A CCSP should also become familiar with the new FIPS 140-3, which will be replacing FIPS 140-2 over the next several years.

Summary

In order to discuss the cloud, each individual must be familiar with the terminology surrounding this technology. This understanding includes characteristics of cloud computing, as well as the service models and deployment models of cloud computing. It also includes the role of the CSP in cloud computing and the shared security model that exists between the CSP and the customer. Finally, the technologies that make cloud computing possible are discussed in this chapter alongside the emerging technologies that will support and transform cloud computing in the future. Understanding this chapter will make it easier to access the discussion in each of the following domains.

DOMAIN 2

CCSP®

Cloud Data Security

RESPONSIBILITY FOR MANY ELEMENTS of cloud security are shared between the cloud service provider (CSP) and the cloud consumer, while each party retains exclusive control over some elements. For example, the CSP is always responsible for the physical security of the infrastructure, while the consumer retains control over the identity and access management concerns of their applications and services. When it comes to securing data stored in the cloud, the most important thing to remember is that the consumer is always ultimately accountable, which means they must not only take steps to secure data but also ensure the CSP is implementing adequate security practices.

DESCRIBE CLOUD DATA CONCEPTS

Data is a crucial element to most modern organizations, so processes to provision and secure the systems that store, process, and transmit that data are essential. The cloud strategy for most organizations will include a variety of personnel in different roles, from the top C-level executives or other executive management all the way to operational personnel responsible for day-to-day functions such as data input and processing. To provide adequate security, it is vital to have a model for understanding how data is created, stored, and used in the cloud environment such as the cloud data lifecycle.

Cloud Data Lifecycle Phases

Unlike valuable physical assets such as precious metals or other objects, information can be hard to definitively identify and secure. Data is constantly being generated, used, stored, transmitted, and, once it is no longer valuable, destroyed. In such a dynamic environment, it can be useful to model the phases that information passes through. This model provides a generic way of identifying the broad categories of risks facing the data and associated protections, rather than trying to identify each individual data element in an organization.

The secure cloud data lifecycle is roughly linear, though some data may not go through all phases, and data may exist in multiple phases simultaneously. Regardless, data should be protected in each phase by controls commensurate with its value. Figure 2.1 illustrates the lifecycle.

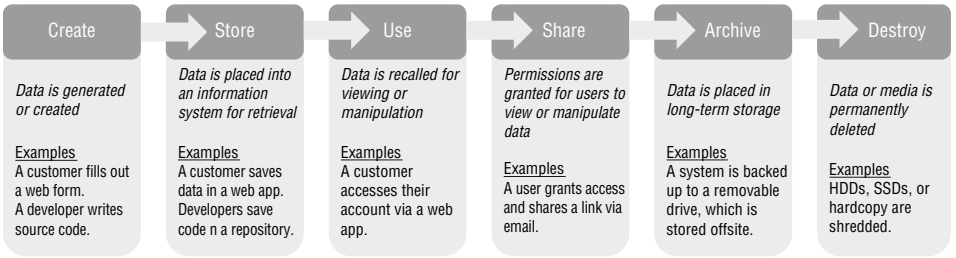

FIGURE 2.1 **The secure data lifecycle**

The cloud data lifecycle is not always iterative; that is, it does not repeat, unlike other lifecycles such as the systems development lifecycle. Data may also exist in one or more of the phases simultaneously; for example, data being created may be shared and stored at the same time if the user saves a file to a shared drive that other users can access. However, data in each phase of the lifecycle faces specific risks, and the cloud security practitioner should thoroughly understand these risks and appropriate mitigations specific to each phase.

- **Create**: Data is created when it is first entered into a system or whenever it is modified. Examples include a user writing a document and saving it, a process collecting data through an API and placing the data into a database, or a user opening a shared document and updating its contents. If data is being modified, the data lifecycle iterates, but it is also normal for data to be created once and never be updated.
  - **Create phase controls**: Data classification is a foundational security control, as it allows the organization to identify data's value and implement appropriate controls. During creation, data should be classified by the creator or owner. In some systems, this might be a manual process such as placing the classification

in a document header/footer. In other cases, the classification of data may be done by a system owner. In this case, the users do not make classification decisions, but all data stored in the system will be classified the same and protected by system-level controls.

- **Store**: Storage is the act of saving data in a retrievable location, such as a solid-state drive (SSD), application database, or hard-copy records. In many cases, the storage phase may occur simultaneously with creation; for example, a user uploads a new document to a fileshare.
  - **Store phase controls**: Data being stored may need protection in transit, especially for cloud services where the data must cross public networks to reach the organization's cloud apps. These protections include the use of Transport Layer Security (TLS), a virtual private network (VPN), Secure Shell (SSH), or other secure data in transit controls. The actual act of storing data should be governed by policies and procedures, such as restrictions on where data of differing classification levels may be stored, access control and granting procedures, and technical controls such as encryption implemented by default for highly sensitive data. Once stored, data should be protected using appropriate access controls and encryption management to preserve confidentiality, as well as adequate backups to preserve both integrity and availability.
- **Use**: The use phase comprises accessing, viewing, and processing data, sometimes collectively referred to as *handling*. Data can be handled in a variety of ways, such as accessing a web application, reading and manipulating files, or fetching data via an API.
  - **Use phase controls**: The use phase is typically the longest lasting phase of the data lifecycle—just as systems spend most of their life in the operations and maintenance phase while being actively used. The number of controls is commensurate, such as managing data flow with data loss prevention (DLP), information rights management (IRM), system access controls such as authorization and access reviews, network monitoring tools, and the like. Accountability controls are also crucial in this phase, which requires adequate logging and monitoring of access. Note the data states, which are data in use, data in transit, and data at rest, are not directly related to this data lifecycle phase. Data being actively "used" by a company may be stored on a laptop hard drive (data at rest) or be sent to a web app for processing (data in transit). The use phase is simply when data is actively being accessed or handled.
- **Share**: The share phase is relatively self-explanatory: access to data is granted to other users or entities that require access. Not all data will be shared, and not all sharing decisions will be made explicitly. Highly classified data may be severely

restricted to access by just a few critical personnel, while less sensitive data may be shared by virtue of being stored in a shared system like a collaboration app.

- **Share phase controls**: Access controls will form the majority of security safeguards during the share phase, both proactive and reactive. Proactive access controls include role-based access authorizations and access-granting procedures. Reactive controls such as DLP, IRM, and access reviews can detect when unauthorized sharing occurs and, in some cases, prevent the shared data from leaving organizational control.

- **Archive**: Data may reach the end of its useful life but still need to be retained for a variety of reasons, such as ongoing legal action or a compliance requirement. Archiving data is the act of placing it into long-term retrievable storage. This may be required depending on the type of organization and data you are processing, such as financial or healthcare data, which has mandated retention requirements. Archiving is also an operational cost benefit, as it moves data that is not actively being used to cheaper, though possibly slower, storage.
  - **Archive phase controls**: Data controls in the archive phase are similar to the store phase, since archiving is just a special type of storage. Due to long timeframes for storage in archives, there may be additional concerns related to encryption. Specifically, encryption keys may be rotated on a regular basis. When keys are rotated, data encrypted with them is no longer readable unless you can access that older key; this leads to solutions such as key escrow or the cumbersome process of decrypting old data and re-encrypting it with a new key. Additionally, storage formats may become obsolete, meaning archived data is not readable, or archival media may degrade over time, leading to losses of integrity or availability.
- **Destroy**: Data that is no longer useful and that is no longer subject to retention requirements should be securely destroyed. Methods of destruction range from low security such as simply deleting filesystem pointers to data and reusing disks to more secure options such as overwriting disks, physical destruction of storage media, or the use of cryptographic erasure or *cryptoshredding*, whereby data is encrypted and the keying material is securely erased. Data destruction can be a particular challenge for cloud security practitioners as the CSP has physical control over media, preventing the use of options like physical destruction. Additionally, cloud services such as Platform as a Service (PaaS) and Software as a Service (SaaS) typically do not provide access to underlying storage for tasks like overwriting.
  - **Destroy phase controls**: Choosing the proper destruction method should balance two main concerns: the value of the data being destroyed and the options available in a cloud service environment. Low-sensitivity data such as

public information does not warrant extraordinary destruction methods, while high-value information like personally identifiable information (PII) does. Various resources provide guidance for selecting a destruction method, such as the NIST SP 800-88, *Guidelines for Media Sanitization*, available here: csrc.nist.gov/publications/detail/sp/800-88/rev-1/final.

Data Dispersion

Data dispersion refers to a technique used in cloud computing environments of breaking data into smaller chunks and storing them across different physical storage devices. It is similar to the concept of striping in redundant arrays of independent disks (RAIDs), where data is broken into smaller segments and written across multiple disks, but cloud-based data dispersion also implements *erasure coding* to allow for reconstruction of data if some segments are lost.

Erasure coding is similar to the idea of a parity bit calculation in RAID storage. In simple terms, data being written is broken into multiple segments, a mathematical calculation is conducted on the segments, and the result is stored along with the data. In the event that some segments are lost, the parity bit and the remaining segments can be used to reconstruct the lost data, similar to solving for a variable in an algebra equation.

Data dispersion in cloud environments can have both positive and negative impacts on an organization's security. The benefit of availability is obvious—even if a physical storage device fails, the data on it can be reconstructed, and if a physical device is compromised, it does not contain a complete, usable copy of data. However, dispersing the segments can cause issues. If data is dispersed to countries with different legal or regulatory frameworks, the organization may find itself subject to unexpected laws or other requirements. Most CSPs have implemented geographic restriction capabilities in their services to allow consumers the benefit of dispersion without undue legal/regulatory complexity.

Properly configuring the cloud service is a crucial task for the Certified Cloud Security Professional (CCSP) to meet the organization's compliance objectives. If the organization is subject to European Union (EU) General Data Protection Regulation (GDPR) requirements, it may be preferable to maintain data in the EU rather than dispersing it to all countries the CSP operates in. This may be a simple configuration for a particular service, or may require complex information technology (IT) project planning and system architecture.

One potential downside of data dispersion is latency, due to the additional processing overhead required to perform the erasure coding and reconstruct data. This is similar to the performance considerations in RAID setups that implemented a parity bit. The additional time and processing capacity may introduce system latency, which can have negative consequences on system availability. This is especially true for high-volume, transaction-based systems or for systems with data that is highly dynamic like fileshares.

DESIGN AND IMPLEMENT CLOUD DATA STORAGE ARCHITECTURES

Pooling resources is one of the key elements of cloud services. Virtualized pools of storage are much more flexible than installing and maintaining individual disks or storage networks, but it can be difficult to identify exactly where and how data is being stored in these broad pools. Understanding the storage options available in different cloud service models is essential. The CCSP should be aware of general storage types, threats, and countermeasures associated with each, as well as the specific offerings of their chosen CSP. Traditional decisions such as directly choosing SSDs for high-speed storage needs may be replaced instead by a set of configurable parameters including the quantity and speed of data access.

Storage Types

Each of the three cloud service models offers unique storage types designed to support the needs and use cases of the particular service model. These options are detailed next. It is important for the CCSP to note that CSPs may offer storage of a particular type but with unique branding or naming conventions. There are also novel storage solutions used for specific industries that are outside the scope of the CCSP exam.

IaaS

Infrastructure as a Service (IaaS) typically offers consumers the most flexibility and also requires the most configuration. Types of storage available in IaaS include the following:

- **Ephemeral**: Unlike the other storage types discussed in the following sections, ephemeral storage is not designed to provide extended storage of data. Similar to random access memory (RAM) and other volatile memory architectures, ephemeral storage lasts as long as a particular IaaS instance is running, and the data stored in it is lost when the virtual machine (VM) is powered down. Ephemeral storage is often packaged as part of compute capability rather than storage, as modern operating systems (OSs) require temporary storage locations for system files and memory swap files.
- **Raw**: Raw device mapping (RDM) is a form of virtualization that allows a particular cloud VM to access a storage logical unit number (LUN). The LUN is a dedicated portion of the overall storage capacity for use by a single VM, and RDM provides the method for a VM to access its assigned LUN.
- **Long-term**: As the name implies, long-term storage is durable, persistent storage media that is often designed to meet an organization's records retention or data archiving needs. The storage may offer features such as search and data discovery as well as unalterable or immutable storage for preserving data integrity.

- **Volume**: Volume storage behaves like a traditional drive attached to a computer, but in cloud data storage, both the computer and drive are virtualized. Volume storage may store data in blocks of a predetermined size, which can be used for implementing data dispersion. Because the disk is virtualized, the data may actually be stored across multiple physical disks in the form of blocks along with the erasure coding needed to reconstruct the data if some blocks are missing.
- **Object**: Object storage is similar to accessing a Unix sharepoint or Windows file server on a network. Data is stored and retrieved as objects, often in the form of files, and users are able to interact with the data objects using file browsers.

PaaS

Some PaaS offerings provide the ability to connect to IaaS storage types, such as connecting a volume to a PaaS VM to provide a virtual disk for storing and accessing data. There are storage types unique to PaaS, however, including the following:

- **Disk**: This is a virtual disk that can be attached to a PaaS instance and may take the form of a volume or object store, depending on the CSP offering and consumer needs. Many PaaS offerings simply offer parameters for storage connected to the PaaS instance, such as speed and volume of I/O operations or data durability, and the CSP provisions appropriate storage space based on the configurations specified by the consumer.
- **Databases**: This is both a storage type and PaaS offering. Platforms that can be delivered as a service include popular database software such as Microsoft SQL Server and Oracle databases, as well as CSP-specific offerings such as AWS Relational Database (RDS) or Microsoft Azure Databases. In most cases, these databases will be offered in a multitenant model with logical separation between clients, and data is accessed via API calls to the database.
- **Binary Large Object (blob)**: Blobs are unstructured data; that is to say, data that does not adhere to a particular data model like the columns in a database. These are often text files, images, or other binary files generated by applications that allow users to generate free-form content; it is possible to apply some loose organization. This is similar to manually organizing files into folders such as word processing files by date of writing or photos by vacation destination. Blob storage services such as AWS Simple Storage Service (S3) and Azure Blob Storage apply these concepts to large volumes of blob data and typically make it available to applications or users via a URL.

Some types of storage platforms or storage types may be specific to a particular CSP's offerings. Examples include blob storage for unstructured data in Microsoft Azure, and a variety of queue services available in AWS that support the short-term storage, queueing, and delivery of messages to users or services.

SaaS

SaaS offerings are the most abstracted service model, with CSPs retaining virtually all control including data storage architecture. In some cases, the data storage type is designed to support a web-based application that permits users to store and retrieve data, while other storage types are actual SaaS offerings themselves, such as the following:

- **Information storage and management**: This storage type allows users to enter data and manipulate it via a web GUI. The data is stored in a database managed by the CSP and often exists in a multitenant environment, with all details abstracted from the users.
- **Content and file storage**: Data is stored in the SaaS app in the form of files that users can create and manipulate. Examples include filesharing and collaboration apps, as well as custom apps that allow users to upload or attach documents such as ticketing systems.
- **Content delivery network (CDN)**: CDNs provide geographically dispersed object storage, which allows an organization to store content as close as possible to users. This offers advantages of reducing bandwidth usage and usually delivers lower latency for end users as they can pull from a server physically closer to their location.

One feature of most cloud data storage types is their accessibility via application program interfaces (APIs). The virtualization technologies in use for cloud services create virtual connections to the pooled storage resources, replacing physical cable connections. These APIs mean the storage types may be accessible across from more than one service model; for example, object storage may be accessed from a PaaS or SaaS environment if an appropriate API call is made. Many CSPs have specifically architected their storage APIs for broad use in this manner, like Amazon's Simple Storage Service (S3), which is an object storage type accessible via a REST API. This enables IaaS, PaaS, SaaS, on-premises systems, and even users with a web browser to access it.

Threats to Storage Types

There are universal threats to data at rest (in storage) regardless of the location, including on-premises or legacy environments, local workstation storage, and cloud services. These affect all three elements of the confidentiality, integrity, availability (CIA) triad: unauthorized access is a threat against confidentiality, improper modification represents a threat to integrity, and loss of a storage device or loss of connectivity is a threat against availability. Tools that are appropriate for on-premises environments may not work in a distributed cloud environment, so the CCSP should be aware of how these threats impact cloud storage and appropriate countermeasures.

- **Unauthorized access**: Any user accessing data storage without proper authorization presents obvious security concerns. Appropriate access controls are required by the consumer to ensure only properly identified and authorized internal users are able to access data stored in cloud services. Because of the multitenant nature of cloud storage, the CSP must provide adequate logical separation to ensure cloud consumers are not able to access or tamper with data that does not belong to them.
- **Unauthorized provisioning**: This is primarily a cost and operational concern. The ease of provisioning cloud services is one of the selling points versus traditional, on-premises infrastructure. This ease of use can lead to unofficial or shadow IT, which drives unrestricted growth in the cloud services and associated costs. Unauthorized storage can also act as a blind spot when it comes to security; if the security team is not aware of where the organization is storing data, the team cannot take appropriate steps to secure that data.
- **Regulatory noncompliance**: Certain cloud service offerings may not meet all the organization's compliance requirements, which leads to two security concerns. First are the consequences of noncompliance like fines or suspension of business operations. Second is the foundation of the compliance requirements in the first place—to protect data. Requirements like the use of a specific encryption algorithm are usually driven by a need to protect data; cloud services, which do not meet the compliance objectives, are also unlikely to offer adequate security for the regulated data being stored.
- **Jurisdictional issues**: The major CSPs are global entities and offer highly available services, many of which rely on global redundancy and failover capabilities. Unfortunately, the ability to transfer data between countries can run afoul of legal requirements, particularly privacy legislation that bars the transfer of data to countries without adequate privacy protections. The features in a particular cloud storage service may support global replication by default, so it is incumbent on the CCSP to understand both their organization's legal requirements and the configuration options available in the CSP environment.
- **Denial of service**: Cloud services are broadly network accessible, which means they require active network connectivity to be reachable. In the event a network connection is severed anywhere between the user and the CSP, the data in storage is rendered unavailable. Targeted attacks like a distributed denial of service (DDoS) can also pose an issue, though the top CSPs have robust mitigations in place and are usually able to maintain some level of service even during an attack.
- **Data corruption or destruction**: This is not a concern unique to cloud data storage. Issues such as human error in data entry, malicious insiders tampering

with data, hardware and software failures, or natural disasters can render data or storage media unusable.

- **Theft or media loss**: This threat applies more to devices that can be easily accessed and stolen, such as laptops and universal serial bus (USB) drives; however, the risk of theft for cloud data storage assets like hard drives does exist. CSPs retain responsibility for preventing the loss of physical media through appropriate physical security controls. Consumers can mitigate this risk by ensuring adequate encryption is used for all data stored in the cloud, which renders the stolen data useless unless the attacker also has the key.
- **Malware and ransomware**: Any location with data storage and processing abilities is at risk from malware, particularly ransomware. Attackers have become more sophisticated when writing ransomware, so it not only encrypts data stored on locally attached drives but also seeks common cloud storage locations like well-known collaboration SaaS apps. Proper access controls and anti-malware tools can prevent or detect malware activities.
- **Improper disposal**: Like physical drive loss, the CSP has the majority of responsibility when it comes to disposal of hardware. Ensuring hardware that has reached the end of its life is properly disposed of in such a way that data cannot be recovered must be part of the CSP's services. Consumers can protect data by ensuring it is encrypted before being stored in the cloud service and that the encryption keys are securely stored away from the data.

DESIGN AND APPLY DATA SECURITY TECHNOLOGIES AND STRATEGIES

Data security in the cloud comprises a variety of tools and techniques. According to the shared responsibility model published by the major CSPs, consumers are responsible for securing their own data. Under most privacy legislation, the data owner, who is usually the cloud consumer, is ultimately accountable and legally liable for data breaches. However, adequately securing data requires actions by both the cloud consumer and the CSP to properly secure elements such as the hardware infrastructure and physical facilities.

Encryption and Key Management

Encryption is the process of applying mathematical transformations to data to render it unreadable. It typically requires the use of a key or cryptovariable, which is a string of data used by the cryptographic system to transform the data. The steps taken to achieve the transformation are known as an *algorithm*.

Modern cryptographic algorithms like Rijndael, which is part of the Advanced Encryption Standard (AES), offer protection for data that could take thousands or millions of years to break. Trying every possible combination of keys and permutation steps in the algorithm would take more time and resources than most attackers have available, but the process of encrypting and decrypting is relatively short if you know the key. Encryption is a foundational element of modern data security, particularly in cloud environments when data is stored outside of the organization's direct control.

Due to the criticality of encryption, organizations should focus attention on properly implementing and managing cryptographic systems. One particular area of focus is managing encryption keys. Auguste Kerckhoffs, a Dutch cryptographer, defined a simple doctrine that underpins key management: a cryptosystem should be secure even if everything about the system, except the key, is public knowledge. Known as *Kerckhoffs's principle*, this simple maxim guides security in cryptographic systems by placing the emphasis on protecting keys.

Keys should be classified at the highest data classification level available in the organization and protected as other high value assets would be. It is appropriate to implement controls across all categories including policies for creating and managing keys, operational procedures for securing and using keys, and tools and systems for handling keys. As a data asset, the keys should be protected at each stage of their lifecycle, including the following:

- Creating strong, random keys using cryptographically sound inputs like random numbers
- Storing keys in a secure manner, whether encrypted inside a key vault or stored on a physical device, and handling the process of storing copies for retrieval if a key is ever lost (known as key escrow)
- Using keys securely, primarily focused on access controls and accountability
- Sharing keys is not as common due to their highly sensitive nature, but facilities should exist for sharing public keys, securely transferring symmetric keys to a communications partner, and distributing keys to the key escrow agent
- Archiving keys that are no longer needed for routine use but might be needed for previously encrypted data
- Secure destruction of keys that are no longer needed or that have been compromised

FIPS 140-3 provides a scheme for U.S. government agencies to rely on validated cryptographic modules and systems, though it has become a globally recognized framework as many tools offer Federal Information Processing Standards (FIPS) validated modes for encryption. As of 2020, this standard is being phased in to replace its predecessor, FIPS

140-2. FIPS 140-3 establishes a framework and testing scheme for validating the strength of protection provided by a cryptographic module and defines levels of protection for such modules including physical tamper-evident hardware security modules (HSMs). Details on the standard can be found here: csrc.nist.gov/publications/detail/fips/140/3/final.

Cloud security practitioners will need to understand where encryption can be deployed to protect their organization's data and systems. Many CSPs offer virtualized HSMs that are validated against the FIPS 140-2 standard and can be used to securely generate, store, and control access to cryptographic keys. These virtual HSMs are designed to be accessible only by the consumer and never by the CSP and are usually easy to integrate into other cloud offerings from the same CSP.

If your organization uses multiple cloud providers or needs to retain physical control over key generation, your apps should be architected to allow for a bring-your-own-key strategy. This is more technically challenging for the organization, as hosting any on-prem systems requires more skills and resources, but it offers more control over the configuration and use of encryption, as well as physical control over the HSMs.

Encryption in cloud services may be implemented at a variety of layers, from the user-facing application all the way down to the physical storage devices. The goals of safeguarding data, such as counteracting threats of physical theft or access to data by other tenants of the cloud service, will drive decisions about which types of encryption are appropriate to an organization. Some examples include the following:

- **Storage-level encryption** provides encryption of data as it is written to storage, utilizing keys that are controlled by the CSP. It is useful in cases of physical theft as the data should be unreadable to an attacker, but CSP personnel may still be able to view data as they control the keys.
- **Volume-level encryption** provides encryption of data written to volumes connected to specific VM instances, utilizing keys controlled by the consumer. It can provide protection in the case of theft and prevents CSP personnel or other tenants from reading data, but it is still vulnerable if an attacker gains access to the instance.
- **Object-level encryption** can be done on all objects as they are written into storage, in which case the CSP likely controls the key and could potentially access the data. For high-value data, it is recommended that all objects be encrypted by the consumer with keys they control before being stored.
- **File-level encryption** is often implemented in client applications such as word processing or collaboration apps like Microsoft Word and Adobe Acrobat. These apps allow for encryption and decryption of files when they are accessed using keys controlled by the user, which prevents the data from being read by CSP personnel or other cloud tenants. The keys required may be manually managed, such as a password the user must enter, or automated through IRM, which can verify

a user's authorization to access a particular file and decrypt it based on the user's provided credentials.

- **Application-level encryption** is implemented in an application typically using object storage. Data that is entered or created by a user is encrypted by the app prior to being stored. Many SaaS platforms offer a bring-your-own-key ability, which allows the organization to prevent CSP personnel or other cloud tenants from being able to access data stored in the cloud.
- **Database-level encryption** may be performed at a file level by encrypting database files or may utilize transparent encryption, which is a feature provided by the database management system (DBMS) to encrypt specific columns, whole tables, or the entire database. The keys utilized are usually under the control of the consumer even in a PaaS environment, preventing CSP personnel or other tenants from accessing data, and the encrypted data is also secure against physical theft unless the attacker also gains access to the database instance to retrieve the keys.

Hashing

Hashing, sometimes known as *one-way encryption*, is a tool primarily associated with the integrity principle of the CIA triad. Integrity deals with preventing, detecting, and correcting unintended or unauthorized changes to data, both malicious and accidental. Cryptographic algorithms called *hash functions* take data of any input length and perform mathematical operations to create a unique hash value. This process can be performed again in the future and the two hash values compared; if the input data has changed, the mismatched hash values are proof that the data has been altered.

Hashes form an integral part of digital signatures, which provide users the ability to verify both the integrity and the source of a message, file, or other data such as an app. The signing party calculates a hash and encrypts the hash value with their private key. A receiving party, who may be a user receiving a message, downloading an app, or pulling software from a repository, calculates a hash of the received data and then decrypts the sender's digital signature using the sender's public key. If the two hashes match, it can be assumed the data is original, as no other user would be able to change the data, calculate a hash, and use the original sender's private key to create the digital signature. A CCSP should be aware of digital signatures as a method for verifying messages and apps used in cloud environments, especially when third-party software is being integrated.

Hashes can provide multiple security services in cloud environments. They can verify copies of data like backups are accurate and can be used to verify the integrity of messages like email. They are also widely implemented in many security tools as a way to detect changes, which can indicate system compromise. File integrity monitoring is used by some anti-malware and intrusion detection systems to identify changes to key system files, and highly secure systems may create hashes of data about system hardware

such as manufacturer, devices connected, or model numbers. In both cases, there is an expectation that these items should not change; comparing a current hash with a previously calculated hash can identify unwanted changes.

When implementing hash functions, it is important to choose a strong function that is collision-resistant. A *collision* occurs when two different inputs produce the same hash value as an output. In this case, it is impossible to rely on the hash function to prove integrity. As with many aspects of information security, there is a U.S. federal government standard related to hashes. FIPS 180-4, *Secure Hash Standard (SHS)*, provides guidance on the Secure Hash Algorithm (SHA-3). As with FIPS 140-2 and 140-3 encryption, many popular tools and platforms provide FIPS-compliant modes for hash algorithms. More details on SHS can be found here: csrc.nist.gov/publications/detail/fips/180/4/final.

Masking

Masking is similar to obfuscation, which is discussed later in the chapter, and both are used to prevent disclosure of sensitive data. Data masking involves hiding specific elements of data for certain use cases, primarily when there is a need for data to be retrievable for some but not all users or processes. As an example, a corporate human resources (HR) system may need to store a user's Social Security number (SSN) for payment and tax purposes. Daily users accessing the HR system do not have a need to see the full SSN, so the system instead displays XXX-XX-1234, as the last four digits are needed to verify a user's identity.

Data masking can be useful in preventing unintended disclosures by limiting the amount of data displayed. It is very granular implementation of minimum necessary access. Although a user may be authorized to view full SSN information, in the daily use case of managing employee records, they do not have a need to see the full information.

Unstructured data can present problems for masking, as well as tokenization, obfuscation, and de-identification. When data is structured in a database, it is easy to identify and apply these techniques. Unstructured data can be stored in files, free-form text or comment fields in databases, or in applications that store data objects without structure. As an example, it would be quite simple to identify and apply masking to a database column labeled Social Security Number, but if some records have that data in a comments field along with other data, those records will obviously not be masked. Data handling and system use policies should dictate the proper use of information, including where to store sensitive data elements. If the organization utilizes unstructured storage or formats, data security tools must also be chosen to deal with unstructured data types.

Tokenization

Tokenization is a process whereby a nonsensitive representation of sensitive data, otherwise known as a *token*, is created and used. The token is a substitute to be used in place of more sensitive data like a credit card number, often called a *primary account number*

(PAN). Rather than storing and using PANs, which is risky due to the value of a PAN, tokens can be used instead.

Tokens can be traced back to the original information by making a proper request to the tokenization service, which usually implements access controls to verify user identities and authorization to view sensitive data. Tokens are implemented for many online payment systems where credit card numbers are stored; they are not really stored in the app but are instead tokenized. When the user makes a purchase, the app supplies the token along with user identity information to the tokenization server, which if it accepts the information provided, accesses the relevant credit card data and supplies it to complete the transaction.

Using tokens instead of the actual data reduces risk by removing sensitive data needed by the application. A database of live credit card PANs would be incredibly valuable to a thief, who could use those cards to make purchases. The same database full of tokens is virtually worthless. Tokenization systems are obviously high-value targets, but due to their specialized function, it is possible to more robustly secure them versus a general-purpose application.

Although tokenization is widely used in credit card processing transactions and is a recommended control for payment card industry data security standard (PCI DSS) compliance, any sensitive data can be tokenized to reduce risk. Implementations vary, but the process of tokenizing data generally follows the same basic steps:

1. An application collects sensitive information when a user enters it.
2. The app secures, often using encryption, and sends the sensitive data to a tokenization service.
3. Sensitive data is stored in the token database, and a token representing the data is generated and stored in the token database along with the sensitive data.
4. The token is returned to the original application, which stores it instead of the original sensitive data.
5. Any time the sensitive data is required, the token and appropriate credentials can be used to access it. Otherwise, the sensitive data is never revealed, as the tokenization service should be tightly access controlled.

Data Loss Prevention

DLP, sometimes also known as *data leakage prevention*, refers to a technology system designed to identify, inventory, and control the use of data that an organization deems sensitive. It spans several categories of controls including detective (identifying where sensitive data is stored and being used), preventative (enforcing policy requirements on the storage and sharing of sensitive data), and corrective (displaying an alert to the user informing them of the policy violation and preventing inappropriate action such as sending sensitive data via email).

Due to the multiuse nature of DLP, many organizations will implement only some of the functions at one time. An organization is unlikely to be successful attempting to

simultaneously perform an organization-wide data inventory, deploy new technology in the form of DLP agents and network devices, and manage process changes due to the new DLP functionality. In cloud security environments, particularly when the enterprise architecture combines on-premises and cloud services, DLP can be useful for enforcing policies on the correct use of various systems, such as storing regulated data only in approved on-premises repositories rather than cloud storage.

A typical DLP installation will comprise three major components:

- **Discovery** is a function of DLP that allows the organization to identify, categorize, and inventory data assets. In large organizations, this may be the first step of a phased rollout, while in smaller organizations with fewer assets, it may not be required. DLP scanners identify data storage locations belonging to the organization, typically by performing network scans to identify targets such as fileshares, storage area networks (SANs), databases, common collaboration platforms like SharePoint, and cloud services like Google Drive or Dropbox. The scan will likely require as input some details of the organization, such as IP ranges or domains, over which it will perform the scans.
  - Once the tool has a created a blueprint of the organization's network and likely storage sources, it will scan the identified targets to identify data based on common formats such as xxx-xx-xxxx, which represents a U.S. Social Security number. Organization-defined sensitive data can also be identified, such as documents that contain strings like "Confidential – for internal use only" in document footers or utilizing regular expressions to identify sensitive data the organization generates. Highly privileged credentials will likely be required, so managing access controls for the DLP scanner is a major focus. Most scanners offer the ability to categorize information based on standard categories such as PII, protected health information (PHI), payment information, etc., and some support organization-defined classification levels. The result of the scan is an asset inventory of organization data sets.
- **Monitoring** is the most important function of a DLP system, which enables the security team to identify how data is being used and prevent inappropriate use. The choice of DLP tool should be made in light of its capability to monitor the platforms in use at an organization; for example, some legacy tools do not provide monitoring for emerging instant message tools like Slack. Another critical concern for monitoring is the placement of the DLP's monitoring capabilities. A network-based DLP monitoring traffic on an organization LAN is unable to monitor the actions of remote workers who are not always connected to the network and may not provide sufficient coverage to mitigate risk.
  - **In-motion** data monitoring is typically performed by a network-based DLP solution and is often deployed on a gateway device such as a proxy, firewall, or email server. Some DLP agents deployed on user workstations can perform

in-motion monitoring as data leaves the particular machine, such as scanning the contents and attachment of an email before it is sent to identify a policy violation. This type of DLP must be placed in appropriate locations to be able to monitor unencrypted data; otherwise, users could create an encrypted tunnel, and the DLP will not be able to scan the data being sent. Workstation agent- or proxy-based tools can prevent this issue by scanning data before it is sent over an encrypted connection.

- **At-rest** monitoring is performed on data in storage and is usually performed by an agent deployed on the storage device, though some network-based DLP can perform scans of storage locations with proper credentials. These can spot policy violations such as sensitive information stored outside of prescribed columns, for example, users entering credit card or PII in unencrypted notes/comments fields rather than fields where encryption, tokenization, or other controls have been applied. Compatibility is a particular concern for agent-based DLP solutions, as the organization's storage solutions must be supported for the DLP to be effective.
- **In-use** monitoring is often referred to as endpoint or agent-based DLP, and it relies on software agents deployed on specific network endpoints. These are particularly useful for monitoring users interacting with data on their workstations or other devices and enforcing policy requirements on the use of those endpoints. Compatibility is a major concern for these agents as the DLP must support the devices and OSs in use. Most DLP solutions offer support for popular business operating systems, but support for some platforms such as the macOS and mobile operating systems like iOS and Android may be limited.

- **Enforcement:** DLP applies rules based on the results of monitoring to enforce security policies. For example, a DLP agent running on user's workstation can generate an alert or block the user from saving sensitive information to a removable USB drive, or from attaching the information to an email. A network-based DLP can look for information by analyzing traffic entering or leaving a network (like a virtual private cloud), or monitor information sent to and from a specific host. If the DLP detects sensitive data that is not being handled appropriately, such as unencrypted credit card information, it can either block the traffic or generate an alert. Alerts and enforcement actions taken by the DLP should be monitored and investigated as appropriate, and are a valuable source of security incident detection.

Deploying DLP in a cloud environment can be a particular challenge, especially in SaaS or PaaS service models where the organization lacks the ability to install software or that do not permit scanning such as DLP discovery. There are many cloud-native DLP tools, and the CCSP must ensure the organization's system requirements are elicited clearly, particularly which operating systems and cloud environments must be supported. DLP can create operational overhead due to the time and resources needed to scan network traffic

or resources consumed on endpoints when monitoring. This impact should be considered as part of the cost-benefit analysis associated with the DLP solution deployment.

Data Obfuscation

Obfuscation is similar to data masking but is more often implemented when sensitive data needs to be used in a different situation. For example, obfuscation can remove or replace sensitive data elements when data from a live production system is copied for testing purposes. Testers are likely not authorized to view the data and can perform their jobs using synthetic data, which is similar to live data. Regulations often require the use of obfuscation or de-identification of data prior to its use for purposes outside of normal operations.

There are a number of ways to perform obfuscation on data, outlined next. These methods are often implemented as part of database export functions or in data management programs like spreadsheets, and they may also be implemented in business applications when users need to access or use a data set without requiring access to all sensitive data elements.

- Substitution works by swapping out some information for other data. This may be done randomly, or it may follow integrity rules if the data requires it. As an example, when substituting names in a data set, the algorithm may have a set of male and female names to choose from. If the data is to be analyzed, this gender information could be important, so gender-appropriate names will be required when substituting.
- Shuffling involves moving data around. This can be done on an individual column; for example, a name like Chris would become Hrsci, though this is fairly easy to reverse engineer. More robust shuffling can be performed by shuffling individual data points between rows; for example, swapping Mary Jones's and Bob Smith's purchase history information. Shuffled data still looks highly realistic, which is advantageous for testing but removes identifiability.
- Value variance applies mathematical changes to primarily numerical data like dates, accounting or finance information, and other measurements. An algorithm applies a variance to each value, such as +/– $1,000. This can be useful for creating realistic-looking test data.
- Deletion or nullification simply replaces the original data with null values. A similar practice is redaction, where a document's sensitive contents are simply blacked out. The resulting document may not be useful if too much of the data has been redacted; nullified data may be problematic for testing as zero values are often unrealistic.
- Encryption may be used as a tool for obfuscation, but it is problematic. Encrypted data is not useful for research or testing purposes, since it cannot be read. This challenge has given rise to the field of *homomorphic encryption*, which is the ability to process encrypted data without first decrypting it. Once returned to the

original data set, the processed data can be decrypted and reintegrated, however, homomorphic encryption is largely experimental and therefore not widely used.

Obfuscating and de-identifying data often incorporates rules to ensure the output data remains realistic. For example, credit card numbers conform to known patterns—usually 16 digits broken into groups of four—so replacing real credit card information must use only numerals. Some other information may also need to be verifiable against external sources. For example, postal codes and telephone area codes indicate geographic locations, so modifying a data set with address, post code, and telephone numbers may need to conform to the rules of these numbering systems or the relationships that data implies. If your business does not serve customers outside the continental United States, then postal codes for Canada, Alaska, and Hawaii should not be allowed when performing substitutions.

A process known as *pseudo-anonymization* or *pseudonymization* (often *pseudonymisation* due to its presence in the EU GDPR and use of British English spellings by European translators) is a process of obfuscating data with the specific goal of reversing the obfuscation later. This is often done to minimize risk of a data breach and is performed by the data owner or controller prior to sending data to a processor. For example, if an organization wants to store large volumes of customer purchase orders in a cloud service, where storage is cheaper and more durable, they could remove PII and replace it with an index value prior to upload. When the data is retrieved, the index value can be looked up against an internal database, and the proper PII inserted into the records. The cost of storing that PII index would be small due to the lower volume of data, and the risk of a cloud data breach is also minimized by avoiding storage of PII in the cloud.

Data De-identification

De-identifying data is primarily used when the data contains PII, known as *direct identifiers*, or contains information that could be combined with other data to uniquely identify an individual, known as *indirect identifiers*. Direct identifiers consist of information such as name and financial account numbers, while indirect identifiers are often demographic information such as age or personal behavior details such as shopping history. Removing these identifiers makes data anonymous rather than identifiable; hence, this process is often known as *anonymization*. This differs from pseudonymization, where the goal is to allow the re-identification of data, unlike anonymization, which is designed to be permanent.

Most privacy regulations require data anonymization or de-identification for any PII use outside of live production environments. For example, U.S. healthcare entities regulated by the Health Insurance Portability and Accountability Act (HIPAA) are required to de-identify medical records information when it is to be used for anything other than patient treatment, such as research studies. Details that could be used to specifically identify a patient must be removed or substituted, such as full name, geographic location

information, payment information, dates more specific than the year, email address, health plan information, etc.

Removing indirect identifiers can be more of a challenge, starting with identifying what information could be uniquely identifiable. Trend information such as frequently browsed product categories at an online shopping site can be cross-referenced to a user's social media posts to identify them with relative certainty. If an online retailer removed all direct identifiers and published their sales trend data, it might still be possible to uniquely identify users. Combining multiple obfuscation and anonymization techniques can be useful to combat this threat, such as deleting names, substituting postal codes, and shuffling rows so that purchase history and location are no longer linked.

IMPLEMENT DATA DISCOVERY

Data discovery has two primary meanings in the context of information security. Discovery of data stored in your environment is the purview of DLP solutions, which helps build an inventory of critical data assets your organization needs to protect. This is not the same as eDiscovery, which deals with collecting evidence in legal situations, but utilizes many of the same principles.

Discovering trends and valuable intelligence within data is the second meaning, and one that is less a dedicated concern of the security department. Analyzing data to uncover trends or make predictions, such as what products are likely to be in-demand or what a given customer might be interested in shopping for, can drive meaningful business improvements like ensuring adequate inventory of goods. In these cases, security is not a primary concern of data discovery, but the business intelligence (BI) is itself a valuable intellectual property (IP) data asset, which requires security commensurate with its value. Supporting elements such as algorithms or proprietary data models used for analysis may also be valuable IP and should be included in the organization's security risk management program.

Many security tools, especially those monitoring large-scale deployments of user workstations, servers, and cloud applications, make use of data discovery. Analysis tools can be used to drive security operations by identifying suspicious events that require investigation, such as system vulnerabilities, misconfigurations, intrusion attempts, or suspicious network behavior. This data, comprising details of an organization's vulnerabilities, is obviously an asset worth protecting as it would be useful to an attacker. It is important for a security practitioner to understand both the value of these tools in supporting security operations as well as how to adequately protect them.

There are a number of important terms associated with data discovery and BI, including the following:

- **Data lake and data warehouse**: These terms are similar but not the same. Both are designed to consolidate large amounts of data, often from disparate sources

inside or outside a company, with the goal of supporting BI and analysis efforts. A lake is an unstructured data storage mechanism with data often stored in files or blobs, while a warehouse is structured storage in which data has been normalized to fit a defined data model.

  - *Normalization* is the process of taking data with different formats—for example one system that stores MM-DD-YYYY and another that uses YYYY-MM-DD—and converting it to a common format. This is often known as *extract*, *transform*, *load* (ETL), as the data is extracted from sources like databases or apps, transformed to meet the warehouse's data model, and loaded into warehouse storage. Normalizing data improves searchability.

- **Data mart**: A data mart contains data that has been warehoused, analyzed, and made available for specific use such as by a particular business unit. Data marts typically support a specific business function by proactively gathering data needed and performing analysis and are often used to drive reporting and decision-making.
- **Data mining**: Mining data involves discovering, analyzing, and extracting patterns in data. These patterns are valuable in some way, much like minerals mined from the ground; they may support business decision-making or enhance human abilities to identify important trends, patterns, and knowledge from large sets of data. Even small organizations with only a few users and systems generate enormous volumes of security log data, and data mining tools can be useful for isolating suspicious events from normal traffic.
- **Online analytic processing (OLAP)**: As the name implies, OLAP provides users with analytic processing capabilities for a data source. OLAP consists of consolidation, drill-down, and slice-and-dice functions. Consolidation gathers multidimensional data sets into cubes, such as sales by region, time, and salesperson. Drill-down and slice-and-dice allow users to analyze subsets of the data cube, such as all sales by quarter across all regions or sales of a particular product across all salespeople. Security incidents that require forensic analysis often make use of OLAP to extract relevant information from log files.
- **ML/AI training data**: Machine learning (ML) and artificial intelligence (AI) are emerging areas of data science. ML is concerned with improving computer algorithms by experience, such as asking a computer to identify photos of dogs and then having a human verify which photos actually contain dogs and which contain other animals instead. The computer algorithm learns and refines future searches; these algorithms can be used across a wide variety of applications such as filtering out unwanted emails by observing which messages users mark as spam, and they are widely implemented in security tools designed to learn and adapt

to an organization's unique environment. AI is a field of computer science with the goal of designing computer systems capable of displaying intelligent thought or problem solving, though the term is often used to describe systems that simply mimic human tasks like playing strategy-based board games or operating autonomous vehicles. Both AI and ML require large sets of data to learn, and there are myriad security as well as privacy concerns, especially when the data sets include personal information like photos used for training ML facial recognition.

As a consumer of cloud services, a CCSP should be aware that their organization retains accountability for protecting data, including data discovery processes to identify and inventory sensitive data. CSPs may be subject to contractual obligations for implementing specific data protections, but they are not the data owners and are therefore not legally liable under most privacy and security laws for data breaches. An adequate inventory of sensitive data is a vital input to security risk assessment and mitigation, so it is essential that a CCSP recognize the capabilities of data discovery and utilize them to support their organization's security risk management process.

Structured Data

Structured data refers to data that has been formatted in a consistent way. This often takes the form of a database where all records conform to a known structure: data is separated into columns, and each row contains the same type of information in the same place. Standalone data may also be structured using a markup language such as XML or JSON, which utilizes tags to provide context around data like <AcctNumber>123456</AcctNumber>.

Data discovery is simplified with structured data, as the process only needs to understand the data's context and attributes to identify where sensitive data exists, such as PII, healthcare data, transaction information, etc. Columns and data attributes are typically named in a self-explanatory way, simplifying the identification of sensitive or useful data. Many security information and event management (SIEM) tools also provide functionality to ingest data from multiple sources and apply structure to the data, which facilitates the process of analysis across disparate sources. As an example, the SIEM tool might normalize log data from multiple operating systems to include human-readable usernames in log files, rather than a numeric global user ID value.

Structured data is often accompanied by a description of its format known as a *data model* or *schema*, which is an abstract view of the data's format in a system. Data structured as elements, rows, or tuples (particularly in relational databases) is given context by the model or schema. For example, defining a particular string as a user ID can be achieved using tags defined in a schema. Understanding the relationship of a user belonging to a particular business unit can be achieved with data in a particular column; for example, the user's business unit designation appears in the "Bus. Unit" column.

These relationships and context can be used to conduct complex analysis, such as querying to see all failed login attempts for users from a specific business unit.

Metadata, or data that describes data, is a critical part of discovery in structured data. *Semantics*, or the meaning of data, is described in the schema or data model and can be useful when analyzing the relationships expressed in data. A particular user record, for example, may contain a tag <BusUnit>EMEA</BusUnit>, which identifies that the user belongs to the EMEA business unit and might be considered sensitive information as it provides some location information for that user. Similarly, column names can also be used to identify specific types of regulated data, such as credit card numbers, which require specific protections.

Unstructured Data

Structured data simplifies the process of analysis by providing context and semantics, which speed up discovery and analysis. Unfortunately, not all data is structured—human beings tend to create data in a variety of formats like documents or photos containing infinite types and configurations of data. Unstructured data refers to information stored without following a common format. For example, credit card data may be stored in tables or as strings inside a word processing document or in a spreadsheet with user-defined columns like CC, Card Number, PAN, etc. The variety of unstructured data makes it harder to identify and analyze, but it is nonetheless valuable and therefore requires protection.

Applying *data labels* is one approach to dealing with unstructured data. Labels can identify the classification level of a particular file and, by extension, the protections required. Files or other objects stored in a particular system may have a label applied by virtue of being stored in that system; for example, all documents stored in a "Restricted" fileshare are also given a "Restricted" classification. Labels may also be applied individually via metadata in a file management system or inside documents as a header, footer, or watermark. Security tools such as DLP should be able to detect these unstructured files based on the labels and take appropriate actions, such as blocking files with "Restricted" in the footer from being sent as email attachments. Note that this is an imperfect approach, as users can often delete this data or use incorrect templates and thereby mislabel the data.

Another approach to unstructured data discovery is content analysis, which requires a great deal of resources to parse all data in a storage location and identify sensitive information. Analysis can be performed using one of several methods, such as the following:

- **Pattern matching**, which compares data to known formats such as credit card numbers that are 16 numeric digits, or unique organization-defined patterns such as user account information like "j.smith." Patterns are typically defined using a regular expression or *regex*, which allows for more powerful search capabilities

by defining not just exact match conditions, but flexible conditions as well. For example, searching for j.smith@company.com would return only exact matches of that email address. If the user has both j.smith and john.smith aliases, a regex can be created to search for j*.smith@companyname.com, which returns both email aliases.

- **Lexical analysis** attempts to understand meaning and context of data to discover sensitive information that may not conform to a specific pattern. This is useful to flag highly unstructured content like email or instant message communications where users may utilize alternate phrasing like "payment details" instead of "card number." However, it is prone to false positives as linguistic meaning and context are quite complex.
- **Hashing** attempts to identify known data such as system files or important organization documents by calculating a hash of files and comparing it to a known set of sensitive file hashes. This can be useful for documents that do not change frequently.

One particular challenge for data discovery is the inclusion of unstructured data inside structured data sets. This often occurs as an unstructured text field in an otherwise structured database, like a free-form notes or comments field into which users can enter any information. Systems that support both types of data are also problematic, like ticketing systems with form fields and file attachments. In both scenarios, users are required to enter information into defined fields, but they may enter or upload anything in free-form text or file attachments. The result is a system with a wide variety of data at differing classification levels, including more sensitive data than originally planned. The organization's data discovery tool must be flexible enough to identify both types of data within the same system, which increases cost and complexity.

IMPLEMENT DATA CLASSIFICATION

Class has many meanings, but as it relates to data, it is a way to identify common attributes that drive protection requirements. These may include regulated categories such as PII, which is covered by a variety of regulatory and legal schemes, such as privacy legislation, or internal schemes such as Business Confidential information, which offers your organization some competitive advantage and therefore should be highly protected.

Classification is the act of forming classes, or groups, by identifying these common attributes. The term *categorization* is often used, especially when discussing systems or large data sets, and describes the process of determining which class a system or data set belongs to by eliciting requirements for protecting the system's confidentiality, integrity,

and availability. Each classification level should have an associated set of control expectations, such as data classified as "Public" does not require encryption in transit, while data classified as "Internal Use Only" must be encrypted both at rest and in transit. These requirements are mitigations for the risk presented to the organization by the data or system, as described by the operational impact of a loss of confidentiality, integrity, and/or availability.

Data classification is a way for organizations to provide a uniform set of controls and expectations, as well as a method for speeding up security decision-making. Creating a classification scheme, such as Low, Moderate, and High, allows the organization to bundle security control expectations and simplify the process of determining required actions. When a new system is brought online, security practitioners do not need to perform exhaustive research to determine the security requirements they need to meet. The classification scheme provides a clear set of risk-based security controls and expectations designed to speed up the process.

Data classification levels and schemes are driven by multiple aspects of data. They may be prescribed by an outside entity such as a regulator or government agency or may be driven by purely internal risk management requirements. Here are some examples:

- **Data type**: Different types of data are regulated by different rules, such as healthcare, sensitive PII, financial, educational, or legal. Data classification schemes based on data type often point to a set of external requirements that must be met if a system or data set includes the given data type.
- **Legal constraints**: If data on EU citizens is collected by a company based in another country, that company may have to either implement privacy protection similar to the EU's GDPR or be based in a country with privacy laws recognized by the EU as equivalent to GDPR. Understanding the legal constraints attribute allows the organization to make decisions such as geolocation of application architecture.
- **Ownership**: Many organizations utilize data that is shared by business partners, customers, or shared sources, all of which may impose requirements such as not sharing the data with third parties or securely destroying data after a specified retention period.
- **Value/criticality**: Some data's value is incredibly context-specific. A database of contact details for restaurant suppliers is of little value to an IT services company, but that same database would be mission-critical to a company operating a chain of restaurants. The data classification scheme must take into account how valuable and critical data is to the organization, often by measuring the impact that a loss of the data would have on operations.

Because of the information provided in a data classification policy, it is often a foundational document for an organization's security program. Rather than specifying a long list of system-specific security requirements for each system individually, such as approved encryption or data retention schedules, a classification label provides a common vocabulary for communicating security needs in a consistent manner. Other information security policies should specify the appropriate controls for data or systems at various classification levels, like approved cryptographic modules, access control procedures, and data retention periods and destruction requirements.

Mapping

Data mapping comprises a number of activities in the overall practice of data science—the application of scientific methods and algorithms to identify knowledge or useful information from data sets. One particular practice related to data mapping is relevant to the role of a security practitioner: identifying the locations of data.

Identifying and mapping the location of data within the organization is a critical inventory task, which is in turn a critical input to risk assessments. Identifying what needs protecting—system and data assets belonging to the organization—and where they exist are crucial to ensure a security program is designed appropriately. Many DLP tools provide this functionality by scanning a network, domain, or other set of organization-controlled resources to identify data storage locations. This mapping may be further extended by identifying metadata such as asset ownership like a person, role, or organizational unit, which provides the organization with key information on responsibility and accountability for security processes.

Labeling

Once a data set or system has been classified, the classification level must be communicated in some way so that users, administrators, and other stakeholders know how to protect it. Again, the use of labels provides simplification. Rather than forcing users to memorize which systems they can print data from and which systems ban printing, users are instead instructed that systems labeled "Internal" allow printing, while systems labeled "Top Secret" do not allow printing.

Labeling data can be tricky, as we typically think of labels in physical terms but obviously are not able to stick a label on a collection of digital 0s and 1s. There are a variety of labeling methods for different types of assets, such as the following:

- **Hard-copy materials**, primarily printed information on paper, which can be labeled with a printed watermark, stamps, or a physical container such as a folder or box. Hard-copy materials are the easiest to affix labels to because they are physical objects and do not change often.

- **Physical assets**, including servers, workstations, disc drives, optical disks, and removable media, which can be physically labeled with a sticker or badge. These are somewhat tricky to label as the information on these devices can change quite easily. It can also be more challenging to identify and label found physical assets, as the user needs to have appropriate equipment to read the data on the asset; there may also be security issues around plugging in found media due to the possibility of introducing malware.
- **Digital files**, which may come from common collaboration tools, databases, or other programs, and can often be labeled with metadata. This may include content inside the document such as a digital watermark or signature, or even text like a document footer with the classification level printed. File metadata such as the filename or document attributes stored in a database can also be used to apply a classification label.
- **Some complex or shared systems and data sets** will have subcomponents that can be labeled, but the overall system cannot. In these cases, labeling of components along with supporting procedures, such as training and reference materials for users or a master organization-wide list of systems, can be useful to ensure users are aware of the protection requirements they must meet.

When it comes to labeling data sets, there are a number of best practices that facilitate the use of the classification level to ensure adequate protection is applied. The first is to ensure labels are appropriate to all types of media in use within a system; for example, if both digital files and hard-copy documents are to be used, a digital watermark that also appears when the document is printed helps ensure the data label is visible across all media states. Labels should be informative without disclosing too much—stamping a folder "Top Secret" makes it easy to recognize for both legitimate users and bad actors! Labels may include only an owner or asset number that can be used to determine the sensitivity of the data in the event it is lost. Finally, when media is found, it should be classified at the highest level supported by the organization until an examination proves otherwise, ensuring sensitive data is not disclosed.

The organization's DLP tool may be a crucial consumer of data labels, as many DLP tools allow organization-defined labels to be used when performing data identification and classification. If DLP is being used, labels should be applied in a consistent and accessible manner, such as text in the file identifying the classification or common filename conventions to facilitate the discovery process.

Sensitive Data

The organization's data classification policy should cover a number of requirements for handling data, many of which will be driven by external laws and regulations. These

external obligations will typically provide guidance for handling sensitive classes of information such as the following:

- **Personally identifiable information (PII)**: Governed globally by privacy laws and often by laws or regulations specific to certain industries covering the collection, use, and handling of PII. Examples include the EU GDPR and Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) laws, which broadly cover PII, and the U.S. Graham-Leach-Bliley Act (GLBA), which covers banking uses of PII.
- **Protected health information (PHI)**: Defined and governed primarily by the U.S. HIPAA, though personal health records are considered PII by most global privacy laws such as GDPR.
- **Cardholder data (often referred to as a *cardholder data environment*, or CDE)**: Defined and regulated by PCI DSS, it provides guidance on the handling, processing, and limited allowable storage of information related to credit and debit cards and transactions.

Data protection should be specified for all sensitive data discovered, and may be a mix of requirements defined in the various laws mentioned earlier as well as an organization's own risk-management goals. Some elements appropriate to specify in a data classification policy include the following:

- **Compliance requirements inherent at various classification levels**: While this may be too complex for an average user, it ensures vital security requirements are not overlooked. As a best practice, points of contact who are skilled at managing sensitive data should be identified in the policy so users can seek assistance as needed.
- **Data retention and disposal requirements**: Many data protection laws specify retention periods, such as customer records must be held indefinitely while the customer is still active and then for five years thereafter. Classification and retention policies and procedures should be tightly aligned and provide guidance on approved disposal or destruction methods for data that has reached the end of its retention period.
- **What is considered sensitive or regulated data**: Some regulations include exceptions for a variety of circumstances, and an untrained individual may not fully understand the subtle nuances of when healthcare data is considered PHI or not, for example. The classification policy should provide clear guidance on what data is considered sensitive or regulated and explicitly state any exceptions that may apply.

- **Appropriate or approved uses of data**: Many regulations provide explicit guidance on approved use and processing of data, frequently related to the intended purpose or consent given by the data subject. Classification policies must provide guidance on how to identify these approved uses, such as with explicit system instructions or a point of contact who can provide definitive guidance.
- **Access control and authorization**: Controlling logical and physical access to assets is, along with encryption, one of the most powerful tools available to security practitioners. Classification can be used to determine access rights; for example, only users in the payments team are allowed to see plaintext payment card data to process customer transactions. This clearly identifies the need for obfuscation, tokenization, or other methods of blocking users on other teams from accessing the data.
- **Encryption needs**: Encryption is a multipurpose tool for security and privacy, so the application and configuration of encryption must be clearly documented for users to ensure it is properly applied.

DESIGN AND IMPLEMENT INFORMATION RIGHTS MANAGEMENT

Since data is highly portable and there is great value in collaborating and sharing access, it is often necessary to ensure an organization's security controls can be extended to offer protection for the data wherever it might be. Data that is shared outside the organization will likely end up on information systems and transit networks not controlled by the data owner, so a portable method of enforcing access and use restrictions is needed: information rights management, sometimes also called *digital rights management* (DRM). There are two main categories of IRM.

- Consumer-grade IRM is more frequently known as DRM and usually focuses on controlling the use, copying, and distribution of materials that are subject to copyright. Examples include music, videogame, and application files that may be locked for use by a specific (usually paying) user, and the DRM tool provides copy protections to prevent the user from distributing the material to other, nonpaying users.
- Enterprise-grade IRM is most often associated with digital files and content such as images and documents. IRM systems enforce copy protection as well as usage restrictions, such as PDFs that can be read but prevent data from being copied or printed, and images that can be accessed for only a certain duration based on the license paid for. IRM can also be a form of access control, whereby users are granted access to a particular document based on their credentials.

IRM is often implemented to control access to data that is designed to be shared but not freely distributed. This can include sensitive business information shared with trusted partners but not the world at large, copyrighted material to which a user has bought access but is not authorized to share, and any information that has been shared under a license that stipulates limitations on the use or dissemination of that information.

Objectives

Most IRM solutions are designed to function using an access control list (ACL) for digital files, which specifies users and authorized actions such as reading, modifying, printing, or even onward sharing. Many popular file sharing SaaS platforms implement these concepts as sharing options, which allow the document owner to specify which users can view, edit, download, share, etc.

IRM systems should ideally possess a number of attributes, including the following:

- **Persistence**: The ACL and ability to enforce restrictions must follow the data. Some tools allow users to set a password required to open a document, but the tools also allow other users to disable this password-based access control, which defeats the purpose.
- **Dynamic policy control**: The IRM solution must provide a way to update the restrictions, even after a document has been shared. If users no longer require access, the IRM solution must provide a way for the document owner to revoke the permission and enforce it the next time the document is opened regardless of its location. This leads to a key usability challenge, as IRM tools often require users to have an active network connection so the policy can be validated before access is granted.
- **Expiration**: IRM tools are often used to enforce time-limited access to data as a form of access control, which reduces the amount of time a bad actor has to exploit a document to which they have gained unauthorized access. While this can be an element of dynamic policy control, which requires the ability to query an IRM server, it may also be done by calculating and monitoring a local time associated with the file. One example is the timer that begins when a user first starts playback of a rented digital movie and restricts the user's ability to play the movie to 24 hours after that initial start time.
- **Continuous audit trail**: Access control requires the ability to hold users accountable for access to and use of data. The IRM solution must ensure that protected documents generate an audit trail when users interact with protected documents to support the access control goal of accountability.
- **Interoperability**: Different organizations and users will have a variety of tools, such as email clients and servers, databases, and operating systems. IRM solutions must offer support for users across these different system types. Document-based

IRM tools often utilize a local agent to enforce restrictions, so support for specific operating systems or applications is a critical consideration. System-based IRM tools, such as those integrated into document repositories or email systems, are capable of broad support despite the user's OS, but may offer limited support for user applications like browsers or email clients. Lastly, sharing documents across systems can be challenging, especially with outside organizations who may utilize different services such as Microsoft Office and Google Apps for collaboration tools.

IRM restrictions are typically provisioned by a data owner, whose responsibilities will vary depending on the access control model being used. In a discretionary access control (DAC) model, the owner is responsible for defining the restrictions on a per-document or data set basis. This may involve manual configuration of sharing for documents, specifying user authorizations for a database, or defining users and their specific rights for a data set. In nondiscretionary access control models such as mandatory access control (MAC), the owner is responsible for specifying metadata like a classification rating or a user role. The IRM system then utilizes this metadata to enforce access control decisions, such as allowing access to users with the same clearance level or denying users who are not assigned specific job roles.

Appropriate Tools

IRM tools comprise a variety of components necessary to provide policy enforcement and other attributes of the enforcement capability. This includes creation, issuance, storage, and revocation of certificates or tokens, which are used to identify authorized users and actions. This requires a centralized service for identity proofing and certificate issuance, as well as a store of revoked certificates that can be used to identify information access that is no longer authorized. This model is used for software distribution via app stores, where developers digitally sign code and user devices validate the signature each time the app is launched. This can ensure that the device is still authorized and the user is still the authentic license holder, but also offers the ability for the entity controlling the app store to prevent apps from running if their certificates have been revoked. Such a solution obviously requires network connectivity between devices and the centralized management system.

Both centralized and decentralized IRM solutions will require local storage for encryption keys, tokens, or digital certificates used to validate users and access authorizations. This local storage requires protection primarily for the integrity of data to prevent tampering with the material used to enforce IRM. Modifying these access credentials could lead to loss of access control over the IRM-protected data; for example, a user might modify the permissions granted to extend their access beyond what the data owner originally specified.

PLAN AND IMPLEMENT DATA RETENTION, DELETION, AND ARCHIVING POLICIES

Data follows a lifecycle starting with its creation and ending when it is no longer needed by the organization. There are many terms used to describe this end-of-life phase including disposal, retention, archiving, and deletion. Although sometimes used interchangeably, they are unique practices, and a CCSP must be aware of requirements that mandate the use of one specific option.

Data disposal is most often associated with the destruction or deletion of data, though the term may be used to mean disposition, which implies a change of location for data such as moving it from active production to a backup environment. While data is still required for use by the organization or must be held for a set period of time to meet a regulatory or compliance objective, the practice of *data retention* will be used. Once data is no longer needed by the organization and is not subject to any compliance requirements for retention, it must be deleted using tools and processes commensurate with its value.

Data archiving is a subset of retention typically focused on long-term storage of data not required for active processing or that has historical value and may therefore have high integrity requirements. Data retention, archive, and destruction policies are highly interconnected, and the required practices may even be documented in a single policy or set of procedures governing the use, long-term storage, and secure destruction of data.

Data Retention Policies

Data retention is driven by two primary objectives: operational needs (the data must be available to support the organization's operations) and compliance requirements, which are particularly applicable to sensitive regulated data such as PII, healthcare, and financial information. Many regulatory documents refer to data as records; hence, the term *records retention* is often used interchangeably. Data retention policies should define a number of key practices for the organization and need to balance organizational needs for availability, compliance, and operational objectives such as cost.

A CCSP should recognize several use cases for cloud services related to backups, which are often a key subset of data retention practices related to availability. Cloud backup can replace legacy backup solutions such as tape drives, which are sent to offsite storage. This offers the benefit of more frequent backup with lower overhead, as sending data over a network is typically cheaper than having a courier pick up physical drives, but the organization's network connection becomes a single point of failure.

Cloud backup can be architected as a mere data storage location, or a full set of cloud services can act as a hot site to take over processing in the event the on-premises environment fails. This scenario may be cost effective for organizations that are unable to use

the cloud as a primary processing site but do not want to incur the costs of a full hot site. Temporary use of the cloud as a contingency helps to balance cost and security.

Cloud services, particularly SaaS and PaaS deployments, may offer intrinsic data retention features. Most cloud storage services provide high availability or high durability for data written into the environment, which allows the organization to retain vital data to meet operational needs. Some services also offer compliance-focused retention features designed to identify data or records stored and ensure compliance obligations are met. In all cases, the CCSP needs to be aware of the features available and ensure their organization properly architects or configures the services to meet internal and compliance obligations.

Storage Costs and Access Requirements

Data retention has associated storage costs, which must be balanced against the requirements of speed to access it. In many cases, older data, such as old emails or archived documents, is accessed less frequently. This data may not be accessed at all for routine business matters, but is only needed for exceptional reasons like archival research or a legal action where it is acceptable to wait a few hours for a file to be retrieved. Organizations may also have compliance or regulatory obligations to retain data long after it is no longer useful for daily operations. The costs associated with this storage can be significant, so CSPs offer a variety of storage services that balance cost and retrieval speeds; typically the solutions offer a combination of either low price/retrieval speed or higher price and quick retrieval. As an example, Amazon Simple Storage Service (S3) offers higher-priced S3 Standard, where retrieval is in real time, or lower-priced S3 Glacier, where retrieval time ranges from 1 minute to 12 hours. Similarly, Microsoft's Azure Blob Storage offers Hot, Cool, and Archive tiers, in order of higher cost/retrieval speed to lower cost/speed.

To model this cost-benefit analysis, consider Alice's Blob Cloud (ABC), which offers the following storage service levels (currency is USD per gigabyte):

- **Rabbit Storage**: $0.5, real-time access (>50 milliseconds)
  - Storing 5TB of data (5000GB) would cost $2,500/month or $30,000/year.
- **Turtle Storage**: $0.005, access times from 1 to 12 hours
  - Storing 5TB of data (5000GB) would cost $25/month or $300/year.

The costs savings of using Turtle are significant: $29,700 per year in this limited example. Most organizations will generate significantly more than 5TB of data. If the data is used infrequently, such as once a quarter, the data retention policy should specify appropriate storage options to balance costs with the access speed requirements. This may be part of the records retention schedule (for example, all data that is older than 12 months is moved to Turtle) or as part of the data classification policy (for example, live production data is encrypted at rest and stored in Rabbit, while archive production is encrypted and stored in Turtle).

Specified Legal and Regulatory Retention Periods

All organizations should retain data for as long as it is functionally useful; otherwise, the organization faces a loss of availability. This may be marked by obvious milestones such as the end of a customer relationship or project, after which the data is no longer valuable to the organization and can be deleted. Any organization dealing with regulated data like PII is also likely to have external obligations for data retention, which are often legally enforceable. Some examples of external regulations include the following:

- **HIPAA**: Affects all U.S. residents and specifies a six-year retention period for documents, such as policies and procedures, relevant to the HIPAA compliance program. Retention of patient medical data is not directly mentioned in HIPAA but is specified by many state-level laws that require medical records be retained for as long as a patient is active and then for a set period of time thereafter.
- **EU GDPR**: Affects data of EU citizens; it does not set a specific retention period but rather provides for indefinite retention so long as a data subject has given consent and the organization has a legitimate need for the data. If consent is revoked or the organization must act on the revocation by deleting, destroying, or anonymizing the data.

Data Retention Practices

The organization's data retention policy should specify what data is to be retained and why. Procedures should also be documented to specify how those retention goals will be met, including details regarding the following:

- **Schedules**: Data and records retention often refers to a schedule of retention, which is the period of time the data must be held for. These are often included in data labels to enable the discovery of data that is no longer required and can be destroyed.
- **Integrity checking**: Whenever data is copied, there is a chance something could go wrong, and data stored may fall victim to environmental issues like heat or humidity, which damage the storage media. Integrity checking procedures should be established to verify data when it is written and periodically thereafter to ensure it is readable and complete.
- **Retrieval procedures**: Data may have different access requirements across different environments. Typically, there will be more users authorized to access data in a live production environment as it is needed to support operations, while data in an archive may only be needed by a more limited set of users, like the legal department when responding to a lawsuit or auditors reviewing regulatory compliance. Retrieval procedures should include proper authorization artifacts like approved access requests and enforce accountability for data usage.

- **Data formats**: The format of data, including programs, apps, and hardware needed to read and write it, requires consideration. Over time, file formats and hardware change, so procedures such as virtualizing legacy environments, purchasing reserves of obsolete equipment, or converting data to new formats may be appropriate.

Data Security and Discovery

Retained data will face unique security challenges, particularly driven by the fact that it is long lived and may be difficult to access or manipulate as threat conditions change over time. For example, data encryption standards evolve quite frequently, so today's state-of-the art cryptography may be trivially easy to crack in a few years. It may not be technically or financially feasible to decrypt data retained in archives and re-encrypt using more robust cryptography. Security practitioners should consider defense-in-depth strategies such as highly secure key storage and tightly limited access control over archival data as a compensating control for weaker legacy encryption standards. Similarly, keys that are no longer actively used for encrypting data in production environments may need to be securely stored to grant access to archival data.

Data retention is often a requirement to support after-the-fact investigations, such as legal action and review by regulators. Data retention methods should support the ability to discover and extract data as needed to support these compliance obligations. For legal actions the requirements of eDiscovery must be considered. eDiscovery is covered in detail in Domain 6: Legal, Risk, and Compliance, but in short, it is the ability for data to be queried for evidence related to a specific legal action, such as all records during a certain time period when fraudulent activity is suspected.

Data Deletion Procedures and Mechanisms

When data is no longer needed for operational needs and has been retained for the mandated compliance period, it can be disposed of. It may, however, still be sensitive, such as a medical records of a patient who is no longer being treated at a particular facility, but is still living and is legally entitled to privacy protections for their medical data. In this case, simply disposing of the data by selling old hard drives or dumping paper files into the trash would open the organization and the patient to risk, so proper controls must be enforced to ensure the confidentiality of information remains intact during destruction.

NIST SP 800-88, *Guidelines for Media Sanitization*, is a widely available standard for how to securely remove data from information systems when no longer required. It defines three categories of deletion actions for various types of media to achieve *defensible destruction*—the steps required to prove that adequate care was given to prevent a breach

of data confidentiality. These categories, in hierarchical order based on protection they provide, are as follows:

- **Clear**: The use of tools to remove or sanitize data from user-addressable storage. Clearing may include standard operating system functions like deleting data from a trash can/recycle bin, which merely renders the data invisible but recoverable using commonly available tools. These options are typically the fastest and lowest cost but are inappropriate for very sensitive data.
- **Purge**: The use of specialized tools like overwriting drives with dummy data, physical state changes such as magnetic degaussing, or built-in, hardware-based data sanitization functions designed to provide secure destruction of data. Purged media can typically be reused, which may be a cost factor to consider, but the time required to perform purge actions like writing 35 passes of dummy data over a modern high-capacity hard drive might make it infeasible. Purging may also shorten the lifespan of media to such an extent that its remaining useful life is negligible. Data may be recovered from purged media using highly specialized tools and laboratory techniques and may be appropriate for moderate risk data where no determined attacker with adequate means exists.
  - *Cryptographic erasure* or *Cryptoshredding* is a form of purging that utilizes encryption and the secure destruction of the cryptographic key to render data unreadable. This is effectively a positive denial-of-service attack and is often the only option available for cloud-based environments due to loss of physical control over storage media, the use of SSDs for storage that cannot be reliably overwritten, and the dispersion of data in cloud environments. Organizations utilizing the cloud can encrypt all data using organization-controlled keys, which can be securely destroyed, rendering data stored in the cloud economically infeasible to recover. Modern smartphone, tablet, and workstation operating systems also implement this feature using technologies such as Apple FileVault or Microsoft BitLocker, which save costs of purging and extends the useful life of storage media.
- **Destroy**: The most drastic option that renders storage media physically unusable and data recovery infeasible using any known methods. Destruction techniques include physical acts like disintegrating, pulverizing, melting, incinerating, and shredding. It is unlikely a CSP will provide the capability for cloud consumers to physically destroy media, but this may be an appropriate control for the CSP to implement for information system components that contain sensitive customer data but that are no longer needed.

The choice of data deletion procedures should be driven by a cost-benefit analysis. Cost including replacement of the media, fines, or legal settlements if a data breach

occurs, and the actual implementation of destruction must all be taken into account. As an example, hard drives containing high-sensitivity information may simply be cleared if they are to be reused in the same environment where the risk of a data breach is low, but may be physically destroyed if they are leaving the organization's control. The full NIST SP 800-88 document covering data destruction and associated risk factors can be found here: csrc.nist.gov/publications/detail/sp/800-88/rev-1/final.

Data Archiving Procedures and Mechanisms

Data archiving refers to placing data in long-term storage for a variety of purposes: optimizing storage resources in live production environments and meeting the organization's retention requirements are both examples. The procedures and mechanisms in place need to ensure adequate security controls are in place for data as it moves from live systems to the archive, which may implement significantly different controls for access control and cryptography.

Access controls for production environments are typically more complex than for archive environments, but that does not mean the archive deserves less rigor. Archivists or retention specialists may be the only users authorized to routinely access data in the archive, and procedures should be in place to request, approve, and monitor access to the data. Procedures governing the handoff between production and the archive should be documented as well, to ensure the change of responsibility is well understood by the personnel assigned.

Cryptography serves multiple important roles for archival data just as it does in other environments. Data in transit to an archive will need to be protected for both confidentiality and integrity; for cloud-based systems or backup tools, this will entail encrypting data in transit to preserve confidentiality as the data moves between on-premises and cloud environments, and verifying the integrity of data copied to the cloud environment.

Hashing may be appropriate for data with high integrity requirements. Data can be hashed at a later date when accessed from the archive, and the values compared to an initial hash from when the data was first stored. This will identify if changes have occurred. In some cases, integrity can also be verified by loading backup data into a production-like environment to verify it is readable and conforms to expectation. This would be particularly appropriate for a cloud hot site where failover needs to happen quickly.

In addition to mandated retention periods, security practitioners must understand requirements for data formats mandated by their legal and regulatory obligations. For high-sensitivity data, particularly in the financial services industry, there may be a requirement for data to be stored immutably; that is, in a format where it cannot be changed. The integrity of the data is required to support investigations of regulated activity such as financial transaction, which will drive decisions on where and how to store the data. Once data is written, it will require adequate physical and environmental protections as well, to prevent theft, tampering, or degradation.

Write once, read many (WORM) media is one example of high-integrity media: data written to the media is physically unalterable, preventing a user from covering up evidence of fraudulent activity. Some cloud storage services implement similar write protections for integrity along with highly durable storage media; a CCSP should ensure proper storage solutions are chosen to meet the organization's need. Blockchain technology is also being used for verifiable integrity, as blockchains rely on hashing to defensibly prove data is legitimate. An organization's storage solution might be integrated with a blockchain ledger to prove data has not been modified since it was written as a way to prove the integrity of the data.

Legal Hold

Legal hold is a simple concept but has important implications for data retention. When data is placed under legal hold, its retention schedule is indefinitely suspended; if it should be retained for seven years but is placed under legal hold, it must be retained until the legal hold is lifted, even if the seven-year retention period passes. Determining legal hold is usually not within the purview of the CCSP, but they should be able to respond to such requests when the organization is involved in legal action such as a lawsuit.

The primary challenges surrounding legal hold include identifying applicable records to be held and implementing a way to exclude records from standard deletion procedures while under hold. Legal requests may often be vague or specifically broad, as they may be related to suspected wrongdoing and seek to obtain evidence; this leads to problems for data archivists and security practitioners when determining which records to place under hold. Internal legal counsel should always be consulted, and in general, it is better to retain more rather than less.

The second challenge of excluding records under legal hold from deletion is a data management problem. Hard-copy records under hold can be easily separated from other records by moving them to a secure facility and ignoring them when performing deletion, but electronic information systems are more complex. Copying records to a separate storage solution may be feasible but introduces challenges of preserving integrity during the process as well as the need to set up an access-controlled legal hold archive location. Another approach is the use of metadata to electronically flag records to exclude them from deletion. Many databases and filesystems support this functionality, but the archivist or security practitioner must also ensure supporting systems such as cryptographic key management are aware of records under legal hold. Otherwise, encrypted records may be retained because they are flagged, but the key needed to decrypt them may be deleted when its retention period expires.

DESIGN AND IMPLEMENT AUDITABILITY, TRACEABILITY, AND ACCOUNTABILITY OF DATA EVENTS

All the topics discussed thus far fall into broad categories of data governance and data management. Just like any other valuable asset, an organization must identify strategic goals and requirements for managing, using, and protecting data. Despite best efforts to safeguard these assets, there will be times when it is necessary to collect evidence and investigate activity that could support assertions of wrongdoing. Gathering this digital evidence requires the existence of relevant data in logs or other sources, as well as capabilities to identify, collect, preserve, and analyze the data. An organization with no preparation for these activities will find it difficult or impossible, and security practitioners should ensure their organizations are not put in situations where they are unable to investigate and hold bad actors accountable.

Definition of Event Sources and Requirement of Identity Attribution

There are formal definitions of events used in IT service management as well as security incident response. In general, an event is any observable action that occurs, such as a user logging in to a system or a new virtual server being deployed. Most IT systems support some ability to capture these events as running logs of past activity, though cloud-based systems can present a number of challenges that a CCSP should be aware of.

The primary concern regarding information event sources in cloud services is the accessibility of the data, which will vary by the cloud service model in use. IaaS will obviously offer the most data events as the consumer has access to detailed logs from network, OS, and application data sources, while in SaaS, the consumer may be limited to only data events related to their application front end with no access to infrastructure, OS, or network logs. This loss of control is one cost factor organizations should weigh against benefits of migrating to the cloud and may be partially offset with contract requirements for expanded data access in exceptional circumstances like a data breach.

The Open Web Application Security Project® (OWASP) publishes a series of guides called *cheat sheets*, including one on logging. The guide covers a variety of scenarios and best practices and also highlights challenges that a CCSP is likely to face when using cloud services, particularly the inability to control, monitor, or extract data from resources outside the organization's direct control. Some of these scenarios include the following:

- **Synchronize time across all servers and devices**: The timestamp of events in a log is crucial to establish a chain of activity performed by a particular user. If a cloud service is hosted and generates logs in a time zone different from the user's workstation, it will require significant mental overhead to trace events.

- **Differing classification schemes**: Different apps and platforms will categorize events with different metadata. For example, one OS may identify user logins as "Informational" events, while apps running on that OS log the same activity as "User Events." Constructing queries on such varied data will be difficult.
- **Identity attribution**: Ultimately logs should be able to answer the basic question "Who did what and when?" Sufficient user ID attribution needs to be easily identifiable; otherwise, it may prove impossible to definitely state that a particular user took a given action. The organization's identity and access management system may offer the ability to utilize a single identity across multiple services, but many cloud services enforce their own user management, complicating the attribution of events to a specific user.
- **Application-specific logs**: Apps may offer the ability to generate logs, but unlike widely standardized tools like operating systems or web server software, they may log information in a unique format. SaaS platforms may offer even fewer configuration options as internal functionality such as event logging configuration is not exposed to the end users. Making sense of this data in relation to other sources, such as an app that uses a numeric ID value for users while others use an email address, can be challenging.
- **Integrity of log files**: App and OS log files typically reside on the same hosts as the software generating the logs and so are susceptible to tampering by anyone with privileged access to the host. If a user has administrative permissions, they may be able to not only perform unauthorized actions but then cover up evidence by modifying or deleting log files.

The full Logging Cheat Sheet is available here: cheatsheetseries.owasp.org/cheatsheets/Logging_Cheat_Sheet.html.

Logging, Storage, and Analysis of Data Events

Logs are made valuable only by review; in other words, they are valuable only if the organization makes use of them to identify activity that is unauthorized or compromising. Due to the sheer volume of log data generated, it is unlikely a human being would be capable of performing log reviews for any computing environment more complex than a single application.

SIEM tools, which are covered in detail in Domain 5: Cloud Security Operations, can help to solve some of these problems by offering these key features:

- **Log centralization and aggregation**: Rather than leaving log data scattered around the environment on various hosts, the SIEM platform can gather log data from operating systems, applications, network appliances, user workstations, etc., providing a single location to support investigations. This is often referred to as

forwarding logs when looked at from the perspective of individual hosts sending log files to the SIEM.

- **Data integrity**: The SIEM platform should be on a separate host with access permissions unique from the rest of the environment, preventing any single user from tampering with data. System administrators may be able to tamper with logs on the systems they are authorized to access but should be denied write access to log data stored on the SIEM platform.
- **Normalization**: The same piece of data can often be formatted in a variety of ways, such as dates written YYYY/MM/DD, DD/MM/YYYY, or even the Unix format, which is a count of the number of seconds measured from a start time of January 1, 1970. SIEMs can normalize incoming data to ensure the same data is presented consistently; for example, all timestamps are converted to utilize UTC rather than the time zone of the originating system.
- **Automated or continuous monitoring**: Sometimes referred to as *correlation*, SIEMs use algorithms to evaluate data and identify potential attacks or compromises. These can often be integrated with other security tools such as intrusion detection systems (IDS) and firewalls to correlate information, such as a large number of failed logins by users after the user visited a particular website. This is indicative that users may have fallen for a phish, and the attackers are now trying to use those credentials. Crucially, this monitoring can be automated and is performed continuously, cutting down the time to detect a compromise.
- **Alerting**: SIEMs can automatically generate alerts such as emails or tickets when action is required based on analysis of incoming log data. Some also offer more advanced capabilities like IPS, which can take automated actions like suspending user accounts or blocking traffic to/from specific hosts if an attack is detected. Of course, this automated functionality will suffer from false positives, but it performs at much greater speed compared to a human being alerted to and investigating an incident.
- **Investigative monitoring**: When manual investigation is required, the SIEM should provide support capabilities such as querying log files, generating reports of historical activity, or even providing concise evidence of particular events that support an assertion of attack or wrongdoing; for example, a particular user's activity can be definitively tracked from logging on to accessing sensitive data to performing unauthorized actions like copying the data and sending it outside the organization.

Chain of Custody and Nonrepudiation

Data that is collected in support of an investigation has unique requirements for integrity, particularly one in which civil or criminal prosecution is called for. It is vital to establish a *chain of custody*, or a defensible record of how evidence was handled and by whom, from its collection to its presentation as evidence. If data has been altered after it was collected, the defendant might make a case that they are innocent because the evidence has been altered to make them look guilty.

Chain of custody and evidence integrity do not imply that data has not changed since collection, but instead they provide convincing proof that it was not tampered with in a way that damages its reliability. For example, data on a user's workstation will physically change location when the workstation is collected from the user, and the data may be copied for forensic analysis. These actions, when performed by properly trained professionals, do not change the underlying facts being presented or the believability of evidence. However, if the workstation is left unattended on a desk after it is taken from the user, then the user can claim the data was altered in a way that incriminates them, and thus the evidence is no longer reliable.

Nonrepudiation can be a challenging term because it is proving a negative. Repudiating an action means denying responsibility for the given action, so nonrepudiation is a characteristic whereby you can definitely hold a particular user accountable for a particular action. In nontechnical terms, it can be difficult to hold anyone accountable for drinking the last of the office coffee from a shared coffee maker; everyone can repudiate the assertion that they finished the coffee without brewing more. If users are required to enter a PIN to unlock the coffee maker, then it is possible to defensibly prove, when the coffeepot is found to be empty, that the last person whose code was used is the culprit.

Information systems enforce nonrepudiation partially through the inclusion of sufficient evidence in log files, including unique user identification and timestamps. System architecture and limitations can pose challenges such as shared user accounts that are tied to a group of users but no single user, as well as processes that act on behalf of users, thereby obscuring their identity. Nonrepudiation is a concern not just for data event logging but also access control and system architecture. Security practitioners must ensure their access control, logging, and monitoring functions support nonrepudiation.

One final note on chain of custody and nonrepudiation: the process of investigating digital crimes can be quite complex and may best be left to trained professionals with appropriate skills and experience. Simple acts like plugging in a drive can cause irreversible data loss or corruption, and destroying evidence may limit the organization's ability to seek legal redress if a crime has occurred. While CCSPs should be familiar with these practices, it is also advisable to know when expert professionals are required.

SUMMARY

Cloud services offer many benefits for accessing, managing, and handling the data that are crucial to modern business operations, but they are not free from risk. The cloud data lifecycle provides a convenient framework for identifying the types of activities, risks, and appropriate security controls required to ensure data remains secure. This, coupled with an understanding of the storage architectures available in different cloud service models, allows a CCSP to design and apply appropriate safeguards such as data discovery, encryption, and tokenization, as well as countermeasures like DLP. Proper cloud data security requires organizations to know what kind of data they handle and where it is stored, and deploy adequate policies, procedures, and technical controls to ensure the business benefits of cloud environments are not offset by increased information-security risks.

DOMAIN 3

CCSP®

Cloud Platform and Infrastructure Security

THE CLOUD SECURITY PROFESSIONAL must understand the cloud infrastructure for each of the cloud delivery models. The infrastructure includes the physical components of the cloud, the services they provide, and the communication infrastructure that allows us to connect with the cloud. Each part of the infrastructure has specific security needs and shared security responsibilities.

In the shared security model of cloud computing, it is easy, but incorrect, to assume that security is the sole responsibility of the Cloud Service Provider (CSP). The cloud security professional needs to understand the unique security needs of the cloud environment and the technologies, such as virtualization, that make the cloud possible. The security professional must understand the nuances affecting cloud security, including which part of security is the responsibility of the CSP and which part is the responsibility of the customer. The security professional must be clear on how the security provided by the CSP and the customer work together to protect the customer's processes and data.

Finally, the cloud security professional must understand how the cloud can support the business. The cloud supports more than day-to-day operations.

The cloud also provides important tools to support business continuity and disaster recovery. In this role, the cloud becomes more of a business continuity and resiliency tool.

COMPREHEND CLOUD INFRASTRUCTURE COMPONENTS

There are common components to all cloud infrastructure. These components are all physically located with the CSP, but many are accessible via the network. In the shared responsibility security model, both the customer and the CSP share security responsibilities for the cloud environment. Those responsibilities are discussed with each component.

Physical Environment

The physical environment is composed of the server rooms, data centers, and other physical locations of the CSP. The identity of the CSP may vary by the different cloud security models, as follows:

- A private cloud is often built by a company on-premise. If this occurs, the company/customer is the CSP. If the private cloud is built using a commercial vendor, such as AWS or Azure, the CSP is the commercial vendor.
- In a community cloud, a member of the community often hosts the space and equipment needed. The community member hosting the cloud is the CSP. If the community cloud is built using a commercial cloud service, such as AWS or Azure, the CSP is the commercial vendor.
- In public clouds, the CSP is the company providing the service, such as AWS, Microsoft, Google, IBM, and so on. This commercial vendor is the CSP.

The physical environment is under the sole control of the CSP. CSPs provide all physical security. They are also responsible for monitoring, auditing, and maintaining a secure environment. The security risks are those common to all data centers, regardless of whether they are on-premise or provided by a third party. These include physical security of the equipment, identity and access management, data integrity and confidentiality, and so on.

The CSP uses the typical methods for these risks. For physical security these measures include locks, security personal, lights, fences, etc. Identity and access management (IAM) will use tools such as SAM or OAuth2 or may use a cloud vendor such as Okta to

provide these services. IAM allows single sign-on (SSO) and is supported by multifactor authentication, tokens, and other authentication tools. Data confidentiality and integrity are supported by IAM and the use of cryptography technologies such as encryption, message digest, digital certificates, and public-key infrastructures (PKI), and transmittal methods such as Secure Sockets Layer (SSL), and virtual private networks (VPNs).

Network and Communications

The networking components housed by the CSP are the responsibility of the CSP, and they maintain security. Components housed at the customer's facility (on-premise) are the responsibility of the customer. The largest area of concern is the public Internet that exists between these two. Who is responsible for the security of the public Internet? The answer would be the individuals using it. Security is a shared responsibility.

For this reason, the CSP must support and the customer must use secure protocols such as HTTPS, encrypt the data prior to transmission, or use a VPN for secure data communication. At each end of the transmission pipeline, both the customer and the CSP are responsible for the firewalls and other systems needed to maintain secure communications.

In the shared security model, the CSP provides tools for secure computing, logging, encryption, and so on. However, it may be the responsibility of the customer to properly set up and configure these services. In all cases, it is the responsibility of the customer to connect to and transmit data to the service securely.

Consider, for example, that you are a company shipping a product. You have many ways in which you can accomplish this.

- You could create your own fleet, leasing aircraft from a supplier, but providing your own pilots and cargo personnel. In this case, it is similar to an infrastructure as a service (IaaS) environment. The supplier is responsible for the aircraft provided, but not how you use the aircraft OR the safety and security of your product.
- In a different approach, the supplier may provide the aircraft and pilots, and you handle the storage and security of the cargo. You remain responsible for the safety and security of your product, and the supplier provides a safe aircraft that is flown safely.
- In the simplest method, you drop off the package with the suppler. Once in their possession, the supplier is responsible for the security of the package. They are not responsible for the manner in which the product is packaged or the condition of the product when it is dropped off with the service.

IaaS, platform as a service (PaaS), and software as a service (SaaS) work similarly to the previous list of examples.

- In an IaaS service model, the customer is responsible for configuring the environment as well as enforcing company policies in the use of systems, as if the systems were on-premise as well as the connection to the CSP. The CSP is responsible for the technology provided but not how it is used.
- In a PaaS module, the CSP is responsible for the physical components, the internal network, and the tools provided. The customer is responsible for the proper use of those tools and the connection to the CSP.
- In a SaaS model, the customer remains responsible for access to the cloud service in a secure manner—using appropriate technologies to connect with and transmit data to and from the service securely.

Compute

The compute resources are the infrastructure components that deliver virtual machines (VMs), disk, processor, memory, and network resources. These resources are owned by and under the direct control of the CSP. The security issues are the same for these resources in a cloud environment as they are on-premise with the additional challenge of multitenancy.

The CSP in every delivery and service model remains responsible for the maintenance and security of the physical components. These are owned and supported by the CSP. The customer in every delivery and service model remains responsible for their data and their users. Between the physical components, there is a vast array of software and other components. Who is responsible for each of these remaining parts varies by service and delivery model and sometimes by the CSP. The contract between the customer and the CSP should spell out the responsibilities for each part of the cloud environment. Typical responsibilities will be described for each service model (IaaS, PaaS, and SaaS).

In an IaaS environment, the CSP provides the hardware components and may provide networking, virtualization, and operating systems. The security of physical components provided by the CSP are the responsibility of the Internet Service Provider (ISP). When the CSP provides software components, such as virtualization and Operating System (OS) software, they are responsible for the versioning and security of those components.

When the CSP provides the virtualization and OS software, some of the software configuration may be left to the customer. When this is the case, the customer is responsible for the security implications of the configuration they choose. In the IaaS service model, the customer has the most responsibility for security. Other than the hardware components and system software provided, the customer remains responsible for all other security for the tools they install, software they develop, and, of course, all identity and access management, customer records, and other data. These responsibilities include the patching and versioning of the software installed or developed by the user and the security of data at rest and in motion.

In a PaaS environment, the CSP is responsible for the security and maintenance of all the components and systems they were responsible for in the IaaS service model. In addition, the CSP is responsible for all additional services they provide. These services can be storage, transform and analysis, and numerous other services. While the customer may have some ability to configure these services, all services provided by the CSP will usually be maintained by the CSP, to include patching and versioning. The customer is responsible for the configuration and use of all CSP systems and services provided as well as the patching and versioning of all software the customer installs or develops. The customer is always responsible for the security of their data and users. The contract with the CSP should address these issues.

In a SaaS environment, the customer is generally responsible for the customization of the SaaS service, as well as the security of their users and the data used or produced by the service. The CSP is responsible for the security of all other compute resources.

Virtualization

There are two types of two types of hypervisors that provide virtualization. These are Type-1 hypervisors (also known as *bare-metal hypervisors*) and Type-2 hypervisors (also known as *hosted hypervisors*). Among the major CSPs, a Type-1 hypervisor is more common. These include the version of the Xen hypervisor provided by AWS and the Hyper-V hypervisor provided in Azure.

In a Type-1 hypervisor, the hypervisor sits on the physical server and its associated hardware components. The Type-1 hypervisor does not sit on an OS. Rather, the hypervisor provides the OS functionality. VMs sit atop the hypervisor. You manage the virtual environment using a management console separate from the hypervisor. In the Type-1 environment, VMs can move between physical machines. When done properly, it is invisible to the end user. In addition to Hyper-V and the Xen Server (now called Citrix) hypervisors, other common Type-1 hypervisors include VMware vSphere with ESX/ESXi and Oracle VM.

In a Type-2 hypervisor, there are typically three layers in the virtualization portion. At the bottom is the host OS, such as Windows macOS or Linux. The next layer is the hypervisor itself. Examples include Oracle VM VirtualBox, VMware Workstation Pro/VMware Fusion, and Windows Virtual PC. These are not usually used for enterprise solutions but are used for individual needs and testing environments. The top layer consists of the individual VMs. Management of the VMs is built into the Type-2 hypervisor.

The security of the hypervisor is always the responsibility of the CSP. The virtual network and virtual machine may be the responsibility of either the CSP or the customer, depending on the cloud service model as described in the earlier "Compute" section.

Hypervisor security is critical. If unauthorized access to the hypervisor is achieved, the attacker can access every VM on the system and potentially obtain the data stored on

each VM. In both Type-1 and Type-2 hypervisors, security of the hypervisor is critical to avoid hypervisor takeover as the hypervisor controls all VMs. In a Type-2 hypervisor, host OS security is also important, as a breach of the host OS can potentially allow takeover of the hypervisor as well. Proper IAM and other controls limiting access to those with both proper credentials (authentication) and a business need (authorization) can protect your systems and data. In a cloud computing multitenant model, the problem has an additional challenge. The attacker may have permission to be on the server hosting the VMs, as the attacker may be another customer of the CSP. Security of the hypervisor is the responsibility of the CSP.

CSP hypervisor security includes preventing physical access to the servers and limiting local and remote access to the hypervisor. The access that is permitted must be logged and monitored. The CSP must also keep the hypervisor software current and updated.

The virtual network between the hypervisor and the VM is also a potential attack surface. The responsibility for security in this layer is often shared between the CSP and the customer. In a virtual network, you have virtual switches, virtual firewalls, virtual IP addresses, etc. The key to security is to isolate each virtual network so that there is no possibility to move from one virtual network to another virtual network. This isolation will reduce the possibility of attacks being launched from the physical network or from the virtual network of other tenants, preventing attacks such as VLAN hopping.

A control layer between the real and virtual devices such as switches and firewalls and the VMs can be created through the use of a software-defined networking. In AWS, when you create a virtual private cloud (VPC), the software-defined networking creates the public and private networking options (subnets, routes, etc.). To provide security to the software-defined network, you will need to manage both certificates and communication between the VM management plane and the data plane. This includes authentication, authorization, and encryption.

The security methods for a virtual network are not that much different from physical networks. But the tools used on the physical networks may be unable to see and monitor virtual traffic. Using security tools designed specifically for virtual environments is recommended.

The final attack surface in the virtualization space is the VM itself. The responsibility for the security of the VM may be shared but is usually the responsibility of the customer in an IaaS model. In the PaaS model, the security of the VM may be the responsibility of either the CSP or the customer, depending on who creates and uses the VM. If the CSP uses a VM to provide a PaaS service, the responsibility for security is the CSP's. However, if the customer is creating VMs, the customer is responsible for the security of their VMs. In a SaaS model, the VM is created and used by the CSP to provide a service, and the responsibility of VM security rests on the CSP.

Storage

Cloud storage has a number of potential security issues. These responsibilities are shared by the customer and the CSP. At a basic level, the CSP is responsible for physical protection of data centers, and the customer is responsible for the security and privacy of data and customer information. The CSP is responsible for the security patches and maintenance of data storage technologies and other data services they provide, such as an AWS S3 bucket. The customer is responsible for using these storage tools securely.

These tools provided by the CSP will provide a set of controls for secure storage. The customer is responsible for assessing the adequacy of these controls and properly configuring and using the controls available. These controls can include how the data is accessed (through the public Internet or via a VPN, for example) and how the data is protected at rest and in motion. For example, the CSP may provide the ability to encrypt data at rest and methods to transfer data securely. It is the customer's responsibility to use these controls to protect their data. Failure to properly configure secure storage using available controls is the fault of the customer.

In a cloud environment, you lose control of the physical medium where your data is stored while retaining responsibility for the security and privacy of that data. These challenges include the inability to securely wipe physical storage and the possibility of a tenant being allocated storage space you once had allocated to you. This creates the possibility of fragments of your data files existing on another tenant's allocated storage space. You retain responsibility for this data and cannot rely on the CSP to securely wipe the physical storage areas.

Compensating controls for the lack of physical control of the storage medium include only storing data in an encrypted fashion and employing crypto shredding when the data is no longer needed. These controls will provide protection against data fragments being retrieved.

Management Plane

The management plane provides the tools (web interface and APIs) necessary to configure, monitor, and control your cloud environment. It is separate from and works with the control plane and the data plane. If you have control of the management plane, you have control of the cloud environment.

Control of the management plane is essential and starts with limiting and controlling access to the management plane. If you lose control of the management plane, you are no longer in control of your cloud environment.

The most important account to protect is root, or any named account that has administrative/superuser functionality. The start of this protection is enforcement of a strong password policy. The definition of a strong password is an evolving question

that has recently been addressed by NIST in SP 800-63, Digital Identity Guidelines. In the final analysis, longer is better, and passphrases are easy to remember and difficult to guess.

A strong password policy needs to be coupled with other measures for the critical root and administrative accounts. They have the keys to the kingdom, so to speak. Multifactor authentication (MFA) should also be implemented. The best method is a hardware token that is stored securely. But other methods also improve on basic password protections, to include hardware tokens kept by individuals, or even SMS texts. In general, software solutions add some protection, but not as much as the hardware solutions. Over time, some methods will be or have been shown to be less secure than others. Where possible, less secure methods should be supplemented by or replaced by more secure methods.

Role-based access control (RBAC) or access groups are other methods to limit access to these sensitive accounts. Using RBAC or access groups makes management of these groups and permissions important. If rights are not deleted when an employee changes positions or employment, access can become too broad very quickly. Another step is to limit access to users on-premise or through a VPN, if remote work is required.

Another method for limiting access is to use attribute-based access control (ABAC), also called policy-based access control. Using this method, a variety of attributes can be used with complex Boolean expressions to determine access. Typical attributes such as username can be used as well as atypical attributes such as geographic and time restrictions. For example, you must be on a corporate endpoint attached to the corporate network locally if accessing the accounts after-hours.

Each of these methods can make accessing critical root or administrative accounts more difficult for both legitimate users and malicious users alike. How tightly you lock down these accounts is in direct proportion to how valuable the information and processes in your cloud are to your business. There is a balance in this, to create as much security as possible while maintaining reasonable business access.

Root and administrative accounts are typically the only accounts with access to the management plane. The end user will generally have some limited access to the service offering tools for provisioning, configuring, and managing resources. The degree of control will be determined by each business. The end user will normally be restricted from accessing the management plane. The separation of management and other workforce uses makes the creation of separate accounts for development, testing, and production an important method of control.

In instances where the management functions are shared between the customer and the CSP, careful separation of those functions is necessary to provide proper authorization and control. In a Cisco cloud environment, the management plane protection (MPP) tool is available. AWS provides the AWS Management Console.

These are some of the methods that can be used to protect the cloud management plane. A layered defense is important, and the amount of work used to protect the management plane is, in the end, a business decision. The cloud security professional must be aware of the methods available for protection in order to be a trusted advisor to the business in this matter.

DESIGN A SECURE DATA CENTER

Designing a secure data center can be challenging with the physical siting, environmental, logical and physical controls, and communication needs. In a cloud environment, many of these traditional concerns are the responsibility of the CSP or cloud vendor as they have physical control and ownership of the data center and the physical infrastructure. The customer may be able to review the physical, environmental, and logical controls of the underlying infrastructure of the vendor in limited cases.

If the vendor uses a CSP such as Google, Amazon, or IBM to provide their infrastructure needs, this becomes logistically impossible as each has more than 60 data centers, located in the four regions of North America; Asia-Pacific (APAC); Europe, Middle East, and Africa (EMEA); and Latin America.

Cloud customers have the ability to create a logical data center. A logical data center is a construct, much like a container, where the customer designs the services, data storage, and connectivity within their instance of the cloud service. However, the physical mapping of this design to the underlying architecture is not controlled by the customer as it would be in an on-premise data center.

Logical Design

The logical design of a data center is an abstraction. In designing a logical data center, the customer utilizes software and services provided by the CSP. If used securely, the logical design can provide a secure data center. The needs of a data center include access management, monitoring for compliance and regulatory requirements, patch management, log capture and analysis, and configuration of all services.

In a logical data center design, a perimeter needs to be established with IAM and monitoring of all attempts to access the data. Access control can be accomplished through various IAM methods, including authentication and authorization, security groups, VPCs, management and other consoles, and so on. The CSP equivalent to software firewalls, traffic monitoring, and similar services can be implemented to monitor data center activities and alert on potentially malicious behavior.

All services used should have a standard configuration. This configuration is determined by the business and specifies how each approved cloud service is to be configured and can be used. Using a standard pattern/configuration makes administering and maintaining cloud services simpler and often more secure. Variations from patterns approved by each business should be approved through an exception process that includes the risk of any variance. Secure baseline configurations can provide a more secure environment for the data center. The configuration of these cloud services can be monitored and changes can be either prevented or alerted and logged.

Connections to and from the logical data center must be secured using VPNs, Transport Layer Security (TLS), or other secure transmission methods. With an increasingly remote and mobile workforce, the need to access the data center becomes more important. A careful logical design can help to ensure that a secure data center in the cloud is possible.

Tenant Partitioning

Multitenant models make cloud computing more affordable but create some security and privacy concerns. If the walls between tenants are breached, your data is at risk. Multitenancy is not a new concept. Many business centers physically house multiple tenants. These business centers may provide some access control to the building and other general security services. But, if another tenant, a custodian, or other service vendor accesses your offices, then your data could be at risk. If the data is on whiteboards or scattered on various desks and in unsecured computing systems, it is at risk.

In a similar fashion, multiple customers share computing resources in a cloud environment. The vendor provides some basic security services, as well as maintenance and other services. However, if you leave your data "lying around," then your data could still be at risk and exfiltrated. However, if you monitor access, provide robust IAM services, and encrypt data in transit and at rest, the risk is greatly minimized.

In our physical business center example, if you lock up your servers and all of your data but leave the keys in a desk drawer or with the business center owner, security is lessened. In the same way, the security provided by encryption is improved if the customer securely maintains their own encryption keys external to the cloud vendor.

Access Control

When creating a logical data center, a primary concern is access. A single point of access makes access control simpler and monitoring better. If you have a physical data center with multiple doors and windows, securing the data center is more difficult. This is no different in a logical data center.

One method of access control is to federate a customer's existing IAM system with access to customer cloud resources. Depending on the sophistication of the customer's IAM and its ability to properly access cloud resources, this can be an appropriate method. This choice allows the customer to control access more directly. It becomes simpler for the business to oversee who has credentials and what resources those credentials can access. It may also be possible to extend a customer's IAM system to the cloud without federation if cross-connection is unwanted.

Another method to prevent cross-connection between cloud and on-premise resources is to use identity as a service (IDaaS) to provide access to a company's cloud services. Gartner refers to this as SaaS-provided IAM or simply SaaS IAM, which keeps with the three basic cloud service models. An IDaaS solution has the benefit of providing a service that is tailored to cloud resources and services. A SaaS IAM may include a CSP service or an independent SaaS IAM.

Regardless of whether a customer's current IAM system can be used, your first line of defense is to educate your workforce to create secure credentials that are different from credentials they use on other corporate and personal accounts to decrease the impact if an IAM system is compromised.

Physical Design

Physical design is the responsibility of the owner of the cloud data center. This is generally the cloud vendor or CSP. Physical design of a data center for the CSP or cloud vendor is the same as for on-premise data centers with the additional complexity of support for multitenancy.

Location

Physical siting of a data center can limit some disaster concerns. Areas commonly impacted by natural disasters, civil unrest, or similar problems should be avoided. Multiple locations for data centers should also be considered. There is no location immune to all disasters, so locations that have different disaster risk profiles will increase availability of cloud resources. The location should also have stable power/utilities and access to multiple communication paths when possible.

Buy and Hold

A buy-and-hold decision is made when considering movement to a cloud-based data center. The considerations include the sensitivity of the data, regulations on the locations of the data, and availability of a skilled workforce capable of maintaining the on-premise location.

Sometimes a customer has data that must be kept on-premise when there are legal reasons that prevent data from crossing borders, unless the CSP contract can satisfy this requirement. Other customers may have data that is sufficiently sensitive that a breach could place individuals in harms' way. If a business does not have data that should stay on-premise, all of the data could be migrated to the cloud if a business case can be made for this migration.

For many business processes, cloud computing provides significant advantages with respect to cost, availability, and scalability. But cloud computing is not always the best approach for all business processes and data.

Environmental Design

The environmental design like the physical location of a data center is also the responsibility of the CSP or cloud vendor, except for a private cloud. Environmental design can impact the availability of cloud resources. For the cloud customer, review of the basic design of a vendor or CSP's environmental design can be part of the risk analysis of the vendor.

Heating, Ventilation, and Air Conditioning

Electrical service is clearly a concern with any computing installation and has been discussed in other sections. Equally important is the ability to provide appropriate heating, ventilation, and air conditioning (HVAC) support for the facility. This can be partly mitigated by the physical siting of the facility and the construction of the facility. A facility in Canada will need to consider heating and ventilation more carefully than air conditioning (AC). Similarly, a data center near the equator will have greater concerns for AC than heat.

Regardless, HVAC concerns should be considered carefully when choosing the site for an on-premise data center or when reviewing potential cloud vendors. An HVAC failure will affect computing resources just as surely as an electrical or communication disruption. Because of the importance of HVAC in a CSP data center, part of the review of a CSP must include capacity and redundancy of HVAC systems as well as the process for movement between data centers when necessary. This includes the geographic location of the CSP data center. A data center in Hawaii may be more concerned with ventilation and humidity. A data center in Austin, Texas, would need to focus on air conditioning, and a data center in Edmonton, Canada, would focus on heating.

A number of documents can help assess HVAC concerns. A CSP SOC-2 report should have a section on availability and the controls that provide it. This should include

environmental systems. SOC-2 reports can be difficult to obtain and may require a nondisclosure agreement (NDA). Other documents that may assist in determining a CSP's environmental design sufficiency would be business continuity and disaster recovery plans.

Multivendor Pathway Connectivity

Connectivity is critical in a cloud computing environment. The ability for a customer to remain connected to all cloud resources requires planning. The issue is rarely with the Internet, but instead is with the connectivity to the Internet of the customer or of the CSP. The solution often requires the use of multiple ISPs or multiple vendors providing connectivity.

These two issues are connectivity from the customer to the Internet and the connection of the CSP to the Internet. The concerns are similar but independent. A communication failure at the customer's end would impact that single tenant while leaving the vendor and all other tenants unaffected.

The cloud customer should consider multiple paths for communicating with their cloud vendor. In an era of increasingly dispersed workforce, often working from locations separate from the business network, strategies to keep the workforce connected to the Internet and the cloud vendors must be developed and tested. The solution often requires multiple vendors.

The vendor must also develop multiple connectivity strategies. If the vendor cannot access the Internet, it does not matter whether all the tenants can because they will be unable to access their cloud resources. The customer also needs a strategy for times that the vendor is unavailable. How will the customer continue to execute critical business functions? One way is to have multiple vendors for critical functions.

For example, if enabling the dispersed workforce to communicate and collaborate is a critical business function, multiple methods can be developed and tested. A simple example is with online meetings.

ANALYZE RISKS ASSOCIATED WITH CLOUD INFRASTRUCTURE

Any data center or cloud infrastructure has risks, whether it's on-premise or hosted by a third party. Many organizations are moving from an on-premise to a cloud-based infrastructure and must properly consider the risks associated with this move. A cloud-based infrastructure is not less risky; it is differently risky. A move to the cloud must consider

these risks and do a cost-benefit analysis to ensure the cloud is the right move. In the end, after analyzing the risks, many organizations will end up with a hybrid environment with some local infrastructure for some processes/solutions and one or more cloud environments for other processes/solutions.

Risk Assessment and Analysis

The risk analysis of any CSP or cloud solution involves many departments. These include business units, vendor management, privacy, and information security. The new risks with a cloud solution are mostly associated with privacy and information security. There are some key issues when conducting a risk assessment for a CSP or cloud solution. Some major issues will be discussed.

Authentication is a key question. Will the cloud solution provide authentication services, or will the cloud solution be authenticated by the customer. If using the cloud solution's authentication, it is unknown if the cloud provider's authentication is secure. As users have the tendency to reuse passwords, a breach of the cloud service's authentication server may also provide the information necessary to breach your on-premise systems.

If the customer provides their own IAM system, it may be accomplished through a SaaS IAM solution, or through federation with the customer's on-premise IAM manager. Each solution has pros and cons. For example, if using a SaaS IAM system, users are likely to use the same username and password for both the cloud and on-premise IAM systems. However, the SaaS IAM has been designed for the cloud and cloud security needs. If federating the cloud service to the on-premise system, cross-connection between the two environments may create a security concern as well. In either case, user education is an important component to the cloud IAM strategy.

Data security is a concern. How a vendor encrypts data at rest is important. In addition, the method used to transfer data to and from the vendor and between the vendor and any third-party services used by the vendor must be investigated. The data remains the responsibility of the customer even when stored on the vendor's system.

It is important to assess any risk posed by the vendor's policies and processes. These include the vendor's privacy policy, incident response process, cookie policies, information security policy, etc. You are no longer assessing only your organizational policies but also the policies of the organizations with whom you are doing business.

For some key systems, it is important to assess the support provided for incident response. This includes an assessment of logging support, vulnerability scans of the vendor, application vulnerability scans, and external assessments of the vendor being considered or currently used.

Many vendors providing services will have a SOC-2 or SOC-3 report. The preferred report is a SOC-2 Type-2 report. Accessing this report, if possible, will usually require an NDA. However, the assurance this type of report can provide is worth an NDA. Other useful attestations are ISO 270017 (Cloud Security) and ISO 27018 (Privacy) in particular and ISO 2700X certifications in general, FISMA, and FEDRAMP. Carefully read the cover letter provided by the third-party assessor for an understanding of any special circumstances governing the findings in the report.

Cloud Vulnerabilities, Threats, and Attacks

The primary vulnerability in the cloud is that it is an Internet-based model. Anyone with access to the Internet has the potential to attack your CSP, your cloud provider, or you. For example, if you are an international business or are involved in a business that is not well regarded throughout the world (even if your business is legal in the country where you conduct it), you may be the subject of an attack that creates a potential data breach and/or a denial of service (DoS).

The attack on your CSP or cloud vendor may be unrelated to you, so the typical threats you are prepared for may not cover the full spectrum of potential threats. The threat may be targeting the vendor, may be targeting another tenant of the cloud provider, or be related to threats in the geographic location of the cloud data center. These attacks may come from anywhere in the world. You may simply be collateral damage. Regardless, the end result may be a DoS—even if you are not the intended target.

Other risks come from the other tenants. If protections keeping customer data separate fail, your data may be exposed to another tenant. This tenant could be a competitor or other bad actor. Encryption may be the best protection, with the customer managing their own encryption keys. The customer may also consider not storing their most sensitive data in the cloud.

There can be an additional risk from the cloud vendor. Employees of cloud vendors have been known to exfiltrate customer data for their own purposes. Contractual language may provide the only remedy once detected. Prevention becomes the best defense. As with other tenants, encryption with the customer managing their own keys (separate from the cloud) prevents data exposure.

Virtualization Risks

Virtualization is a powerful tool and, as such, has specific risks. The hypervisor is under the control of the CSP in the shared security model. If the hypervisor is compromised, all VMs on the hypervisor may be compromised. So, the task of the CSP protecting the hypervisor is critical.

VM sprawl is also a risk. Members of the workforce may create VMs for projects and forget to close them down when done. VM sprawl increases the attack surface as these unused VMs may not be actively monitored, so malicious use may go unnoticed. The increase in the overall number of VMs can also balloon costs to the organization unexpectedly.

Another risk with VMs is with the data stored in each VM. Sensitive data and data at different trust levels can exist in each VM unless care is taken to manage and monitor what data may be used and where it is stored. The risk associated with VM sprawl and sensitive data storage is a management issue. If not carefully monitored and managed, organizations can easily lose control of the VMs they own, putting data and budgets at risk.

Countermeasure Strategies

There are a number of ways to mitigate risks in the cloud environment. The start of security is with the selection of the CSP. This is the exercise of due diligence in selection. A careful risk assessment and analysis of CSPs will eliminate some of the riskier players in the cloud space.

Using a trusted CSP is a good start. The next step is the design of systems. Security should be designed in at every step. When using cloud services, the configuration of each service can be an important design step to ensure the most secure configuration.

The next countermeasure is encryption. Encryption should be enabled for all data at rest and data in motion. CSPs provide encryption services. Sometimes the only thing missing is the customer enabling encryption. Data in transit must be protected using TLS, IP security (IPSec), VPNs, or another encrypted transmission method. In addition, limiting the ingress/egress points in each cloud service can enhance monitoring.

Each major CSP provides the ability to manage your secure configuration, monitor changes to cloud services, and track usage. For example, AWS provides Inspector, CloudWatch, CloudTrail, and other tools to assist your management of your cloud environment. However, the best monitoring is worthless without regular and consistent review of logs.

DESIGN AND PLAN SECURITY CONTROLS

The risks associated with cloud computing can be mitigated with the proper selection of controls. This is the same approach used in traditional risk management tailored to the risks associated with cloud computing.

Physical and Environmental Protection

The location housing the physical servers in a cloud environment must consider physical and environmental controls. This may be a CSP for a public or other third-party cloud. For a private cloud, it may be the company employing the cloud.

Site location has an impact on both physical and environmental protections. A data center along the waterfront in an area subject to regular hurricanes or in a location with frequent tornados, flooding, or possibly earthquake activity may be ill advised or may lead to a strategy with multiple data centers in different locations to provide redundancy. If each data center has a different risk profile, they provide greater assurance of availability.

Other physical and environmental requirements are typical in all data centers, including physical cloud data centers. These include the ability to restrict physical access, reliable power and other utilities, and the availability of an adequate workforce.

A cloud data center also has significant network capability requirements. All data centers require network capabilities. But, as a CSP may provide services to a large number of customers over a geographical area that can be worldwide, these requirements may be more substantial for the cloud data center than a single tenant data center on-premise. More than one ISP may improve connectivity redundancy in many scenarios.

The customer has no control over the physical siting of a cloud data center except for a private cloud located on-premise. That does not relieve the customer of responsibility in this regard; it simply changes the customer responsibility. To the extent possible, the customer should review the location of cloud data centers and be aware of the cloud vendor's business continuity and disaster recovery plans. The ability of the cloud vendor or CSP to respond to disasters directly affects the ability of the cloud customer to serve their customers.

System and Communication Protection

There are a number of controls available for system and communications protection. One source for controls is NIST Special Publication 800-53: Security and Privacy Controls or Information Systems and Organizations; specifically, Section 3.18 System and Communications Protection (SC). Similar controls can be found with ISO and other control sets. The SC Controls Set includes 51 specific controls. A few of the major controls include the following:

- **Policy and Procedures**: This is a primary control. Policies and procedures (P&P) are a primary foundation to all areas of security. P&P provide the foundation for all security actions by setting the purpose, scope, roles, and responsibilities.

- **Separation of System and User Functionality**: This is a core control. Separation of duties is a fundamental security principle and prevents users from altering and misconfiguring systems and communication processes.
- **Security Function Isolation**: Just as we want to separate the user and system functions, separating security and nonsecurity functions allows for cleaner interfaces and the ability to maintain the security function.
- **Denial-of-Service Protection**: A DOS attack is a primary concern of all communication systems. Preventing a DOS attack involves dealing with bandwidth and capacity issues, detecting attacks, and decreasing pivot risk to prevent one system from attacking another.
- **Boundary Protection**: This includes preventing malicious traffic from entering the network, but preventing malicious traffic from leaving your network, data loss (exfiltration) protection, and other boundary needs of the organization.

These are just a few of the controls in the SC Controls Set. As mentioned, there are 51 potential controls in the NIST publication for System and Communication Protection. It is not expected that an organization will implement all 51 controls. This is more of a collection of potential controls to review and consider.

Cloud computing has a shared security model, and which controls are the responsibility of the CSP and which are the responsibility of the customer should be carefully reviewed and understood. For example, the CSP should have a strategy to keep your business functioning in the event of data center disasters and communication disruptions including DOS attacks.

In addition to these controls mentioned, the customer should assure the potential to communicate in the event of an ISP failure by ensuring alternate methods of communicating with the CSP. This can be through dual ISPs or similar strategies to make communication with the vendor possible when a communication interruption occurs.

Protection of business data remains a critical issue. Controls to protect data in transit must be available and utilized. The primary control for data in motion (transit) is encryption using protocols/tools such as TLS, IPSec, or VPNs. This control is supported by a robust identification and authentication system and the creation of authorization mechanisms that enforce access controls to the system, process, and data.

Virtualization Systems Protection

The controls for VMs and virtualization systems start with secure IAM. On its own, IAM is not sufficient for full protection of your VMs. However, controlling access to VMs is your single most important control. If you get IAM wrong, other controls will be less effective. Privileged access must be strictly limited and should enforce least privilege and separation of duty controls.

Creating standard configurations for VM use in your organization will also help protect them. Variability adds complexity and makes monitoring for unauthorized use more difficult. If the organization uses standard patterns, it is much simpler to identify malicious behavior and to enforce configuration policies.

Configuration guides can provide guidance on the use of cloud services. This will assist workforce members new to the cloud in creating secure solutions. For example, an S3 bucket has many configuration options. Which options are recommended or required and how they should be set can be explained. Cloud tools can then enforce the configuration, prevention, or alerts on changes outside of business guidance.

Other controls exist in the cloud environment to assist with monitoring performance, identifying malicious behavior, and enforcing or identifying variance with policy. Many tools are available to provide these controls—including tools integrated with major CSPs and tools that can be added to a customer's cloud environment.

For example, Amazon provides CloudWatch. This tool can monitor EC2 instances and other data storage systems. CloudWatch can also set alarms, store logs, view graphs, as well as respond to changes in AWS resources. The early alerting provided by this tool can prevent small changes from becoming large problems and may alert you to attacks, misuse of resources, and variations of approved customer configurations.

Azure provides a similar tool in Microsoft Cloud Monitoring (MCM). This tool will monitor Azure applications, analyze log files, and alert customers to potential threats. Like CloudWatch, MCM will monitor utilization, performance, and workloads. In both cases, the software is built into the services each CSP provides. Other CSPs provide similar tools.

Finally, the ability to run multiple VMs from multiple locations must find a balance between security and availability. VM usage can get out of control if not carefully managed. If VM usage is not carefully managed, the attack surface for the organization can actually grow over time.

Identification, Authentication, and Authorization in Cloud Infrastructure

Many organizations will want to extend their current IAM system into the cloud. They may also choose a SaaS IAM and use it for both their cloud environment and their on-premise environment. They may also maintain their legacy IAM on-premise and use a separate SaaS IAM for cloud services. This is really a business decision. An SSO environment for the workforce is generally desirable but may not be possible to achieve.

Another option is to use the vendor's IAM system for each cloud resource. Using a vendor's IAM system introduces a number of risks. First, the security of the vendor's IAM system is unknown. If you are relying on the vendor to provide access to your business data, you should have confidence in their IAM system. For some cloud services, this

is the only option. Careful vetting of the vendor IAM system would be necessary if not providing your own IAM system.

For each possible solution, user education is important. End users often reuse passwords. An employee may use the same username and password on all work-based systems. If a vendor is compromised, then other corporate systems may be similarly compromised. The cloud vendor may not notify you of a compromise of the IAM for an extended period of time, if they are even aware of a compromise. This puts your business systems at further risk.

Audit Mechanisms

It can be more difficult to audit either a CSP or a cloud vendor. A customer will not have broad access to the physical security of cloud data centers or the vendor's networks. This is both due to logistics and due to the presence of other tenants. A CSP may have a number of data centers that are geographically dispersed. Having a customer audit all physical controls is essentially impossible.

Broad access to the vendor network is also a security risk. If you have access to the network, you may intercept privileged information belonging to another vendor, and multitenancy boundaries could be violated. If the vendor provides you with broad access, they would also provide other customers with the same. This would be a security concern.

Vendors may also not share logs. If they are not properly disaggregated, log entries for other tenants could be exposed. However, using CSP tools allows you to monitor your own resources and provide alerts tailored to your needs. For example, AWS CloudWatch monitors operation data and can be used to create logs and generate alerts.

Log Collection

One common problem with logs is the vastness of data collected in logs. If using cloud services it is important to tune the events logged to those that matter. You may set the logging broadly, and as you become familiar with normal operations, you can tune to identify those activities most essential to log. Even without tuning, changes to privileged user accounts should always be logged and alerted. Privileged accounts do change, but it should be authorized, usually in writing. Any changes made using a privileged account should also be logged.

A log aggregator can ingest the logs from all the on-premise and cloud resources for review in the SOC. These logs must be monitored daily with privileged account alerts getting immediate attention. If not reviewed, logs become much less valuable. Log collection will remain valuable for incident response and audit purposes. But without regular and consistent log review, the value of log collection is not fully realized.

Packet Capture

To resolve many security issues and truly know what is going on within your network, packet capture is necessary. However, packet capture on a cloud vendor or CSP using traditional tools like Wireshark is not generally possible. Such packet capture would require access to a vendor's internal network beyond what would be permissible. A vendor must protect the data of all tenants, and allowing one or more tenants to perform packet capture can put other tenants at risk. A vendor may perform packet capture and could make it available for incident response or audit purposes in some circumstances. But this should not be expected and would need to be explicitly agreed to in vendor contracts.

To address this security concern (the lack of packet capture in general), some CSPs provide tools that allow packet capture functionality to some degree. Two examples will be discussed with AWS and Azure. If the customer is using a hybrid cloud, where some portions are on the customer's network or in the customer's data center, packet capture of those segments is possible with traditional tools.

Amazon provides AWS VPC Traffic Monitoring. This tool allows a customer to mirror the traffic of any AWS network interface in a VPC they have created and to capture that traffic for analysis by the customer's security team. This can be done directly in AWS using CloudShark to perform network analysis on the packet capture. This tool creates what is essentially a virtual tap. In this way, the customer can monitor all network traffic through their VPCs.

Microsoft provides a similar capability with Azure Network Watcher. This tool allows packet capture of traffic to and from a customer's VMs. Packet capture can be started in a number of ways. One nice feature for security purposes is the ability to trigger packet capture when certain conditions exist.

The specific tools available and the use of these tools will change over time. For security purposes, the ability to capture packets in the cloud services used by the customer is important. This is an area that should be fully investigated when moving to the cloud.

PLAN DISASTER RECOVERY AND BUSINESS CONTINUITY

The cloud has transformed both disaster recovery (DR) and business continuity (BC) by providing the ability ability to operate in geographically distant locations and by providing greater hardware and data redundancy. All of this leads to lower recovery time objectives (RTOs) and recovery point objectives (RPOs) at price points businesses could not achieve on their own. DR and BC can be planned into the system rather than trying to bolt it on after the fact in largely unsatisfactory ways.

Risks Related to the Cloud Environment

There are risks to cloud computing. The first one that generally comes to mind is that the business no longer owns or has full control over system hardware assets. Computing becomes an operational expense (OPEX) rather than a capital expense (CAPEX). This change in ownership and control makes many people uncomfortable. However, with careful selection of CSPs and the development of SLAs and other contractual agreements, these concerns can be addressed. In addition, the service model used affects the amount of control that is retained. For example, in an IaaS environment, the business retains a large amount of control. It may be determined that the control a business wants is not necessary to ensure business continuity and to address potential disasters. The decrease in cost for DR and BC may also mean greater ability to respond to disasters.

The geographic dispersion of the CSP data centers may also mean that the disaster risk profile may be unknown to the customer and may be very different from their own. For example, the CSP data center may be in an area subject to hurricanes, while the customer is located in an area far from the coast. This risk requires the cloud customer to review the ability of a CSP to address disasters relevant to their locations. One reason it is important to use a large CSP having multiple regions is that the risk profile will be different in each region, providing greater continuity benefits. One CSP data center may be in an area subject to ice storms, and another is in a more temperate location. This allows the CSP to move a customer from one area to another when natural disasters are more likely in one area than another area if the contract permits this.

Downtime is a concern for companies, and Internet-based CSPs are not immune. It is also an issue that can affect a company not using the cloud. If the Internet is down, many companies will be unable to do business. Fault tolerance can be built in with a CSP using availability zones and automatic failover. If ISP connectivity is the concern, the major CSPs provide direct access, such as AWS Direct Connect or Azure ExpressRoute. If fault tolerance is designed into the services a company delivers, localized connectivity failures will only impact those in the region impacted and not customers in other regions.

Compliance can be another risk in the cloud environment. Some regulations have strict rules on where PII, PHI, and other sensitive information can be stored. If a company is not careful, they may violate these restrictions. This is an area that contracts must address. Fortunately, the CSP has many customers, and legal requirements will be common among them. The CSP will need to address compliance concerns for many or all of their customers. Cross-border transfers of data are of particular concern in many areas, such as European Union (EU) countries covered by the General Data Protection Regulation (GDPR).

Another compliance concern may be availability of data. In a disaster scenario, a company may still need access to critical business processes and data even if the systems

are down. Procedures for accessing this data in alternate ways (other than through the CSP) must be planned for and tested regularly. Regardless of the disaster, responsibility for the data remains with the customer, so careful selection of CSPs and alternative access methods will need to be addressed.

There is also some concern that multitenancy boundaries can be breached, APIs can be exploited, and the hypervisor may become compromised in a disaster scenario if safeguards fail. Disasters can affect more than availability. They can impact integrity and confidentiality as well. These issues can affect the ability of a business to function. Each of these must be addressed. A BC and DR plan cannot simply point to the cloud for every potential disruption. Instead, comprehensive plans are still necessary and must be developed and tested.

One risk-reducing feature that makes the cloud particularly beneficial for BC and DR is that the Internet and geographic dispersion means that the customer processes are available, widely leading to a highly available architecture. Both highly available because the network is global and highly available because all of a major CSP's availability zones will not be directly affected by the same disaster. Disasters such as storms, terrorist attacks, and other infrastructure failures are usually geographically localized.

Business Requirements

Business-critical systems require more recoverability than is often possible with local resources and corporate data centers without expending enormous resources. The cloud environment provides options to support high availability, scalable computing, and reliable data retention and data storage. Three ways we measure the business capabilities are RTO or how long are you down, RPO or how much data may you lose, and recovery service level (RSL), which measures how much computing power (0 to 100 percent) is needed for production systems during a disaster. This does not usually include compute needs for development, test, or other environments. These are usually nonessential environments during a disaster.

Recovery Time Objective

RTO is the amount of time in which a business process must be restored to a specific service level. Missing the RTO can lead to contractual violations and business consequences. The RTO is usually measured in minutes or hours. However, for some business processes, the RTO can be days. One of the authors of the book once worked for a company that had a 72-hour RTO for the network to be backed up and the data backup to be restored. Clearly, this customer did not have the transactional load of an Amazon or eBay type of business and could catch up in a reasonable period of time. The company had a manual process during that 72 hours, and then manually entered the transactions after the backup was restored.

At the end of the day, RTO is a business decision and not an IT decision. The role of IT is to support the business with options and costs. The business needs full information on options and costs to make the best business decision. Once the decision is made, the role of IT is to implement the decision and to make every effort to meet the business RTO.

Recovery Point Objective

The RPO is a measure of the amount of data that the business is willing to lose if a disaster or other system stoppage occurs. Once the business makes this decision, you have the appropriate RPO, and a backup frequency can be selected. If the business is willing to risk losing 24 hours of data, the systems need to be backed up daily. If the business determines that a potential loss of one hour of transactions is acceptable, an hourly backup is needed. If the business maintains paper copies of transactions between backups, these transactions may be recoverable with time and effort. For an Amazon or eBay type of business, a manual process is not a practical solution. The number of transactions per minute are simply too large for these major online businesses. Other methods to maintain data continuity in the event of a disaster will be necessary.

The RPO is usually measured by the amount of time during which the business chooses to risk losing transactions. The RPO could be the number of transactions that may be lost during a disaster. Either way, the RPO is tightly coupled with system backup strategies and/or redundancy strategies. With a seamless failover in a cloud environment, the RPO can essentially be zero (either zero seconds of lost transactions or zero transactions lost) for all but the most catastrophic events. For organizations with a high level of transactions, the ability to seamlessly failover can be essential. In a cloud environment, maintaining a copy of data in another region (a mirror, for example) can support an RPO of near zero. If multiple regions fail at the same time—a truly catastrophic event—the ability to maintain business processes may not be the primary consideration.

The customer is responsible for determining how to recover in the case of a disaster. The customer can use backups, availability zone, load balancers and other techniques to provide disaster recovery. A CSP can help support recovery objectives by not allowing a data center to have two availability zones. Otherwise, if you are using two zones and they are in the same data center, a single disaster can affect both availability zones. Cloud services can also provide the monitoring needed for this high availability. Other major CSPs provide similar capabilities. It is important to note that while the major CSPs provide these capabilities, a customer must choose to use and properly configure these capabilities. They do not come automatically and often have additional cost.

Recovery Service Level

RSL measures the compute resources needed to keep production environments running during a disaster. This measure (0 to 100 percent) gives an indication of the amount of computing used by production environments when compared to development, test, and other environments that can be shut down during a disaster. During a disaster, the focus is on keeping key production systems and business processes running until the disaster is largely resolved and other activities can restart to pre-disaster levels.

Business Continuity/Disaster Recovery Strategy

The difference between a business continuity plan (BCP) and disaster recovery plan (DRP) will drive the strategy. A BCP may be executed following an event that falls short of being a disaster. A BCP will always be initiated following a disaster. The BCP's purpose is to keep business running after an event, especially a disaster. The BCP may use different locations or use different systems or processes.

For example, a BCP may be initiated following an unscheduled disruption of utility services. If the power utility must shut down a facility for an unscheduled upgrade, the BCP can guide how the business will proceed. One company the author worked for had a failing electrical system at the point utilities were provided to the business. The company scheduled a repair with the utility provider at a convenient time. However, the system suffered a complete failure prior to the scheduled maintenance, leading to the invocation of the BCP. The BCP's purpose is to keep business running in another location or using different systems or processes until such time that the power returns. In this chapter, the BCP will be part of the response to a disaster.

While a BCP keeps business operations going after a disaster, the DRP works on returning business operations to normal. For example, a tornado may make a facility unusable. The BCP will move operations to alternate sites. In a cloud environment, this may mean bringing additional availability zones online to ensure high availability. It may also involve relocated key personnel and the use of cold/warm/hot sites for the workforce.

The purpose of the DRP will be to restore operation in the destroyed building or new permanent facilities. Once systems are restored to their original facilities or new facilities, the disaster recovery effort is over. Restoration of systems includes normal operations, new development, regular maintenance, and other support functions.

Another way of looking at this is the BCP is concerned with critical business processes and keeping them running. The DRP is focused on returning operations to normal, which includes infrastructure and facilities. In a cloud environment, normal operations will be defined as returning to pre-disaster levels of function and not simply continuation of critical business functions.

The cloud supports the BCP/DRP strategy. The high availability of cloud services provides strong business continuity and can serve as a core part of the BCP strategy. In effect, the cloud allows the business to continue regardless of disasters affecting the customer's business facilities. The high availability of cloud services also impacts DRP strategy. As the cloud can provide resilient services at attractive price points, a company may focus more on resilient services that survive a disaster rather than processes to recover from disasters.

Creation, Implementation, and Testing of Plan

There are three parts to any successful BCP/DRP. Each is vital to the success of the organization. These include the creation, implementation, and ongoing maintenance and testing of the plans.

Because of their inherent differences, BCPs have different responsibilities within the business. Sometimes the BCP is seen as part of the DRP. However, it is really an independent plan that supports the DRP. If the business does not continue to function, the responsibility of the BCP, there is no reason to have a DRP as there will be nothing to return to normal. The BCP supports the DRP.

The plans should be developed with knowledge of the other as they must work together. These plans will include many of the same members of the workforce. But they are separate plans.

Plan Creation

The first step in any comprehensive BCP is to do a business impact analysis (BIA). The BIA identifies the impact of process/system disruption and helps determine time-sensitive activities and process dependencies. The BIA identifies critical business processes and their supporting systems, data, and infrastructure. The BIA is essential to determining requirements and resources for achieving the RTO and RPO necessary for business continuity.

The systems can be grouped in a number of ways. One way is to identify critical processes, important processes and support processes, and the systems and data that support these processes. Critical business processes are those that impact the continued existence of the company. Email, for example, is rarely a critical business process. For a company like Amazon, inventory, purchasing, and payment may be critical processes. The selection of a critical process is a business decision.

Along with the selection of critical processes is a prioritization of systems. If you can bring up only one system at a time, which is first? Some processes will not work until other processes are back online. In that case, they are prioritized after the processes

they depend on. There is no point in bringing a process online if it cannot function until other processes are available. Because of this, identification and authentication services are often among the first processes to restore. All critical processes must be prioritized over other processes.

After critical processes are other important processes. These are processes that are a value-add and can have an impact on the business profitability. Once again, these must have a prioritization or ordering.

After the important processes are online, other useful and beneficial processes are brought back online. It is possible that a business decision may be made that chooses to not restore some processes until restoration of normal operations, if they are not critical to the business.

Processes based on legacy systems can be difficult to return to pre-disaster configurations. The technical debt of continuing to run legacy systems unless essential to the business may make them expendable. If there are systems that may be in this category, it may be worthwhile to consider replacing and retiring them prior to an emergency. Legacy systems often have legacy hardware and software requirements that are not easily restored after a disaster. Legacy systems may also have information security risks that are best avoided. If the legacy system supports a critical business process, the need to replace the legacy system may be urgent.

One advantage to a cloud-based BCP/DRP is the expectation of a modern and up-to-date infrastructure. While a legacy system could potentially be hosted in the cloud, the move to a cloud position may provide the opportunity to modernize.

The creation of the DRP often follows the BCP. Knowing the critical business processes and the supporting infrastructure provides a roadmap to returning the system to pre-disaster operations.

Both the BCP and DRP should be created considering a range of potential disasters. History is a good starting point for environmental disasters. Government agencies such as FEMA and private/public partnerships like InfraGard can also provide guidance on the likely disasters that need to be considered in any geographic location.

The CSP is also a source for BCP/DRP planning. Recommendations and solutions for surviving specific business disruptions and returning to normal operations will be available from the larger CSPs, such as Amazon, Google, IBM, etc.

The creation and implementation of a DRP or BCP is a lengthy process. It is simplified by having the system criticality and prioritization determined prior to beginning plan development. The identification, criticality, and prioritization of business processes, systems, and data are a necessary first step to creating a complete and functional plan. If this prerequisite step is omitted, the creation of the plan can be disruptive. Often, every

business unit and all process owners view their processes, systems, and data as important, if not essential. It is important for senior leadership to make these decisions in advance so that each group knows their place within the plan for bringing the business back online.

BCP/DRP Implementation

As part of the creation of a BCP/DRP, the identification of key personnel is important. The corporate sponsor provides key resources for creating the plan. Implementation of the BCP/DRP can include identifying alternate facilities, contracts for services, and training of key personnel.

The BCP/DRP identifies critical processes. To implement the plan, a customer may implement critical services in the cloud to take advantage of multiple availability zones, automatic failover, and even direct connection to the CSP. These choices come with a cost. The cost of high availability in the cloud is generally less than a company trying to achieve high availability on their own. In addition, the cost of building resiliency may be far less than the cost of business interruption. By identifying the critical business processes, a business can also avoid the cost of implementing high availability for noncritical systems. In many conversations on availability with business process owners, it has become clear that everyone wants high availability. But, once the costs associated with that decision are understood, few continue to believe that they have a requirement for high availability.

Critical business processes can be supported through component/data center redundancy and may run as an active-active configuration. This allows near instantaneous continuation of critical processes. Important but noncritical business processes may use the less expensive active-passive configuration. This allows a more rapid restoration of services when a specific region or zone becomes unavailable. Less important business processes may be run in a single region or zone, may take more time to return to service, and may operate at a lower cost.

Methods to provide high availability and resiliency continue to evolve. The ability to automate monitoring and deployment through orchestration and other methods supports high availability in the cloud. New tools and methods will continue to be developed that should lead to an ever-resilient cloud environment at attractive prices.

Implementing a BCP/DRP will also set the schedule for training and testing. Generally, plans should be reviewed and tested at least annually.

BCP/DRP Testing

A BCP and DRP should be tested at least annually. This test should involve the key personnel needed for all disasters. In addition, many scenarios will involve members of the workforce not directly involved in the creation and implementation of the plans.

There are some basic scenarios that apply to most if not all businesses. It is not necessary to test all of them each year. But a robust plan should test all likely scenarios over time. A well tested plan will function even for an unexpected disaster. Common disaster scenarios include the following:

- Data breach
- Data loss
- Power outage or loss of other utilities
- Network failure
- Environmental (e.g. fire, flooding, tornado, hurricane, or earthquake)
- Civil unrest or terrorism
- Pandemics

The plan should test likely scenarios and can be tested in a number of ways. The maturity of the plan and the people implementing the plan will determine the type of testing that takes place. There are a variety of standard test methods.

Tests should be both scheduled and a surprise. Particularly for new plans and organization immature in the BCP/DCP space, a scheduled test ensures that key personnel are available and will begin the maturation process. Tests that have the potential of being very disruptive should also be scheduled to minimize disruption.

Eventually, some tests should be unexpected. Surprise tests do not mean unscheduled. Instead, only high-level approval and a few key individuals are aware of an upcoming test so that the organization can be tested in more realistic ways. The high-level approval is essential in case some amount of unexpected business disruption occurs. Executive approval includes a review of potential disruption and benefits so that a business decision can be made by those responsible for the business. Key personnel who are part of the test planning can be positioned in advance of the test in order to monitor performance metrics.

The simplest test method is the tabletop. A tabletop is usually performed in a conference room or other location around the "table." Key personnel are presented with a scenario and then work through the plan verbally. This is usually the first step for a new plan. It identifies missing pieces in the plan or steps that need greater detail. The next step may be a walk-through where key personnel move to the appropriate locations and verbally verify the steps, sometimes even performing some parts of the plan. While developing a plan, regular tabletops and walk-throughs can help flesh out a more robust plan. A tabletop or walk-through can also be useful for new members of the team to assist them in identifying their responsibilities in the plan.

A more substantial test is a simulation. Like a fire drill or a shelter-in-place activity, a disaster is simulated. The plan is exercised while normal operations continue. More

robust and detailed than a walk-through, those performing certain steps are involved in simulating their response to a disaster.

The next level of testing is a parallel test. Care must be taken not to disrupt normal business operations if possible. In a parallel test, key personnel and workforce members perform the steps needed in case of a disaster. More than simulating the steps, they actually perform the steps to ensure that they can accomplish the critical business processes if the existing systems were disrupted by a disaster. It is parallel because the critical systems continue to run, and some or all of the data is also run at the alternate site. It is then possible to compare the results of the alternate methods to the critical systems to determine any gaps in capabilities.

The most robust level of testing is a full cutover test. In this test, the disaster is simulated in full. The primary system is disconnected, and the business attempts to run critical functions. This is a high-risk test, as the critical functions may fail. Only the most mature organizations and plans should attempt a full cutover test.

SUMMARY

Moving to the cloud is a business decision. Moving to the cloud changes capital expense to operational expense and can add high availability and resiliency at a price that is attractive to businesses of all types. It also provides capabilities not available to all businesses. To achieve these benefits, a customer must have an understanding of the key infrastructure pieces of a cloud environment. The customer must also understand the shared security responsibility between the CSP and the customer. Finally, a customer must understand how to properly configure and use the cloud resources they purchase to ensure appropriate controls are in place to secure the process and data.

DOMAIN 4

CCSP®

Cloud Application Security

CLOUD APPLICATION SECURITY CAN be a neglected part of cloud security. Often, security focuses only on the controls associated with identity and access management (IAM), networking, servers, and other infrastructure components. But if the application software running on these components is insecure, then the entire enterprise is insecure. This chapter will discuss the processes needed to secure the software through the application development lifecycle.

ADVOCATE TRAINING AND AWARENESS FOR APPLICATION SECURITY

Secure software development begins with the development of a culture of security and the implementation of a secure software development lifecycle (SSDLC). Without an appropriate SSDLC, the development of secure applications is posed for failure. A secure culture is not developed through occasional efforts but through regular, deliberate action. The three parts identified by the Software Assurance Forum for Excellence in Code (SAFECode) are executive support and engagement, program design and implementation, and program sustainment and measurement. Training and awareness is an important step in developing a security culture and the development of secure cloud applications sometimes called DevSecOps.

Cloud Development Basics

According to the Cloud Security Alliance (CSA) and SAFECode, the development of collective responsibility is essential and challenging when building a safety conscious application development program. This effort can be broken down into these three parts:

- **Security by design:** This implies that security is part of every step in the process. Security is not a bolt-on after application release or in response to a security flaw, but part of the process from application feasibility to retirement.
- **Shared security responsibility:** The idea is that security is the responsibility of everyone from the most junior member of the team to senior management. No one says that security is not their responsibility. In fact, security is the result of individual responsibility and trust.
- **Security as a business objective:** Security is not something in addition to what we do. Instead, secure computing is what we do. Security is part of all business objectives.

Each of these basic concepts requires a clear understanding of organizational culture, security vulnerabilities, and organizational vision and goals.

Common Pitfalls

The first pitfall is not getting senior management's initial and ongoing support. Efforts to instill a security culture are unlikely to be successful unless senior leadership supports these efforts—through their approval and adherence to security policy as well as funding and staffing security initiatives that are part of organizational goals.

The next pitfall is failing to understand organizational culture. Even companies in the same industry have very different cultures. Building and sustaining a security culture requires careful planning that recognizes an organization's culture and then designs program-level efforts that work within that culture. In addition, program-level efforts must be reviewed regularly to ensure that they remain aligned with business objectives and cultural realities.

Another pitfall is staffing. The size of an organization may limit the number of security experts. These experts may then serve as resources to multiple development processes rather than as a member on one team. If the organization is large, development teams may use security experts as resources or may include security experts as team members. In either situation, it can be challenging to ensure the development of security-aware applications will be consistent across the organization.

Not having a framework for secure software development will also cause difficulties. A mature organization will have a well-defined process that integrates security as part of each step. The lack of a framework or process leads to ad hoc security and does not support a security-aware organization.

Finally, in times of budget constraints, the training budget is often trimmed of nonessential training. It is easy to cut training without immediate impact. However, a security-conscious organization will find ways to continue security awareness training and secure computing practices. There are many options that can reduce training costs while retaining many training options.

Potential security training options include subscriptions to online security courses, as well as online asynchronous and synchronous training. Each of these options eliminates travel expenses. This can trim the budget significantly, while maintaining training operations. One other option is to bring the trainer to the work location. If a large number of people need the same training, bringing the trainer to you rather than sending the people to training can be cost-effective. These options can keep training active while trimming the budget.

Common Cloud Vulnerabilities

Common cloud vulnerabilities include data breaches, data integrity, insecure Application Programming Interfaces (APIs), and denial-of-service (DoS) attacks. Each of these is present because of the extensive use of networks as part of cloud computing. Anytime an application is accessed or transmits data over the Internet, it is doing this in a hostile environment.

Two organizations that provide information on security threats are the CSA and the Open Web Application Security Project (OWASP). Both publish annual research on top threats in cloud computing and related technologies. Both organizations are worth regular review. The top four risks from each organization are provided as an example of their work.

CSA Top Threats to Cloud Computing

For the past several years, the CSA (`cloudsecurityalliance.org`) has published the top threats to cloud computing. The number of threats varies, and the publications are entertainingly named. In 2016 to 2018, it was the Treacherous 12. From 2019 to the present, it is the Egregious 11. Regardless of the name, these lists are a great security resource and should be reviewed each year. Here is an example of the top four threats identified in 2020's Egregious 11:

- Data breaches
- Misconfiguration and inadequate change control
- Lack of cloud security architecture and strategy
- Insufficient identity, credential access, and key management

None of these threats should be surprising. Protection of data, which may be put at risk through misconfiguration, poor access control, and other failures, tops the list. The top items on the list also suggests that cloud security needs to become more mature in many organizations.

OWASP Top 10

The OWASP Top 10 (`owasp.org`) is a periodically updated list and features the top threats in web application security. The last version is from 2017, with the previous version in 2013. The top 10 security risks lead to a variety of issues that include data breach, data integrity, and DoS. Essentially, each item of the CIA triad can be affected by one or more of these risks. The following were the top four in 2017:

- Injection flaws, including SQL, NoSQL, OS, and LDAP injection
- Broken authentication
- Sensitive data exposure
- XML external entities

The most disturbing thing about this list is that the items on the list rarely change. Sometimes all that changes is the order. For example, numbers 1 and 2 in 2013 and numbers 1 and 2 in 2017 are the same. In four years, the top risks were unchanged. Number 3 in 2013 moved to number 7 in 2017, while number 6 became number 3.

We know what the risks are. Organizations like the CSA and OWASP publish what the risks are. As expected, the overlap between the lists is high. Now, it is the responsibility of security professionals to address these risks.

DESCRIBE THE SECURE SOFTWARE DEVELOPMENT LIFECYCLE PROCESS

The software development lifecycle (SDLC) has been understood for several years. The SSDLC enhances the SDLC process. The SDLC has several phases. These steps are Requirements. Design, Development, Testing, Deployment, and Operations and Maintenance (O&M). In the SSDLC, these phases are enhanced to include specific security-focused steps to allow security by design. There are many resources for implementing an SSDLC. These include the Microsoft SDL and the NIST SP 800-160, Systems Security Engineering. Two resources that this section will discuss are the NIST Secure Software Development Framework and the OWASP Software Assurance Maturity Model (SAMM). This section will also explore business requirements, phases, and methodologies related to the SSDLC.

NIST Secure Software Development Framework

Similar to the popular NIST Cybersecurity Framework (CSF), the NIST Secure Software Development Framework (SSDF) defines and describes secure software development practices (`nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04232020.pdf`).

This framework is useful for developing secure traditional IT systems, as well as Industrial Control Systems (ICS), IoT systems, and cyber-physical systems (CPS).

The SSDF can be adapted to existing SDLCs, supports the use of modern software development techniques such as agile, and leverages guidelines and standards from other organizations.

The SSDF is organized into these four groups:

- **Prepare the Organization:** This includes people, processes, and technology.
- **Protect the Software:** Protect it from tampering and unauthorized access.
- **Produce Well-Secured Software:** This means software with minimal vulnerabilities.
- **Respond to Vulnerabilities:** This includes preventing them in future releases.

OWASP Software Assurance Security Model

The OWASP SAMM can also be implemented into an existing SDLC. The OWASP (`owaspsamm.org`) consists of these four steps:

1. Assess the current security posture.
2. Define a strategy (security target).
3. Implement a roadmap.
4. Offer prescriptive advice on activity implementation.

SAMM provides assessment tools and guidance to improve an organization's security posture and supports the development of secure software and systems.

Business Requirements

Mature software development shops utilize an SDLC because it saves money and supports repeatable, quality software development. Studies have been conducted that show the later in the development phase issues are found, the more expensive it is to fix those issues. Adding security to the SDLC benefits secure software development. While the SSDLC adds some up-front cost to the development of new software applications and the modification of existing applications, identifying security vulnerabilities early will lower overall development costs. The expected return is software solutions that are more secure against attack, reducing the exposure of sensitive business and customer data. Bolting on security after the coding or deployment phases simply increases the cost of that security while limiting its effectiveness.

However, an SSDLC is fully successful only if the integration of security into an organization's existing SDLC is required for all development efforts. Only when secure development is a requirement of a business will security by design occur consistently.

Phases and Methodologies

There are many models for SDLCs, from linear and sequential approaches such as waterfall to interactive and incremental approaches such as spiral, agile, and most recently Development and Operations (DevOps), which increase the speed and frequency of deployment.

There are several primary stages to any SDLC, including an SSDLC. These are as follows:

- **Requirements:** This phase includes all parts of the planning and can include feasibility studies, gathering business and security requirements, and some high-level design of the software solution.
- **Design:** The design step starts as a high-level design and gets increasingly detailed as this stage progresses. The design must include the security requirements identified in the requirements phase. Design can also include the design of testing requirements including test cases and acceptance thresholds to ensure all business and security requirements are met. Test cases should be tied to specific requirements identified in the requirements stage. Tests should be developed to meet and implement all requirements.
- **Development:** The coding phase is the creation of the software components as well as the integration or build of the entire solution. Unit testing is generally part of the coding phase.
- **Testing:** This phase is the initial testing of the solution built as well as more focused testing that occurs to validate the solution before the final phase.
- **Deployment:** Deployment is the work associated with the initial release of the software. Part of the effort in this stage is to ensure that default configurations conform to security requirements and best practices, including configuration of APIs and IAM. These steps reduce the risk of credential compromise. These steps also protect the processes for account creation, maintenance, and deletion. Finally, the deployment of a securely developed application service will often use multiple cloud services and APIs. It is a good practice to consider each service that is approved for use by a business and create a standard configuration guide for each of those services to ensure a secure configuration for each service that adheres to company standards.
- **O&M:** This is often the longest phase as it encompasses everything that happens after the release of a software solution. This stage includes any operational, monitoring, and maintenance needs of the solution.

It is easy to see how each phase of the SDLC can be divided into ever finer phases. But these steps capture the flow of the work. At each step, security requirements are an

important part of the overall solution. Following these steps leads to a consistent and repeatable process. These phases are supported in older development methodologies such as the waterfall and spiral models as well as more modern methodologies such as agile and DevSecOps.

APPLY THE SECURE SOFTWARE DEVELOPMENT LIFECYCLE

SSDLC is a collection of best practices focused on adding security to the standard SDLC. Applying an SSDLC process requires dedicated effort at each phase of the SDLC, from requirements gathering to deployment and maintenance. An SSDLC requires a change of mindset by the development teams, focusing on security at each phase of the project instead of just focusing on functionality.

Identifying security issues early helps reduce the risk of identifying security vulnerabilities late in the development process and minimizes the impact when these are found.

Having an SSDLC is beneficial only if it is implemented and used consistently and does not eliminate traditional security tests, such as penetration tests. Instead, it empowers developers to build secure applications from the very beginning of a project. Additionally, some standards and regulations, such as the General Data Protection Regulation (GDPR), Payment Card Industry (PCI), ISO27001, and others require security (data safeguards) to be incorporated in the development process.

Avoid Common Vulnerabilities During Development

Common vulnerabilities and risks are listed in many places. Perhaps the most common list of risks in web-based development are the OWASP Top 10. These are updated regularly to ensure the most common risks are known to developers. By learning this list, developers can actively design and develop systems that have reduced vulnerabilities. The latest OWASP Top 10 2017 list is discussed in the "Common Cloud Vulnerabilities" section. The first five vulnerabilities are the following:

- **Injection:** SQL, NoSQL, LDAP, and OS injection errors allow the attacker to run unintended commands and access data without authorization.
- **Broken authentication:** When authentication and session management are incorrectly implemented, it leads to compromise of passwords, keys, and tokens.
- **Sensitive data exposure:** Web applications and APIs may poorly protect sensitive data, both at rest and in transit when encryption is not used.

- **XML external entities:** Evaluation of external entities in XML documents may disclose internal files, allow remote code execution, and lead to DoS attacks.
- **Broken access control:** Poor access control may allow authenticated users to view unauthorized and sensitive data, execute unauthorized functions, change access rights, and so on.

This OWASP Top 10 list does not change significantly at every release and has a few common themes. This includes misconfiguration, improper use of tools, and poor input validation. Each of these can then lead to data breaches, account takeover, and disclosure of sensitive data. Keeping this list in mind can lead to the identification of security vulnerabilities and the development of code that has fewer vulnerabilities. This in turn will lead to the development of secure software and systems.

Cloud-Specific Risks

There are a number of additional risks that apply specifically to cloud computing. Some specific information on cloud-specific risks or security issues is provided in the Egregious 11 from the CSA (`cloudsecurityalliance.org`).The Egregious 11 includes the following:

1. Data breaches
2. Misconfiguration and inadequate change control
3. Lack of cloud security architecture and strategy
4. Insufficient identity, credential, access and key management
5. Account hijacking
6. Insider threat
7. Insecure interfaces and APIs
8. Weak control plane
9. Metastructure and applistructure failures
10. Limited cloud usage visibility
11. Abuse and nefarious use of cloud services

Cloud-specific vulnerabilities identified by the CSA's Egregious 11 can lead to violations of one of more of the CIA triad requirements of confidentiality, integrity, and availability. The risk of data breach, compromise, and exposure of sensitive data can lead to violations of regulations such as HIPAA and may lead to fines and other penalties. At the least, a well-publicized data breach can lead to a loss of reputation and potentially a loss of business and market share. Loss of data integrity and availability have similar consequences.

Risks beyond those discussed by the CSA will exist for specific companies at different levels of security maturity, in various industries, and in different geographic locations. Each organization must know their specific additional requirements.

The first few CSA Egregious 11 security issues will provide the framework for our discussion. Each organization should review all the CSA security issues and geographic, industry, and regulatory requirements as relevant.

CSA Security Issue 1: Data Breaches

Confidential and restricted data is a key target of hackers. It can lead to loss of trade secrets and other intellectual property (IP), strategic processes, and customer data. This can harm the organization's reputation, lead to regulatory fines, and expose the company to legal and contractual liability.

In the requirements gathering phase, the data to be protected and the users that need access are considered. Involvement of a security architect can be beneficial at this stage.

In the design and development phases, specific strategies to protect the data can be developed, such as data encryption and methods of authentication and authorization. Using existing infrastructure to provide these services leverages existing company assets.

During testing, each requirement, including security requirements, must be exercised. Traditional testing can be extended with penetration testing and vulnerability scanning in our effort to develop secure software and systems.

During deployment, the default setting in your system must be set for maximum security. The default state for any deployment must be a secure state. During O&M, organizations may modify the default settings and introduce vulnerabilities. This potential must be monitored and reported to ensure the system remains in a secure state and any changes are made with full knowledge of potential risks. In addition, monitoring for new vulnerabilities, threats, and the resulting risks is part of the O&M process. It is also during O&M that incident response and regular IAM reviews become part of the process.

CSA Security Issue 2: Misconfiguration and Inadequate Change Control

At deployment and during O&M, securely developed software can become insecure. The areas to watch are data storage and transmission, excessive permissions, default credentials and settings, and standard security controls being disabled.

Storage will be considered at the requirements, design, development, and test stages. But at deployment and O&M, the actual storage, software, and systems must be configured and monitored. The default configuration must be secure. Any changes to the default configuration must be monitored and the user provided with warnings when changed. A cloud environment should be configured to enforce these configurations.

This issue is made even more serious by the reduced visibility of a cloud environment. When all components are on-premise, the infrastructure is known and is easily reconfigured, monitored, and maintained. In the cloud, some of this visibility and control transfers to the cloud service provider (CSP). Extra steps must be taken to use the cloud services to maintain as much visibility and control as possible.

Many developers are new to the cloud environment, and even experienced developers can make errors in configuration that expose systems and data to unauthorized users. This exposure, when occurring in a cloud environment, can lead to worldwide exposure of sensitive data. Continuous automated scanning of cloud resources can prevent or mitigate changes in real time.

Change control is a key part of preventing configuration changes that make a system less secure. Security expertise on the change control committee is an important step in the secure software process. In addition to introducing new vulnerabilities, change control prevents the reintroduction of previously corrected vulnerabilities.

CSA Security Issue 3: Lack of Cloud Security Architecture and Strategy

Companies are moving to the cloud with increasing speed. Often the migration outpaces an organization's ability and preparation for cloud deployment. It is advisable to pause and ensure that the organization has a security-focused architecture in place and has developed a cloud strategy before cloud development and migration begins. If the strategy and architecture are determined after cloud development or migration, similar to adding security to an application that is already deployed, the ability to secure software and systems and bring them into compliance will be more expensive and less effective.

The size of the organization does not affect this issue. Organizations of all sizes need a cloud security architecture and strategy to migrate, develop, and operate in the cloud securely and safely. Not doing so can have severe operational and financial repercussions. This control exists prior to phases of the SSDLC and will impact each phase by requiring all phases conform to the organization's security architecture and strategy policies.

CSA Security Issue 4: Insufficient Identity, Credential, Access, and Key Management

This issue includes several concerns such as a scalable IAM system, the use of multifactor authentication, protection and rotation of cryptographic keys, protection of credentials, and enforcement of password policies.

Addressing this issue begins at the requirements phase. Here, the data and IAM considerations are first addressed. A requirement for strong passwords, cryptographic protections, multifactor authentication, and so on should be part of the requirements. The design phase continues this process by ensuring that all requirements are part of the

design. Development then implements all parts of the secure design, and the test phase ensures that all requirements are met, designed fully, and implemented correctly.

During deployment and O&M, it is important to consider that a cloud solution may have thousands of individuals capable of establishing accounts through the on-demand self-service model over time. In a business environment, the turnover will require the creation of new accounts and the disabling and eventual deletion of other accounts. The use of the cloud service must be carefully configured and monitored to prevent escalation of privileges and access to sensitive data by unauthorized individuals.

Quality Assurance

In a traditional development process, quality assurance (QA) was the testing phase. Separate from the development team, QA was the check before deployment that ensured that requirements were met and that the code was bug-free. QA was also part of the configuration management process, testing patches prior to deployment. In many organizations, that remains the role of QA.

In the modern DevOps or DevSecOps, QA is part of the process and is embedded within the DevOps team. QA at this point isn't about the application service developed. Instead, QA is centered around service delivery of the application service developed. QA occurs at each phase, ensuring continuous improvement and quality tracking. Testing (often automated testing) is tied to both functional and security requirements developed in the requirements phase and specified by the security architecture and strategy.

For QA to be effective, further functional and requirements testing should be performed. QA should be involved in load testing, performance testing, stress testing, and vulnerability management. To be more effective, testing can be automated.

In some DevOps environments, it is the QA team that pushes out the code. So, that team is responsible for ensuring that the code delivered to the customer through the cloud environment is quality code and is both defect-free and secure. The QA team then takes a pivotal role in developing secure software and systems. Doing this requires that they have visibility into all phases of the process.

Threat Modeling

Threat modeling has five major steps that must be integrated. This can be greatly enhanced by using a threat model, such as STRIDE. These steps should be performed early in the SSDLC (in the Requirements phase) and updated throughout the SSDLC. These steps are as follows:

1. **Define security requirements.** If requirements are known, clear, and focused, the remaining steps are simpler to implement. Vague and broad requirements are difficult to work to and achieve.

2. **Create application overview.** In this step, we look at the application architecture, application characteristics, and the users of the system to more easily identify threats.
3. **Identify threats.** When we understand the characteristics of our application and our security requirements, identification of threats is more easily accomplished.
4. **Mitigate threats.** Once we identify threats, we can start identifying controls that will aid in threat mitigation.
5. **Validate threat mitigation.** It is important to monitor and review controls to ensure they adequately address identified threats and decrease threat below the organization's risk appetite.

Throughout the five phases of the SSDLC, these five steps are refined. One way to identify security requirements is through the use of a threat model, such as the STRIDE model. STRIDE is an acronym for Spoofing, Tampering, Repudiation, Information disclosure, DoS, and Elevation of privilege. These are common threat categories to consider in this process. See Table 4.1.

TABLE 4.1 The STRIDE Model

LETTER	THREAT	PROPERTY VIOLATED	DEFINITION
S	Spoofing identity	Authentication	Pretending to be someone or something else
T	Tampering with data	Integrity	Modifying data or memory
R	Repudiation	Nonrepudiation	Claiming to not have done something
I	Information disclosure	Confidentiality	Providing information to unauthorized parties
D	Denial of service	Availability	Exhausting resources needed for a service
E	Elevation of privilege	Authorization	Allowing someone to perform unauthorized tasks

Software Configuration Management and Versioning

The purpose of software configuration management (SCM) and versioning is to manage software assets. This can be challenging as software is almost always developed in a team setting. These teams may be geographically dispersed and may be working in multiple time zones. Multiple people may be making changes to both the config and source code files. Managing all of this is essential to ensure that changes are current and accurate.

SCM is important as it can make rolling back changes possible. For example, if a deployed version has a significant flaw, it is possible to redeploy an earlier version of the software while addressing this flaw in development. Versioning can also maintain copies of the configuration and software for deployment to different machines and operating system versions. There may also be different versions for different customers, different countries, different regulations, and so on.

A key role of SCM occurs at deployment and during O&M release management (updates or patches). Prior to formal SCM tools, there were instances where the wrong version of software was released. This can be a costly error, exposing the organization to reputational, operational, regulatory, and contractual issues.

Additionally, SCM allows for configuration audits and reviews by providing the artifacts necessary to ensure processes are followed. Compliance with requirements may be related to an organization's policies, regulatory requirements, or contractual obligations. The ability to quickly and accurately perform audits is an important role of SCM.

Configuration management and versioning is a common practice in all software development environments and is aided through the use of a configuration management database (CMDB). It may be possible to federate the CMDB with both on-premises and cloud-based solutions. The federated CMDB synchronizes across multiple systems and can store the database on premises or in the cloud.

Another approach is to have a single DB, but each system's application services are saved separately. However, a single corporate process is used to manage this process.

It is possible to do both approaches in a hybrid manner. Each company will have to decide how to manage configuration management. It is as important to perform configuration management on premise as it is for software development in the cloud.

APPLY CLOUD SOFTWARE ASSURANCE AND VALIDATION

Software assurance defines the level to which software is free from vulnerabilities and operates as intended. This assurance is a level of confidence as the absence of errors cannot be proven. We can also test compliance to requirements. But the possibility exists that a software solution will exhibit unintended behaviors as well. We use methods like an SSDLC to design security into the software solution from the beginning and to implement testing to ensure security goals are met and the software functions as designed and according to requirements.

In the following sections, functional testing is described first. These tests ensure that the software meets functional requirements by doing what it was designed to do. Security testing is discussed next. This testing validates the security requirements of the software

and operates within the architecture security strategy of the organization. The goal of this testing is to determine if the software is secure. Again, this is a confidence level as to the security of the software system developed prior to deployment.

Functional Testing

Functional testing is used to test that the functional specifications of the system, linked to system requirements, are met. The execution of a robust set of test cases, linked to functional requirements, will create a level of confidence that the software operates as intended.

There are many ways to test software. However, there are some common tests that are consistently used leading up to functional testing. The primary categories of testing that lead up to functional testing are unit testing, integration testing, and usability testing.

- **Unit testing:** This is testing by a developer on modules being developed as part of a larger system. All paths through the module need to be tested.
- **Integration testing:** As modules are combined, integration testing ensures that the modules work together. As additional modules are added, we get ever closer to functional testing.
- **Usability testing:** This testing uses customers in a production-like environment to get feedback on the interaction between the user and the system.

As the modules are tested and then are integrated (and tested) and the user's feedback is obtained and incorporated into the system, we get to the point where we are ready to perform functional testing on the entire system. When conducting functional testing, there are important considerations that include the following:

- **Testing must be realistic:** Many development shops have Dev, Test, Stage, and Prod environments. These environments are called *lanes* in some organizations. The Dev or development environment can be set up to suit the developers' needs. However, for the greatest assurance, the Test and Stage environments should be set up as closely as possible to the Prod environment. In many cases, the Test and Stage environments will have live data or older copies of live data to ensure functional requirements are met. Once development is complete, the application moves to the Test environment for testing. Upon successful testing, the application will move to the Stage environment for configuration management and potentially additional testing. Once the next release cycle occurs, the software in the Stage environment moves into production. Any environment with live data (current or older) must be protected as well as the Prod environment to prevent data loss.
- **Acceptance:** Testing must be sufficient to guarantee that the application service meets the requirements of the customer and the organization (sometimes they are the same). This means that testing must be designed to exercise all requirements.

- **Bug free:** Testing must be sufficient to have reasonable assurance that there are no major bugs in the software. If there are any remaining bugs, they need to be small, rare, and inconsequential.

Once the system passes functional testing, it is ready to follow a QA process to deploy the system. Once deployed, enhancements and bugs will lead to further development. It is important to use the SSDLC process for all further development. This leads to another form of testing, which is regression testing.

Regression testing is done during the maintenance phase of software development to ensure that modifications to the software application (for example, to fix bugs or enhance the software) do not reduce current functionality, add new vulnerabilities, or reintroduce previous bugs and vulnerabilities that have been fixed.

Testing is the way we obtain the confidence or assurance that our software is free of vulnerabilities and functions as required and designed. Testing allows for quality assurance. Adequate testing is important. This requires that adequate time be allocated to conduct testing to find, fix, and test again in an iterative process. In addition, automated testing tools can improve the efficiency and completeness of testing. In a continuous integration/continuous deployment (CI/CD) environment, automated testing becomes a required feature.

Security Testing Methodologies

Security testing is conducted to provide assurance that the organization's security strategy and architecture are followed and that all security requirements have been met. Testing is usually one of three types:

- **White-box testing:** Tests the internal structures of the software. This requires access to the software. Static application security testing (SAST) is a form of white-box testing.
- **Gray-box testing:** Tests a system with limited information about the application. The tester does not have access to the code but will have knowledge of things such as algorithms and architectures. It is primarily used in integration and penetration testing.
- **Black-box testing:** Tests a system with no knowledge of the code, algorithms, or architecture. Dynamic Application Security Testing (DAST) is a form of black-box testing.

There are common tests used in security testing. These happen at different stages of the development process. These include the following:

- **Static Application Security Testing (SAST):** This test is able to do a static analysis of source code. Source code is available for internally developed software systems. Static testing will not find all vulnerabilities. SAST is a good initial test to eliminate common vulnerabilities that can be found in this manner. As the code is known, this is a form of white-box testing. SAST tests can be run prior

to deployment once a testable amount of code is available and can be run throughout the remaining steps in the SSDLC.

- **Dynamic Application Security Testing (DAST):** This tool is used primarily as a web application vulnerability scanner. It is a form of black-box testing. DAST is known for having poor risk coverage, unclear reporting, and slow performance. So, it should not be the only testing tool used. When used, it should be used as early in the development process as practical. Once an application is deployed, a DAST is not your best choice.
- **Interactive Application Security Testing (IAST):** IAST is newer than SAST and DAST and provides a gray-box testing approach. IAST provides an agent within an application and performs real-time analysis of real-time traffic application performance, detecting potential security issues. It can also be used to analyze code as well as runtime behavior, HTTP/HTTPS traffic, frameworks, components, and back-end connections. IAST can be used at every phase of the SSDLC.

Another tool often discussed with SAST, DAST, and IAST is Runtime Application Self-Protection (RASP). RASP is less a test and more of a security tool. RASP runs on a server and works whenever the application is running. RASP intercepts all calls to and from the application and validates all data requests. The application can be wrapped in RASP and provides additional system security. In a layered defense, this is an additional layer and should not replace secure development and testing.

Security testing provides assurance that the software has been developed securely. While SAST and DAST may have their place, if you could use only one security testing tool, IAST would be the best choice. However, new methodologies are always being developed, and care should be taken to consider and use new tools as they become stable and available to improve your security assurance testing.

USE VERIFIED SECURE SOFTWARE

The only type of software that a security conscious organization should use is software that has been verified as secure. Verification generally comes from a third party that will perform testing on software and validate that it has no discernable vulnerabilities. When there are no verified secure options, a customer must do their own due diligence to ensure security. In this section, we will discuss some major components of secure software.

Approved Application Programming Interfaces

API development and deployment in custom applications requires the same SSDLC as other software development projects. The requirements can specify the methods used in the API to monitor access. API access monitoring is often done through authentication or keys.

If not securely developed, a custom API can be vulnerable, leading to the compromise of the system it fronts. Deployment should focus on API configuration and automate the monitoring of that configuration.

APIs can control access to software or application services in a SaaS solution, to back-end services in a PaaS solution, or even to computing, storage, and other infrastructure components in an IaaS solution. For each of these, an approved API is important to ensure security to the system components with which we are interacting. In addition, when possible, enforcing the use of APIs to create a minimum number of methods for accessing an application service simplifies monitoring and protection of these application services.

A CSP or other vendor will provide an API or services that allow the use of an API. It is important to use the APIs as defined and to configure them carefully. An organization can develop approved configurations for APIs used commonly within that customer's organization, and policy can enforce the use of those standard configurations.

In addition, vulnerability scanning of APIs can test the adherence to standards and can provide assurance that known vulnerabilities do not exist. It is impossible of course to ensure that unknown vulnerabilities such as zero-day vulnerabilities do not exist.

Supply-Chain Management

There are two parts to supply chain management. First, it can refer to the needs of a CSP and potentially the customer to use third parties to provide services. For example, the CSP may rely on other vendors to provide services used by their customers. Management of this supply chain is a form of vendor risk management. When creating relationships with these vendors, both operational and security concerns should be addressed.

Additionally, traditional supply chain management is moving increasingly to the cloud. As the life of many organizations is tightly coupled with their supply chain management, the risks of cloud computing are important to consider. However, as the supply chain becomes increasingly global and sourcing of goods requiring primary and secondary sources, cloud computing increases the reach and benefit of supply chain management. Cloud computing can optimize infrastructure to provide operational and financial benefits.

In recent years, supply chain risk has become an increasingly common theme. An earthquake in one country will affect the availability of a resource in another. Large-scale cyberattacks like the SolarWinds hack can impact critical infrastructure. A global pandemic leads to shortages worldwide. Even a simple problem, like a ship blocking the Suez Canal, can interrupt the global supply in unexpected ways.

So, supply-chain management has a complex set of risks that include cloud computing. Cloud computing cannot overcome the risks of a pandemic or a ship running aground in the Suez Canal. But, securing the supply-chain management software in the cloud and securely connecting vendors globally through cloud services reduces the IT-related risk.

Third-Party Software Management

The use of third-party software adds additional risk. A third party may have limited access to your systems but will often have direct access to some portion of your data. If this is sensitive data, a careful review is necessary and should involve the vendor management office (VMO) if your organization has one. Specific language regarding security should be part of every vendor contract.

The typical issues that are addressed include the following:

- Where in the cloud is the software running? Is this on a well-known CSP, or does the provider use their own cloud service?
- Is the data encrypted at rest and in transit, and what encryption technology is used?
- How is access management handled?
- What event logging can you receive?
- What auditing options exist?

In addition to basic security questions, a review of the third-party SOC-2 report, recent vulnerability scans and penetration tests, and security and privacy policies will provide an assessment of the security maturity of the organization and whether you should entrust them with your sensitive data. While you may delegate some processes, you cannot delegate responsibility for your data.

Another risk that will occur with some third-party vendors is fourth-party risk. *Fourth party* refers to a third party's third party, such as if your vendor uses a separate, independent vendor to provide you a service. For example, when a SaaS solution uses an independent CSP for some of their storage needs, your risks include the risks associated with the SaaS solution as well as any additional risks created by the CSP they use. In essence, your infrastructure (computing, storage, and so on) is hosted on the Internet, and any services used by these parties increase the perimeter of the risk that must be considered.

There are advantages to using third-party services. A third party may supply software to address in-house needs when there is no in-house expertise. Third-party solutions can also provide cost and tax advantages. However, these resources must be understood and managed just as much as in-house application solutions.

Validated Open-Source Software

All software, including open-source software (OSS), must be validated in a business environment. Some argue that open-source software is more secure because the source code is available to review, and many eyes are upon it. However, large and complex solutions are not simple to review. So, validation through sandbox testing, vulnerability scans, and third-party verification is required.

The common belief that there is less risk from OSS because it is inexpensive shows an incomplete understanding of risk. Risk is about the asset that is being protected and not about the cost of the software used. In most business, the data is a primary asset. Losing your data to inexpensive software does not lessen the cost associated with the data breach and exfiltration. OSS must follow the same risk-based steps of verification that commercial software undergoes.

When using OSS, there are steps you can take to validate this resource. The easiest method is to use well-known and well-supported products in the OSS space. For example, there are many versions of Linux available for use, but not all versions are equal. A well-supported version with a proven track record is preferable to a less known and less supported version.

One method for validation that can be used would be to perform code analysis on the open-source code. The advantage of OSS is that the code is available. SAST tools find security vulnerabilities in the code. Static analysis of an application will only get you some value, but will not get you all of the way there. IAST can be used in conjunction with SAST. An agent runs on the application server and analyzes traffic and execution flow to provide real-time detection of security issues.

These methods can also be utilized together. You can use a well-known and well-supported OSS, perform SAST to reveal initial vulnerabilities, and then implement IAST for real-time detection of additional security issues.

COMPREHEND THE SPECIFICS OF CLOUD APPLICATION ARCHITECTURE

The traditional application architecture is a three-tier client-server module. In cloud computing, we have some additional choices. These include microservices, cloud native, serverless, and cloud-based architectures.

The microservice application designs a complex architecture as a collection of services and data. This follows an old software engineering principle of cohesion. Each microservice performs a single business function. Each microservice uses the appropriate language and tools for development and can be combined as needed to provide a complex system. Microservices application architectures are a natural for containers running on virtual machines or physical machines, so they are well suited to the cloud. Container management is managed through services like Kubernetes (K8s) or Docker.

A cloud native architecture is for applications deploying to the cloud. The applications exploit the cloud computing delivery model and can be run in any type of cloud (public, private, community, or hybrid) and can assist in getting applications to market.

A cloud native architecture can deploy in a DevOps, and a CI/CD process and can use microservices and containers.

Serverless environments use an event-driven architecture. Events trigger and communicate between decoupled services. Because they are serverless, these architectures scale well using a REpresentational State Transfer (REST) API or event triggers.

Cloud-based architectures are well suited to building and deploying web applications. Using an API Gateway, a secure API is a front door to web applications providing access to data and business logic.

Considering these cloud architectures, there are a number of tools and services that can support the security needs of both new software solutions as well as legacy devices. These services provide enhanced security and deal directly with common security issues. These services may be supplemental services that protect certain parts of the application architecture, encryption services that protect data at rest or in motion to ensure confidentiality of the data, methods to test securely, and services that tie all these services, web services, and application services together.

Supplemental Security Components

Supplemental security components provide service to your cloud environment to solve specific security concerns. For example, database monitoring works to ensure the integrity of our databases, while XML firewalls support application services through XML messages. Each supplemental security service will describe the problem solved by this service.

Web Application Firewall

A web application firewall (WAF) protects HTTP/HTTPS applications from common attacks. Usually, a WAF protects an Internet-facing application, but it can also be used internally on an intranet. The WAF can be a hardware device, a software device, or both. The WAF monitors GET and POST requests. The requests are then compared to configured rules. A WAF may look for specific signatures or apply heuristics.

By filtering HTTP/HTTPS traffic, a WAF helps protect against SQL injection, cross-site scripting (XSS) and cross-site forgery, and other attacks. The WAF specifically addresses attacks on application services and external sources.

The WAF differs from an Intrusion Detection System (IDS), which monitors specific traffic patterns on a network. A WAF works at the application level and focuses on specific web application traffic and is often employed as a proxy, with one or more websites or web applications protected behind the WAF. The CSPs and third-party vendors provide many WAF options.

Database Activity Monitoring

Database activity monitoring (DAM) refers to a set of tools that supports the identification and reporting of fraudulent or suspicious behavior in the databases used by your application services. This real-time monitoring may use but is independent of native DBMS

auditing and logging tools. DAM analyzes and reports on suspicious activity and alerts on anomalies. In addition to application monitoring and protecting from web attacks, DAM also provides privileged user monitoring.

Like other services, there are third-party vendors providing DAM services and CSPs providing services that are configured for their database offerings. These tools do more than monitor database usage. They can monitor privileged use, data discovery, data classification, and other database needs. Some DAM toolsets also provide assistance in compliance to contractual and regulatory requirements such as PCI DSS, HIPAA, and GDPR.

Like all services provided by third parties and CSPs, the tools change over time, adding breadth and functionality, and sometimes even the service name. This toolset is designed to provide cloud native data activity monitoring and works in major CSPs. DAM tools can be deployed inline to monitor traffic like a WAF or IDS. They can also be used as a detective tool to scan log data and identify issues.

Extensible Markup Language Firewalls

While beneficial for application integration, security is a concern when deploying Extensible Markup Language (XML) services. XML provides a standard way to do data interchange between applications. XML can also be used to perform XML external entity processing, which is one of the OWASP Top 10.

XML firewalls work at the application layer to protect XML-based applications and APIs over HTTP, HTTPS, and other messaging protocols. XML messaging and APIs between these services are an area of security concern. An XML firewall can solve this problem. All service requests pass though the XML application firewall. As an XML firewall must inspect traffic, they are generally implemented as proxies and stand in front of the web application server. An XML firewall can implement complex security rules through Extensible Stylesheet Language Transformations (XSLT).

A number of common web-based attacks can be launched through XML. These attacks include SQL injection and cross-site scripting (XSS). This is done through misuse of input fields and can be prevented through data validation and verification on input fields and schema verification. The use of an XML firewall can support the security needs of an application but should not be a substitute for developing secure software and systems. Instead, it should be an added level of protection. An XML firewall can benefit legacy code that was not designed with security. This becomes a compensating control until the development and deployment of a secure system. By dropping inappropriate traffic, it can also decrease the likelihood of DoS attacks. Firewall as a Service is one of the many cloud services provided by vendors for the major CSPs.

Application Programming Interface Gateway

An API gateway allows traffic to your application backend services. The services provided by an API gateway include rate limiting, access logging, and authorization enforcement. For secure computing, there should be limited doors into your application service. For example, some SaaS providers provide an API to access your account from your PC, Apple device, or Android device. These gateways control the way a user accesses and interacts with the SaaS solution and allow securing and monitoring this traffic. API gateways provide authentication and key validation services that control who may access the service, ensuring confidentiality of data.

Amazon Web Services (AWS) provides this service through Amazon API Gateway. AWS provides both RESTful for serverless computing and WebSocket APIs for real-time communication. Google Cloud provides an API gateway for REST APIs to provide serverless computing and a consistent and scalable interface. Azure API management provides a REST-based API for legacy systems. Essentially, all CSPs provide an API to allow the customer to monitor and control access to data and services to their workforce, partners, and customers in a way that provides a layer of protection and access control.

Cryptography

Cryptography is a key technology for encrypting and protecting sensitive data in the cloud. Encryption is the first line of defense for data confidentiality. Encryption requires the use of keys. Management of the keys is a critical security vulnerability. There are three primary parts of encryption: data at rest, data in motion, and key management.

Encryption for data at rest is a standard practice for all sensitive information. Many CSP services have encryption as a standard option. Some CSPs provide standard APIs to allow encryption to be added to any CSP service or customer application. CSPs generally also provide encryption options for their storage and database services. Naturally, encryption tools are standard in all large CSPs.

Data-in-motion encryption is accomplished in standard ways to include TLS, HTTPS, and VPNs. The ability to work with standard secure data transmission methods is provided by all mature CSPs. In addition, larger CSPs can accommodate data sharing between software solutions, even in multiple regions, without ever transiting the public Internet. However, this promise of not transiting the Internet does not necessarily mean that data is transiting a trusted network. Even when staying off the Internet, encrypted data transmission should be expected.

With all of this encryption, there are encryption keys to manage. The major CSPs all provide key management services (KMS) as do some third-party vendors. The first decision is who will manage the keys. Solutions exist to allow the vendor to manage the keys. However, for sensitive data, it is preferable that the customer use a commercial KMS rather than the vendor in order to improve key security. This provides a separation of duties between the service managing the encrypted data and the service managing encryption keys.

Sandboxing

A sandbox is an environment with restricted connectivity and restrictions or functionality. Sandboxes provide two primary security benefits. These include sandboxes for developers and sandboxes for secure execution of code. Both provide benefits to the cloud customer.

Sandboxes for developers allow the development of code and, more importantly, the testing of code in an isolated environment. A sandbox is a temporary environment to build and test code. Any problems generated will not impact anything outside of the sandbox. This is also a suitable environment for developers new to cloud application development to try and test services provided by cloud providers. This is a safe way to learn and test cloud tools and services.

Secure evaluation of code provides several options, each of which is to protect you from malicious code or poorly developed or misconfigured code. The purpose of the sandbox is to allow the execution and evaluation of code without impacting the rest of the customer (or CSP's) environment. This is especially valuable when executing code that may be malware. The customer can evaluate the effect of the code and can determine if it is dangerous to run in the customer's nonsandbox environment.

Application Virtualization and Orchestration

Virtualization of infrastructure is an essential core technology for cloud computing, and application virtualization through containerization is part of several cloud application architectures such as microservices. For example, a customer may run a virtualized desktop or container on their personal laptop, a virtual machine, or a mobile device for business applications. The entire device is not virtualized, but those applications run separately from the host OS, providing secure access to the business application and data.

Application virtualization through containers allows an application to be packaged with all dependencies together. The container can then be subject to strict configuration management, patch management, and repeatable build processes. These containers can segregate organizational software and data from the user's device in a BYOD environment. This may allow a remote wipe capability the employer can use to remove corporate access, applications, and data without impacting the remainder of an employee's personal device. Application virtualization can make the trend for BYOD more secure for the employer and safer for the employee.

Containerization orchestration is commonly done through K8s. There are many containerization technologies, but K8s was built for container orchestration. Containers provide security to the organization through configuration management, patching, and dependency update support.

Containers do not come without a cost. The containerization technology used must be configured and patched. If the technology is compromised, so is the container.

The orchestration software must also be patched and configured following secure computing policy and standards set by the organization.

Cloud orchestration allows a customer to manage their cloud resources centrally in an efficient and cost-effective manner. This is especially important in a multicloud environment. Management of the complexity of corporate cloud needs will only increase as the move to the cloud accelerates. Orchestration allows the automation of workflows and management of accounts, and the deployment of cloud applications, containerized applications, and services in a way that manages cost and enforces corporate policy in the cloud.

DESIGN APPROPRIATE IDENTITY AND ACCESS MANAGEMENT SOLUTIONS

Identity and access management solutions encompass a range of activities. The IAM solution properly begins with the provisioning of users. Provisioning includes the creation of credentials as well as authorization for the systems to which a user needs access. IAM solutions also include the ongoing maintenance of access, such as adding and deleting access to systems as user roles change within the organization. Finally, an IAM solution includes the deprovisioning of a user when an account is no longer needed.

IAM solutions also perform the identification and validation of users and provide access to systems. Once an IAM solution identifies and authenticates a user, it can perform authorization. Authorization determines which resources an authenticated user can access. Authentication can use role-based, attribute-based, or any other form of authentication. In addition, many IAM providers also support password management, while other solutions can be integrated with a customer's current on-premise authentication systems.

There are a variety of options for cloud-based IAM. This section will discuss the methods of authentication available and the security issues that exist with each of these.

A variety of protocols are available to IAM solutions, to include OAuth2, SAML, LDAP, etc. Each protocol provides different capabilities. For example, OAUth2 was developed to provide authorization with web applications and mobile devices. SAML is an XML-based authentication service well-suited for authentication between the identity provider and a service provider. LDAP is designed to work well with directory services, like Active Directory (AD). Which protocol is used varies by authentication provider and use and is a choice determined by each business.

Federated Identity

Federated identity is related to single sign-on (SSO). In a federated identity, a particular digital ID allows access across multiple systems. With federated identity, a digital ID can access applications across CSPs (a multicloud) and on-premise resources.

Federation also allows SSO for systems across multiple organizations. These may be subsidiaries of the same company, multiple CSPs, cloud vendors, or multiple organizations.

There are security challenges with federated identifications. The primary issue is that once someone compromises a digital ID on one system, the ID is compromised on all systems that are federated. This makes the security of IAM systems equally important on all federated systems. The security of the IAM system for each cloud provider and cloud vendor as well as the on-premise IAM system must be equally protected and monitored for malicious use.

The drawback to a federated identification is similar to SSO. When one system is compromised, all federated systems are compromised. The complicating factor is that the level of trust between systems of multiple organizations may not be the same as between systems of a single organization. Additionally, no organization controls the system protection across all organizations or the evaluation of all users in each organization.

Identity Providers

Identity providers can be CSP services or the services of a third party. In a multicloud environment, a third-party solution may be the best choice as it provides a single solution to the customer. Identity providers do not replace other cloud security tools, such as a Cloud Access Security Broker (CASB), but work together to provide a layered security defense. In this case, the IAM ensures that users are authenticated and their access to cloud resources is authorized. The CASB monitors and protects those cloud resources.

The major CSPs provide IAM services. These include Azure Active Directory, AWS Identity and Access Management, and Google Cloud Identity and Access Management. There are also many good third-party choices for identity management. This would be considered identity as a service (IDaaS). IDaaS is a specialized SaaS. This can be especially advantageous in a multicloud environment.

Single Sign-On

Single sign-on allows access to all authorized systems that fall under a single IAM system after authenticating once. The user can move freely between systems without having to reauthenticate each time. It is important to monitor all access within an SSO system. If a digital ID is compromised, all systems within that IAM environment that the user has access to are also compromised.

The advantage is that an SSO limits the number of credentials that must be changed when compromised and allows for simpler central monitoring and access maintenance. When each user has multiple credentials, monitoring is more complex, and when a user's access must be modified or removed, it is easy to miss one or more sets of credentials. Access and use of all identities must be monitored within an SSO environment for malicious activity. In addition, an organization must have processes in place to react to malicious activity.

SSO is extended through the use of federation. Federation allows SSO for resources across multiple IAM systems, such as multiple cloud and on-premise environments. This increases the risk caused by a compromised identity as the number of systems that may be compromised is greater. This increases substantially the importance of monitoring for and responding to malicious activity.

Multifactor Authentication

Multifactor authentication (MFA) adds a level of security to standard user IDs and passwords when used appropriately. Two-factor authentication (2FA) and MFA are often used interchangeably. There is a subtle difference. 2FA refers to using two factors. MFA may use two or more factors. 2FA is a subset of MFA.

Authentication factors are as follows:

- **Something you know:** This includes answers to security questions and identification of previously selected photos/pictures, PINs, and passwords.
- **Something you have:** Examples include a hardware token, smartphone, or a card, such as a debit card or smart card.
- **Something you are:** This category generally refers to biometrics. This is generally fingerprint, facial recognition, or iris scans. Of the three, these are the most challenging to do reliably and at a reasonable cost.

Often what is described as MFA is simply multiple instances of the same factor. An example is when a system requires a user ID, a password, and the answer to a security question—these are all things you know. This is single-factor authentication masquerading as MFA.

One use of MFA is to limit the potential damage caused by a compromised account in an SSO or federated system. If, when moving from one federated system to another or one federated resource to another, a token is required from something you have, then the ease of SSO or federation is decreased slightly while increasing security. The method chosen will be a balanced decision based on available options, the cost of available options, the value of the asset being protected, and the organization's risk tolerance.

Another approach would be to simply require a token from a device periodically throughout the day; for example, every two hours. This limits the time a compromised identity can be used. The additional burden of authenticating two to three times a day may be an acceptable price to pay to limit the damage of a compromised identity in a federated or SSO environment.

Cloud Access Security Broker

A CASB is an important addition to cloud security. A CASB sits between the cloud application or server and the customer. This service, which may be software or hardware-based,

monitors activity, enforces corporate security policies, and mitigates security events through identification, notification, and prevention. A part of a layered security strategy, a CASB is not meant to replace firewalls, IDS/IPS systems, or similar security systems. Instead, a CASB is meant to enhance the security provided by these other devices.

A CASB that provides security must be in the path of user activity with the application. These CASBs may be agent-based or agentless. In either case, they are inline. There are also out-of-band CASBs that receive all cloud traffic for security analysis. These are somewhat analogous to an IPS and an IDS. The inline CASB enforces policy. The API-based CASB monitors for violation and analyzes cloud traffic broadly. In this section, we discuss inline CASBs.

Agent-based CASBs face friction from users when a bring your own device (BYOD) policy is in place. When the customer owns the device, it may not be possible to install an agent on the devices. Even with customer permission, the wide variety of mobile devices may be more extensive than available agents. So, there may not be a compatible agent. In addition, an agent-based CASB may severely impact the performance of a customer's mobile device. When the organization owns the devices, an agent-based approach can be more effective.

An agentless CASB uses an API on the cloud resources to inspect all traffic to and from that resource to perform its responsibilities. This allows access to all cloud resources to be monitored regardless of endpoint ownership. It always can limit inspection to organizational data, eliminating some privacy concerns. The other advantage is that agentless CASBs can be quickly deployed and more easily maintained.

SUMMARY

Cloud application security requires many things working together. The first step is to choose to develop secure applications by policy. All other decisions flow from this one decision. Once this decision is made, an SSDLC must be adopted with training on secure software development and the tools and processes used in this approach to application development. An SSDLC must be implemented, with development leading to assurance and validation activities to ensure that developed solutions are secure. The next step is to develop mechanisms to ensure verified software is distributed securely and that modified software can be detected to prevent the insertion of malicious code in your securely developed software solutions. Finally, the software solutions must be deployed using the architected secure solution. This approach implements secure API gateways, XML firewalls, web application firewalls, DAM, IAM, CASB, and other tools as necessary. These tools monitor the applications in use, with the ability to respond to potentially malicious behavior and potentially prevent malicious use.

DOMAIN 5

CCSP®

Cloud Security Operations

CLOUD SECURITY OPERATIONS COMPRISE a mix of old and new practices for understanding, mitigating, and monitoring security risks in an organization's cloud environments. Old practices include standard activities that apply to legacy or on-premises IT, such as legal and regulatory compliance management, as well as novel activities like orchestrating cloud infrastructure by writing virtual machine (VM) definitions instead of physically installing new hardware and software.

Key to cloud security operations are the two main roles in cloud computing: the cloud service provider (CSP) and the cloud consumer. The CSP and consumer share responsibilities for securely operating and using the cloud, respectively, and require clearly defined, agreed-upon objectives documented in contracts and service level agreements (SLAs).

IMPLEMENT AND BUILD PHYSICAL AND LOGICAL INFRASTRUCTURE FOR CLOUD ENVIRONMENT

It is important to bear in mind that many aspects of secure cloud operations will be handled by the CSP and therefore may be largely invisible to the cloud consumers. As security professionals, it is critical to understand the importance of both the provider and consumer roles; particularly important is including adequate oversight of the CSP

in third-party security risk management activities. From the CSP's perspective, proper isolation controls are essential due to the multitenant nature of the cloud, as well as appropriate capacity, redundancy, and resiliency to ensure the cloud service meets the availability requirements that customers demand.

Hardware-Specific Security Configuration Requirements

In public cloud deployments, hardware configuration will be handled by the CSP rather than the cloud consumer. Obviously, private and community clouds will require the security practitioner to properly configure and secure hardware, and in some cases, public clouds may offer a virtual private cloud (VPC) option, where some of these elements are configurable by the consumer. The rules for *hardening*, or securely configuring, systems in cloud environments are the same as they are for on-prem systems, though the methods may be different.

There are several targets for hardening hardware, including the following:

- **Basic Input Output System (BIOS)**: The BIOS is a form of firmware, that is to say, software that controls and provides interaction capabilities for low-level hardware components like processors. It is typically stored in read-only memory, though upgradeable or "flashable" BIOS does exist. BIOS is responsible for critical tasks such as verifying the correct configuration and presence of expected hardware, performing self-testing for functionality, and handing control of the system over to the hypervisor or OS once initialized.
  - BIOS is crucial to secure booting operations, where the hardware and firmware configuration of a system is verified before the OS or apps are allowed to run. To achieve this, checks are performed to compare the installed hardware and firmware against a known good configuration, often using fingerprint techniques like hashing or digitally signed software.
  - Without proper security, the BIOS could be tricked into accepting unwanted hardware that is added to a system or modifications of the BIOS software, allowing an attacker to gain highly privileged access to systems and data in use.
  - Major areas of concern for BIOS security include the following: authentication of updates utilizing digital signatures from the BIOS Original Equipment Manufacturer (OEM), integrity protections for the BIOS software to ensure no unwanted changes are made, and non-bypass features such as hardware or software that block unapproved access.
  - NIST SP 800-147B, *BIOS Protection Guidelines for Servers*, provides guidance and details on implementing secure BIOS configurations for servers. It details four security features for BIOS, including authenticated update, secure local updates, firmware integrity, and non-bypassability requirements.

- **TPM**: The Trusted Platform Module (TPM) is a dedicated module included in a computing system with specialized capabilities for cryptographic functions, sometimes referred to as a *cryptographic coprocessor*. It has dedicated components including a processor, persistent storage memory, and volatile memory for performing cryptographic processing, and it is used to support cryptographic functions and enable trust in a computing system.
  - The TPM typically provides a number of services related to cryptography, including random or pseudorandom number generators, asymmetric key generation, and hash generators. TPMs are often used for highly secure storage of limited data as well, such as the cryptographic keys used in full disk encryption solutions like Microsoft BitLocker.
  - TPMs are often used to form roots of trust, since they are highly specialized and secured. For example, a hash of hardware component versions installed in a system can be relied upon if it is digitally signed by the TPM. This hash can be compared with a hash of the current system state to determine if any changes have been made, allowing for the verification of a system's hardware integrity.
  - TPMs are implemented in a variety of form factors. Dedicated hardware may be used to provide tamper resistance or tamper evidence. Integrated and firmware TPMs may be included as part of another chipset or run in a dedicated trusted execution environment of a specific chip but are generally less resistant to tampering.
  - Virtual TPMs are part of the hypervisor and provided to VMs running on a virtualization platform. Since this is a software solution, it cannot implement the same level of tamper resistance as a hardware-based TPM. The hypervisor is responsible for providing an isolated or sandboxed environment where the TPM executes, which is separate from the software running inside a particular VM. This isolated environment must implement robust access controls to block inappropriate access to the TPM.
  - A definition of TPM is provided in ISO/IEC 11889, which specifies how various cryptographic techniques and architectural elements are to be implemented. It consists of four parts including an overview and architecture of the TPM, design principles, commands, and supporting routines (code). An industry consortium known as the Trusted Computing Group publishes a specification for TPMs, which currently stands at version 2.0.

Storage Controllers

Storage controllers are hardware implemented to, as their name implies, control storage devices. This may involve a number of functions including access control, assembly of data to fulfill a request (for example, reconstructing a file that has been broken into

multiple blocks and stored across disks), and providing users or applications with an interface to the storage services. Several standards exist including iSCSI and Fibre Channel/Fibre Channel over Ethernet (FCoE). These are storage area network (SAN) technologies that create dedicated networks for data storage and retrieval. Security concerns for SANs are much the same as for regular network services, including proper access control, encryption of data in transit or at rest, and adequate isolation/segmentation to address both availability and confidentiality.

Network Configuration

For public cloud consumers, the majority of network configuration is likely to happen in a software-defined network (SDN) management console rather than via hardware-based network device configuration. It is the responsibility of the CSP to manage the underlying physical hardware including network controller devices such as switches and network interface cards (NICs). Concerns for this physical hardware include the following:

- Providing adequate physical and environmental security for ingress and egress points to the facility such as point of presence (POP) rooms where ISP connectivity is established. This includes physical access control devices such as locks, adequate and redundant electrical power, and appropriate environmental controls like cooling to deal with the heat generated by these devices.
- Designing for resiliency and redundancy of network infrastructure is essential, due to the heavy reliance on virtualization in cloud environments. A single network cable being unplugged or severed can lead to thousands of hosts losing connectivity. The obvious advantage to virtualization is that a redundant network connection is also shared, so dual NICs and ISP connections can be easily shared by hundreds or even thousands of hosts. Performing single point of failure (SPOF) analysis is crucial for the CSP to ensure they meet requirements for availability.
- Establishing flexible and scalable network architecture is key. In physical terms, creating a LAN requires physically interconnecting users to a single switch. In SDNs, this same functionality needs to be available without physically changing any device connections. The CSP's network must provide sufficient physical connectivity and support software-defined virtual local area networks (VLANs) through the use of 802.1Q VLAN tags.
- CSPs must provide appropriate security capabilities for the virtualized and distributed environment. Many traditional security tools rely on techniques like monitoring traffic via a switched port analyzer (SPAN) or mirror port, which sends a copy of all traffic to the monitoring device. Inter-VM network communication on a virtualized host generally does not leave the physical hardware and traverse network connections and could therefore bypass monitoring tools that

rely on mirroring. VM-specific tools exist to overcome this limitation, and secure configuration of hypervisors must be enforced to ensure they provide appropriate monitoring capabilities.

- CSPs may have business drivers to allow for customer-managed network security controls. Many public clouds offer a virtual private cloud (VPC), which is essentially a sandboxed area within the larger public cloud dedicated to use by a specific customer. These take the form of a dedicated VLAN for a specific user organization, which means other cloud tenants are blocked from accessing resources in the VPC since they are not members of the same VLAN.

Installation and Configuration of Virtualization Management Tools

Virtualization management tools require particular security and oversight measures as they are essential to cloud computing. Without these in place, compromising a single management tool could lead to further compromise of hundreds or thousands of VMs and data. Tools that fall into this category include VMware vSphere, for example, as well as many of the CSP's administrative consoles that provide configuration and management of cloud environments by consumers.

Best practices for these tools will obviously be driven in large part by the virtualization platform in use and will track closely to practices in place for other high-criticality server-based assets. Vendor-recommended installation and hardening instructions should always be followed and possibly augmented by external hardening standards such as Center for Internet Security (CIS) Benchmarks or Defense Information Systems Agency (DISA) Security Technical Implementation Guides (STIGs). Other best practices include the following:

- **Redundancy**: Any critically important tool can be a single point of failure (SPOF), so adequate planning for redundancy should be performed. High availability and duplicate architecture will likely be appropriate due to the nature of these tools.
- **Scheduled downtime and maintenance**: Patching is crucially important for virtualization management tools, as is routine maintenance. Downtime may not be acceptable, so these tools may be patched or taken offline for maintenance on a rotating schedule with migration of live VMs to prevent loss of service.
- **Isolated network and robust access controls**: Access to virtualization management tools should be tightly controlled, with adequate enforcement of need-to-know restrictions and least privilege to ensure authorization is not excessive to a user's job function. Where possible, isolated networks, communication channels, or encryption should be utilized, such as a VPN or dedicated administrative network.

- **Configuration management and change management**: These tools and the infrastructure that supports them should be placed under configuration management to ensure they stay in a known, hardened state. When changes are necessary, formal change management should be utilized to ensure the impact of any changes are understood and new risks adequately mitigated.
- **Logging and monitoring:** Logging activities can create additional overhead, which may not be appropriate for all systems. Highly critical assets like virtualization management tools are likely to warrant the extra operational overhead and should be configured to log all activity.

Virtual Hardware–Specific Security Configuration Requirements

The cloud's heavy reliance on virtualization and multitenancy creates a new risk of data breaches when multiple users share a single piece of hardware. In a nonvirtual environment, it is comparatively difficult to leak data between systems, but a VM shares physical hardware with potentially hundreds of other machines.

The biggest issue related to virtual hardware security is the enforcement, by the hypervisor, of strict segregation between the guest operating systems running on a single host. The hypervisor acts as a form of reference monitor by mediating access by the various guest machines to the physical resources of the hardware they are running on and, in some cases, inter-VM network communication. Since most hypervisors are proprietary and produced by software vendors, there are two main forms of control a CCSP should be aware of:

- **Configuration**: Ensure the hypervisor has been configured correctly to provide the minimum necessary functionality, such as disallowing inter-VM network communications if not required or ensuring that virtualization tools like guest snapshots are encrypted.
- **Patching**: This control applies for any software. Monitor vulnerabilities disclosed that impact the virtualization tools in use by your organization, and ensure all patches are applied in a timely manner.

Virtual hardware is highly configurable and, especially in general-purpose environments like platform as a service (PaaS) or infrastructure as a service (IaaS) public cloud offerings, may not be configured for security by default. It is the responsibility of the cloud consumer to properly configure the cloud environment to meet their specific needs, such as segmenting networks and configuring network access controls to permit

only appropriate host-to-host communications. Particular concerns for virtual network security controls include the following:

- **Virtual private cloud (VPC)**: This is essentially a carved-out section of a public cloud dedicated to use by a particular customer. Similar to the way a VPN creates a virtual connection over the public Internet, a VPC gives the customer a greater level of control, including managing private nonroutable IP addresses and control over inter-VM communication, such as allowing only specific hosts in a middleware VPC to communicate on specific ports to databases in a database VPC. VPCs are essentially a hybrid of the public and private cloud models designed to balance cost savings of the public cloud with greater security in a private cloud.
- **Security groups**: In clouds, a security group is similar to an access control list (ACL) for network access. Security groups can be configured to control access to various elements of the cloud environment, and then new VMs instantiated have the security group applied to manage their network access.

A partial security concern related to virtual hardware configuration is the amount of virtual hardware provisioned. Many tools will allow the user to specify quantities, such as amount of memory, speed, and number of processing cores, as well as other attributes such as type of storage for a VM. Availability is obviously one concern—sufficient quantities of these virtual resources must be provisioned to support the intended workload. From a business perspective, these virtual resources should not be overprovisioned, however, because they do have associated costs.

An emerging trend in cloud environments is the provisioning of hardware using definition files, referred to as *infrastructure as code*. These definition files are read by the CSP and used to specify virtual hardware parameters and configurations, simplifying the process of setting up and configuring the environment. In many organizations, this changes the old paradigm of developers writing code and operations personnel configuring the hosts, because developers can package infrastructure definitions with their application code. This requires adequate training for developers to ensure they understand the business needs, security requirements, and configuration options available to them.

The ability to deploy infrastructure using a definition file enables a feature of cloud computing known as *auto-scaling*. Resources deployed in the cloud environment can be monitored for utilization and, as resources reach their limit, additional resources are automatically added. For example, a web server hosting an online ordering app may come under increased traffic when a celebrity endorses a product; in an auto-scaling cloud environment, new instances of the web server are spun up automatically to deal with the increased traffic. *Serverless computing* is another feature of cloud computing

that can support availability. Serverless environments like Azure Functions and AWS Lambda allow developers to deploy their application code without specifying server resources required. When the application is run—for example, a customer wants to place an order—the CSP provides sufficient resources to handle that demand, supporting availability. The cloud consumer pays for the resources only when they are running the particular application, saving costs.

Installation of Guest Operating System Virtualization Toolsets

Toolsets exist that can provide extended functionality for various guest operating systems including Unix, Linux, and Microsoft Windows. These toolsets provide specific or enhanced capabilities for a particular operating system (OS), such as support for additional devices, driver software, or enhanced interactivity. In a public cloud, these toolsets will typically be provided by the CSP; if a customer requires functionality that depends on a particular OS toolset, it is up to the customer to verify if the CSP can support it before using that provider's cloud.

When managing your own virtualization environment, installing these toolsets should follow the concept of minimum necessary functionality. If a virtualization cluster will not be running virtual Windows servers, then the Windows OS toolset does not need to be installed. The personnel responsible for the virtualization cluster need to understand the business requirements and build the solution appropriately.

OPERATE PHYSICAL AND LOGICAL INFRASTRUCTURE FOR CLOUD ENVIRONMENT

Cloud computing has shifted many responsibilities for managing physical and logical infrastructure away from the users of the corresponding services. When organizations hosted their own infrastructure, it was essential to have adequate processes to assess risks and provision adequate security controls to mitigate them, but in the cloud, many of these tasks are the responsibility of the cloud provider instead. However, the consumers must still be aware of the risks inherent in using the cloud. It is essential for a CCSP to understand these matters to adequately assess cloud services, and doubly important if the organization is providing a cloud service. In the case of a private cloud host or a security practitioner working for a CSP, these controls will be directly under the purview of the organization.

Configure Access Control for Local and Remote Access

In most instances, access to cloud resources will be done remotely, so adequate security controls must be implemented in the remote administrative tools implemented to support these functions. Protocols for supporting remote administration a CCSP should be familiar with are as follows:

- **Secure Shell (SSH)**: As the name implies, this standard provides a secure way for an administrator to access and manipulate a remote system. This is often achieved by interacting with a local command-line interface (CLI), which sends commands to the remote host for execution. SSH can be configured to implement encryption using either symmetric (shared) or asymmetric (public/private key pair) cryptography. In both cases, cryptography is used for both protection of data via encryption as well as authentication of users based on the assumption that a user has maintained control of their key. Only users with the appropriate key(s) will be granted access.
- **Remote Desktop Protocol (RDP)**: RDP was initially a technology specific to Microsoft Windows but is now widely available across Windows, macOS, Linux, and mobile operating systems including iOS and Android. A typical session requires both an RDP server, which provides remote access and control to the machine it is running on, and an RDP client through which the user interacts with the remote machine. Security features available in RDP include encryption of data in transit, user authentication, and bandwidth management to ensure remote desktop sharing and interaction are adequate to support user interactions. As a key element of the Windows OS as well as remote access, RDP functionality is often a target of hackers, so it is critical to maintain up-to-date versions as part of patch management.

  Access to RDP can be controlled in a variety of ways. As a Microsoft standard, Active Directory is often utilized for identification and authentication, and the RDP standard also supports smart card authentication.
- **Virtual Network Computing (VNC)**: This may be considered analogous to RDP and is often implemented for remote access and control of Unix- or Linux-based systems where RDP is not native.

In situations where local administration is being performed, a Secure Keyboard Video Mouse switch (KVM) may be utilized. This is a device that allows access to multiple hosts using a single set of human interface peripherals such as a keyboard, mouse, and monitor—a user does not need to have multiple keyboards on their desk physically attached to different computers.

A basic KVM allows the user to switch their peripherals to interact with various computers; a secure KVM adds additional protections for highly secured environments that primarily address the potential for data to leak between the various connected systems. Attributes of secure KVMs include the following:

- **Isolated data ports**: The physical construction ensures that each connected system is physically isolated from others.
- **Tamper-evident or tamper-resistant designs**: Secure KVMs may be manufactured to make physical tampering extremely difficult or impossible without destroying vital hardware, such as permanently soldered circuit boards or components. They may also implement tamper-evident stickers/labels that make it obvious if the KVM has been opened because the sticker is torn.
- **Secure storage**: KVMs may implement a buffer to store data, but switching between connected systems causes any data in the buffer to be erased. The buffer may additionally have only limited capacity, which reduces the usefulness but also reduces the amount of data that could potentially be leaked.
- **Secured firmware**: The software required by the KVM to run may often be fixed—it can't be changed without rendering the unit inoperable—or require signed firmware updates from the original manufacturer. Both are designed to prevent tampering.
- **Physical disconnects**: A KVM typically contains buttons on the front that allow the user to switch between the various connected systems. A physical disconnect physically breaks the connection between each system when a new system button is pressed, preventing data leaks from the currently connected system to the next system being connected.
- **USB port and device restrictions**: KVMs typically offer USB ports that are used to connect the peripherals being shared. Highly secured KVMs may implement restrictions on the type of USB devices that can be connected, such as allowing keyboards and mice but blocking mass storage devices, as a way to restrict what a malicious user can do if they gain physical access.

All CSPs provide remote administrative access to cloud users via an administrative console. In AWS this is known as the Management Console, in Azure as the Portal, and in Google Cloud as the Cloud Console. All three offer a visual user interface (UI) for interaction that allows for creation and administration of resources including user accounts, VMs, cloud services such as compute or storage, and network configurations. These functions are also available via CLI tools, most of which call APIs to perform these same administrative functions manually and which enable automation such as creating resources based on infrastructure as code definition files.

The CSP is responsible for ensuring access to these consoles is limited to properly authenticated users, and in many cases, the CSPs restrict access to consumer resources entirely. For example, it is possible for a user in Azure to create a VM and remove all network access from it, effectively locking themselves out as well. Even Microsoft cannot reconfigure that VM, because allowing that level of access to consumer resources is incredibly risky. By contrast, the cloud consumer is responsible for implementing appropriate authorization and access control for their own members in accordance with access management policies, roles, and rules. Due to the highly sensitive nature and abilities granted by these consoles, they should be heavily isolated and protected, ideally with multifactor authentication. All use of the admin console should be logged and should be routinely reviewed as a critical element of a continuous monitoring capability. Access to the UI or CLI functions is a key area for personnel security and access control policies, similar to any system or database admin position in an on-prem environment.

Secure Network Configuration

One aspect of cloud services is their broad network accessibility, so it is virtually impossible to find a cloud security concern that does not in some way relate back to secure network connectivity. Several protocols and concepts are important to understand as they relate to securing networks and the data transmitted.

Virtual Local Area Networks

VLANs were originally designed to support the goal of availability by reducing contention on a shared communications line. They do so by isolating traffic to just a subset of network hosts, so all hosts connected to the VLAN broadcast their communications to each other but not to the broader network. Communication with other VLANs or subnets must go through a control device of some sort, often a firewall, which offers confidentiality and integrity protections by enforcing network-level access control. For example, in a multitiered application architecture, the web servers will traditionally be isolated from database servers in separate VLANs and the database layer will implement much more restrictive access controls.

VLAN network traffic is identified by the sending device using a VLAN tag, specified in the IEEE802.1Q standard. The tag identifies the VLAN that the particular data frame belongs to and is used by network equipment to determine where the frame will be distributed. For example, a switch will not broadcast frames tagged VLAN1 to devices connected to VLAN2, alleviating congestion. Firewalls that exist between the two VLANs can make decisions to allow or drop traffic based on rules specifying allowed or denied ports, protocols, or sender/recipient addresses.

An extension of VLAN technology specific to cloud computing environments is the virtual extensible LAN (VXLAN) framework. It is intended to provide methods for designing VLANs utilizing layer 2 protocols onto layer 3; effectively, it allows for the creation of virtual LANs that may exist across different data centers or cloud environments. VXLAN is more suitable than VLANs for complex, distributed, and virtualized environments due to limitations on the number of devices that can be part of a VLAN, as well as limitations of protocols designed to support availability of layer 2 devices such as Spanning Tree Protocol (STP). It is specified in RFC 7348.

From the standpoint of the cloud consumer, network security groups (NSGs) are also a key tool in providing secure network services and provide a hybrid function of network traffic isolation and filtering similar to a firewall. The NSG allows or denies network traffic access based on a list of rules such as source IP, destination IP, protocol, and port number. Virtual resources can be segmented (isolated) from each other based on the NSGs applied to them. For example, a development environment may allow inbound traffic only from your organization's IP addresses on a broad range of ports, while a production environment may allow access from any IP address but only on ports 80/443 for web traffic. Resources in the development environment could also be prevented from communicating with any resources in production to prevent an attacker from pivoting.

Transport Layer Security

Transport Layer Security (TLS) is a set of cryptographic protocols that provide encryption for data in transit, and it replaced a previous protocol known as Secure Sockets Layer (SSL). The current version of TLS is 1.3; versions below this either are considered less secure or have demonstrated compromises. TLS provides a framework of supported cryptographic ciphers and keylengths that may be used to secure communications, and this flexibility ensures broad compatibility with a range of devices and systems. It also means the security practitioner must carefully configure their TLS-protected systems to support only ciphers that are known to be secure and disable older options if they have been compromised.

TLS specifies a handshake protocol when two parties establish an encrypted communications channel. This comprises three steps:

1. The client initiates a request with a ClientHello message. This provides a list of cipher suites and TLS version it supports.
2. The server (if configured properly) chooses the highest-supported TLS version and cipher suite and communicates that choice to the client, along with the server's certificate containing a public key.

3. Depending on the cipher suite chosen, the client and server may then exchange other data, including a pre-master secret key used to negotiate a session key. The session key is utilized to encrypt all data that is to be shared.

TLS can be used to provide both encryption of data as well as authentication/proof of origin, as it relies on digital certificates and public key cryptography. In many cases, such as publicly available web apps, one-way authentication is used for the client's browser to authenticate the server it is connecting to, such as an online banking application. In *mutual authentication*, both client and server (or server and server) are required to exchange certificates. If both parties trust the issuer of the certificates, they can mutually authenticate their identities. Some high-security or high-integrity environments require mutual authentication before data is transmitted, but the overhead makes it largely infeasible for all TLS encryption (imagine the PKI required to authenticate every Internet user, web app, and IoT device on the planet).

Dynamic Host Configuration Protocol

Dynamic Host Configuration Protocol (DHCP) enables computer network communications by providing IP addresses to hosts that dynamically join and leave the network. The process of dynamically assigning IP addresses is as follows:

1. A DHCP server running on the network listens for clients.
2. When a new client machine joins a network, it sends out a DHCPDISCOVER request on UDP port 67, to which the DHCP server responds with a DHCPOFFER on port 68.
3. The client responds with a DHCPREQUEST indicating it will use the address assigned, and the server acknowledges with a DHCPACK.

An easy mnemonic to remember this DHCP process is DORA for Discover, Offer, Request, Acknowledge. As with many network protocols, security was not originally part of the DHCP design. The current DHCP version 6 specifies how IPSec can be utilized for authentication and encryption of DHCP requests. An improperly configured DHCP server can lead to denial of service if incorrect IP addresses are assigned or, worse, can be used in a man-in-the-middle attack to misdirect traffic.

Domain Name System and DNS Security Extensions

Imagine how difficult it would be if you were required to remember every company's and person's IP address in order to send them data. Human-readable addresses like

www.isc2.org are used instead, but these need to be converted to a machine-readable IP address. DNS does this by a process known as *resolving*.

1. A user initiates a network communication using a fully qualified domain name (FQDN). This might be done by entering a URL into a web browser or sending an email to a particular domain.
2. The user's machine queries a Domain Name System (DNS) resolver service on UPD port 53, which provides the IP address associated with the FQDN. The user's machine uses this information to address the communications and carry out the user's action—for instance, loading a web page or sending an email.

DNS operates using records, definitive mappings of FQDNs to IP addresses, which are stored in a distributed database across zones. DNS queries are facilitated by DNS servers sharing information from time to time, known as *zone transfers*, which enable resolution of client requests without having to iterate a request.

DNS Attacks

As with many network protocols, DNS was originally designed without security in mind, which leaves it open to attack. *Cache poisoning* is an attack where a malicious user updates a DNS record to point an FQDN to an incorrect IP address. One way of doing this is to initiate a zone transfer, which by default does not include authentication of the originator. This poisoned record causes users to be redirected to an incorrect IP address, where the attacker can host a malicious phishing site, malware, or simply nothing at all to create a denial of service.

DNS spoofing is another attack against DNS. In this case, an attacker spoofs a DNS service on a network with the goal of resolving a user's requested FQDN to an attacker-controlled IP address. Some attacks against DNS also abuse graphical interfaces instead of the underlying infrastructure. This can be achieved using characters that appear identical to users, such as zero and the letter *o*, sending users to an illegitimate site.

DNS Security Extensions (DNSSEC) is a set of specifications primarily aimed at reinforcing the integrity of DNS. It achieves this by providing for cryptographic authentication of DNS data using digital signatures. This provides proof of origin and makes cache poisoning and spoofing attacks more difficult if users cannot create a proper digital signature for their DNS data. It does not provide for confidentiality, since digital signatures rely on publicly decryptable information, nor does it stop attacks using graphically similar domain names, as these may be legitimate records registered with an authoritative DNS zone.

Virtual Private Network

Resources inside a network can be protected with security tools like firewalls and access controls, but what happens if users are not on the same network? A VPN gives external users the ability to virtually and remotely join a network, gaining access to resources

hosted on that network and benefiting from the security controls in place. This security is achieved by setting up a secure tunnel, or encrypted communication channel, between the connecting host and the network they want to join. On the target network, there is often a VPN device or server that authenticates the user connecting and mediates their access to the network.

Commonly used to provide remote workers access to in-office network resources, VPNs can also be useful in cloud architectures to safely share data between offices, cloud environments, or other networks. These are often implemented at the edge of networks to allow secure communication between any host on connected networks and are called *site-to-site* or *gateway-to-gateway* VPNs.

There are a variety of VPN protocols and implementations that a CCSP should be familiar with, such as the following:

- **OpenVPN**: This is an open-source VPN built on encryption in the OpenSSL project, which can be deployed and run on your own infrastructure, as well as commercial service options. These consist of packages that can be easily deployed in cloud environments as a virtual VPN server. OpenVPN can be used to establish site-to-site VPN connectivity and provides client-server functionality across a wide variety of Linux and Unix operating systems, as well as several versions of the Windows and macOS operating systems. Details can be found at `openvpn.net`.
- **Internet Key Exchange v2 and IPSec (IKEV2/IPSec)**: This protocol utilizes the Security Associate (SA) features of IPSec to establish the encrypted communications channel. Public keys are exchanged and Diffie–Hellman is used to independently calculate the shared session key. It is widely built into Microsoft products and Apple iOS devices, which may simplify deployment as no additional software is required.
- **SSL VPN**: Although the use of SSL has largely been replaced with TLS, these are still commonly referred to as SSL VPNs. These are often implemented in a browser due to ubiquitous support for TLS encryption and can provide remote connectivity for users without requiring the installation of additional software. This ubiquity is also a potential downside, as a user might be connecting from any Internet-connected machine, which may lack standard security software, be missing patches, or have malware installed. In this case, compensating controls such as tightly restricted access for SSL VPN users or some form of network access control (NAC) may be required.

Software-Defined Perimeter

A software-defined perimeter (SDP) is an emerging concept driven by the decentralized nature of cloud applications and services, which have upended the traditional model of a

network with a perimeter (secure boundary). Since cloud applications may reside in data centers anywhere in the world, and users may also be connecting from anywhere in the world, it is no longer possible to define a perimeter.

1. One or more SDP controllers are created, which are connected to an authentication service to enforce access control.
2. Accepting SDP hosts are brought online and authenticate to the SDP controller(s). By default, accepting hosts do not accept communication from any other host.
3. Initiating SDP hosts connect to the SDP controller(s) for authentication and can request access to resources on an accepting host. The SDP controller makes authorization decisions and provides details to both the initiating and accepting hosts. These details include authorization and encryption policies to establish a VPN.
4. A mutual VPN is established between the initiating and accepting hosts, and the user is able to interact with the resource.

For further reference, see the CSA SDP site: `cloudsecurityalliance.org/research/working-groups/software-defined-perimeter`.

Operating System Hardening through the Application of Baselines

Hardening is the configuration of a machine into a secure state. Common hardening practices include changing default credentials, locking/disabling default accounts that are not needed, installing security tools such as anti-malware software, configuring security settings available in the OS, and removing or disabling unneeded applications, services, and functions.

Prior to virtualization, the process of hardening an OS was often a manual task; once a system's hardware was installed, an administrator had to manually install and configure the OS and any application software. The advent of virtualization introduced machine images, which are essentially templates for building a VM. All VMs created from the same image will have the same settings applied, which offers dual benefits of efficiency and security. Of course, the image cannot remain static—patches and other security updates must be applied to the image to ensure new VMs remain secure. Existing VMs must also be updated independently of the image, which is a concern of patching and configuration management disciplines.

A modern OS has thousands of configuration options, so to speed this up, an organization may choose to create or use a baseline configuration. Baselines are simply a documented, standard state of an information system, such as access control requiring

multifactor authentication, vulnerable services such as File Transfer Protocol (FTP) disabled, and nonessential services such as Windows Media Player removed. Each of these configuration options should match a risk mitigation (security control objective).

This baseline and corresponding documentation may be achieved in a number of ways.

- **Customer-defined VM image**: A customer spins up a VM and configures it to meet their specifications. Virtualization tools allow you to create an image from an existing machine, and from then on, this image may be used to create secure VMs. This option may also be built on top of one of the other options in this list. For example, an organization might use a CIS Benchmark as a starting point, tailor it to their specific needs, and create an image from that. This is similar to word processing software that allows you to create a template from an existing document.
- **CSP-defined images**: CSPs may offer one or more of the following as a PaaS solution: popular operating systems like Microsoft Windows and various Linux distributions, databases and big data tools, or virtualization platforms like Hyper-V and VMWare. These images often incorporate the latest patches and may have some security configuration already applied, but they should always be evaluated for appropriateness against the organization's own security needs and tailored where needed.
- **Vendor-supplied baselines**: Microsoft, VMware, and some Linux creators offer configuration guidelines for their products that point out specific security options and recommended settings. As with any external source, it is imperative the organization evaluate the recommendations against their own needs.
- **DISA STIGs**: The US Defense Information Systems Agency (DISA) produces baseline documents known as Security Technical Implementation Guides (STIGs). These documents provide guidance for hardening systems used in high-security environments, and as such, may include configurations that are too restrictive for many organizations. They are available for free and can be tailored, obviously, and also cover a broad range of OS and application software. Many configuration and vulnerability management tools incorporate hardening guidance from the STIGs to perform checks.

  STIGs and additional information can be found here: `public.cyber.mil/stigs/downloads`.
- **NIST checklists:** The National Institute of Standards and Technology (NIST) maintains a repository of configuration checklists for various OS and application software. It is a free resource.

  NIST Checklist repository (which also includes access to DISA STIGs) can be found here: `nvd.nist.gov/ncp/repository`.

- **CIS Benchmarks**: The Center for Internet Security (CIS) publishes baseline guides for a variety of operating systems, applications, and devices, which incorporate many security best practices. These can be used by any organization with appropriate tailoring and are also built into many security tools such as vulnerability scanners.

  More information can be found here: `www.cisecurity.org/cis-benchmarks`.

Availability of Stand-Alone Hosts

Stand-alone hosts are isolated, dedicated hosts for the use of a single tenant. These are often required for contractual or regulatory reasons, such as processing highly sensitive data like healthcare information. The use of nonshared resources carries obvious consequences related to costs. A CSP may be able to offer secure dedicated hosting similar to colocation, which still offers costs savings to the consumer due to shared resources including physical facilities and shared resources like power and utilities. A CCSP will need to gather and analyze the organization's requirements to identify if the costs of stand-alone hosting are justified. In some CSPs, the use of virtual private resources may be an acceptable alternative with lower costs, due to the use of shared physical infrastructure with strong logical separation from other tenants.

Availability of Clustered Hosts

Clusters are a grouping of resources with some coordinating element, often a software agent that facilitates communication, resource sharing, and routing of tasks among the cluster. Clustered hosts can offer a number of advantages, including high availability via redundancy, optimized performance via distributed workloads, and the ability to scale resources without disrupting processing via addition or removal of hosts to the cluster. Clusters are a critical part of the resource pooling that are foundational to cloud computing and are implemented in some fashion for most resources needed in modern computing systems including processing, storage, network traffic handling, and application hosting.

The cluster management agent, often part of hypervisor or load balancer software, is responsible for mediating access to shared resources in a cluster. *Reservations* are guarantees for a certain minimum level of resources available to a specified virtual machine. The virtualization toolset or CSP console is often where this can be configured, such as a certain number of compute cores or RAM allocated to a VM. A *limit* is a maximum allocation, while a *share* is a weighting given to a particular VM that is used to calculate percentage-based access to pooled resources when there is contention.

Maintenance mode refers to the practices surrounding the routine maintenance activities for clustered hosts. Although taking a system offline primarily impacts availability,

there are considerations for all three elements of the confidentiality, integrity, availability (CIA) triad related to maintenance mode.

- **Availability**: Maintenance usually involves taking a system offline, which is obviously counter to the goal of availability. To meet obligations for uptime and availability, the CSP should migrate all running VMs or services off a cluster prior to entering maintenance mode. Most virtualization toolsets automate this process, which is referred to as *live migration*. In some limited cases, a CSP may also take a service completely offline; in these cases, all consumers must be made aware of the outage beforehand, either through ad hoc communication or through a published maintenance window. SLAs should be documented that take into account the need for both routine and emergency maintenance tasks.
- **Confidentiality**: Many live migration tools transmit data in cleartext due to the operational overhead incurred by trying to encrypt all the data related to a running OS and applications. If this is the case, compensating controls may be required for the migration, especially if the live migration data will be transmitted across untrusted network segments.
- **Integrity**: During maintenance mode, customers are not able to access or get alerts regarding the environment's configuration. The CSP's change management process needs to include robust integrity controls such as change approvals and documentation, and systems should still generate logs when in maintenance mode to support after-the-fact investigation if needed.

High Availability

Availability and uptime are often used interchangeably, but there is a subtle difference between the terms. Uptime simply measures the amount of time a system is running. In a cloud environment, if a system is running but not reachable due to a network outage, it is not available. Availability encompasses infrastructure and other supporting elements in addition to a system's uptime; *high availability* (HA) is defined by a robust system and infrastructure to ensure a system is not just up but also available. It is often measured as a number of 9s; for example, five nines or 99.999 percent availability. This equates to approximately 5 minutes of downtime per year and should be measured by the cloud consumer to ensure the CSP is meeting SLA obligations.

Organizations can implement multiple strategies to achieve HA. Some, detailed in the following sections, are vendor-specific implementations of cluster management features for maintaining system uptime. Other strategies include redundancy of infrastructure such as network connectivity and utilities. The Uptime Institute publishes

specifications for physical and environmental redundancy, expressed as tiers, that organizations can implement to achieve HA.

- Tier 1 involves no redundancy and the most amount of downtime in the event of unplanned maintenance or an interruption.
- Tier 2 provides partial redundancy, meaning an unplanned interruption will not necessarily cause an outage.
- Tiers 3 and 4 provide N+1 and 2N+1 levels of redundancy, which results in increased availability. Tier 3 allows for planned maintenance activities without disruption, but an unplanned failure can still cause an outage. Tier 4 is known as fault-tolerant, meaning it can withstand either planned or unplanned activity without affecting availability. This is achieved by eliminating all single points of failure and requires fully redundant infrastructure such as dual commercial power feeds and dual backup generators. This redundancy provides very high availability but also comes at a higher cost.

Distributed Resource Scheduling

Distributed Resource Scheduling (DRS) is the coordination element in a cluster of VMware ESXi hosts, which mediates access to the physical resources and provides additional features supporting high availability and management. It is a software component that handles the resources available to a particular cluster, the reservations and limits for the VMs running on the cluster, and maintenance features.

DRS maintenance features include the ability to dynamically move running VMs from one physical hardware component to another without disruption for the end users; this is obviously useful if hardware maintenance needs to be performed or additional capacity is added to the cluster and the workload needs to be rebalanced. This supports the element of rapid elasticity and self-service provisioning in the cloud by automating dynamic creation and release of resources as client demands change.

DRS can also handle energy management in physical hardware to reduce energy consumption when processing demands are low and then power resources back up when required. This enables cost savings for both the CSP and consumer, as hardware that is not actively being used does not consume energy.

Microsoft Virtual Machine Manager and Dynamic Optimization

Similar to VMware's DRS, Microsoft's Virtual Machine Manager (VMM) software handles power management, live VM migration, and optimization of both storage and compute resources. Hosts (servers) and storage capacity can be grouped into a cluster; options available to configure in the VMM console include the movement of VMs and virtual hard disks between hosts to balance workloads across available resources.

Storage Clusters

Storage clusters create a pool of storage, with the goal of providing reliability, increased performance, or possibly additional capacity. They can also support dynamic system availability by making data available to services running anywhere—if a data center in one part of the country fails and web hosts are migrated to another data center, they can connect to the same storage cluster without needing to be reconfigured. There are two primary architectures for storage clusters:

- Components in a *tightly coupled* cluster are often all provided by the same manufacturer, and updates or expansion must come from that same manufacturer. The advantage of a tightly coupled cluster is generally better performance due to the division of data into deterministic blocks. When a file is written to such a storage cluster, it is broken down into blocks that are faster to read and write to disks than the entire file. This is especially relevant if the data is to be mirrored, as writing multiple copies of a long file will take longer than duplicating blocks, and blocks can be written across multiple nodes or disks simultaneously.
- *Loosely coupled* storage offers more flexibility and lower cost at the expense of performance. Components can usually be added using any off-the-shelf parts, but the use of file-level storage means operations will be slower.

Availability of Guest Operating Systems

The availability of a guest OS in a CSP's environment is generally the consumer's responsibility, as the CSP only provides a base image. Once a VM is created in IaaS the CSP no longer has direct control over the OS, while in PaaS the CSP maintains control. In the software as a service (SaaS) model, the consumer only needs to plan for availability of the data their organization puts into the app.

Ensuring the availability of guest OSs in a cloud environment may involve planning for backup and restoration, which will be similar to traditional on-prem backup and recovery planning, or it may involve utilizing cloud-specific features to design resiliency into the system. Details of these two approaches are presented here:

- **Backup and recovery**: This is a more traditional method that assumes a system will be built, may be interrupted, and will then need to be recovered. In virtualized cloud infrastructure, this might involve the use of snapshots, which are provided by virtualization toolsets. Snapshots typically capture the state of a VM's primary storage, random access memory (RAM), and software configurations, which can be used to re-create the VM—at the point of the snapshot—on another

physical host. Obviously, this approach could involve loss of data between the snapshot creation and the failure, as well as the time required to manually create the new VM instance.

As with all backup and restoration activity, there are concerns across all three CIA triad elements. Backup integrity should be routinely tested to ensure recovery, and the backups should not be stored on the same physical hardware as the primary systems since this single point of failure could impact availability. Snapshots will contain data with the same sensitivity level as the systems they are made from, so adequate access controls and other measures to enforce confidentiality are required.

- **Resiliency**: Building resiliency is achieved by architecting systems to handle failures from the outset rather than needing to be recovered. One example is using a clustered architecture with live migration; in this case, if a physical hardware failure is detected, all running VMs are transferred to another physical host, with little or no interruption to the users.

  Many cloud services also have resiliency options built in, such as worldwide data replication and availability zones or regions that can transfer running apps or services transparently in the event of a failure. The cloud consumer is responsible for choosing and configuring these resiliency options, and in some cases will need to make trade-offs. For example, some CSPs offer database encryption that makes it harder to perform data replication. In traditional on-prem architecture, building such a resilient app would have been cost prohibitive for all but the largest organizations, but the inherent structure of cloud computing makes this type of resiliency broadly available.

MANAGE PHYSICAL AND LOGICAL INFRASTRUCTURE FOR CLOUD ENVIRONMENT

Although many elements of physical and logical infrastructure in cloud environments will be under the direct control of the CSP, it is essential for cloud consumers and the CCSP practitioner to be aware of these practices. In some cases, there will be shared responsibilities that both parties are required to perform, and in others, the consumer must adequately understand these practices and conduct oversight activities like SLA reviews to ensure security objectives are being met.

Access Controls for Remote Access

Remote administration is the default for a majority of cloud administrators from both the CSP and the consumer side. Tools including Remote Desktop Protocol (RDP), used

primarily for Windows systems, and Secure Shell (SSH), used primarily on Unix and Linux systems, must be provisioned to support this remote management.

Secure remote access is a top-level priority in many security frameworks due to the highly sensitive nature of operations it entails and its inherent exposure to attacks. Remote access often relies on untrusted network segments for transmitting data, and in a cloud environment, this will entail users connecting via the Internet. Physical controls that could prevent unwanted access in a data center will be largely missing in a cloud as well; not that the CSP is ignoring physical security controls, but the inherently network-accessible nature of the cloud means most administrative functions must be exposed and are therefore susceptible to network-based threats.

There are a number of concerns that should be addressed to reduce the risk associated with remote access, including the following:

- **Session encryption**: Remote access always requires additional security because it happens outside an organization's perimeter, and cloud-based applications often do away with secure perimeters altogether. Data transmitted in remote access sessions must be encrypted using strong protocols such as TLS 1.3 and should implement cryptographic best practices such as session-specific cryptographic keys, which reduce the possibility of replay attacks.
- **Strong authentication**: Users performing remote administration present higher risk to the organization, so more robust authentication is appropriate for these users. This may be combined with cryptographic controls such as a shared secret key for SSH, assuming the user retains control over their key, as well as the use of two or more authentication factors for multifactor authentication (MFA), such as a one-time code or hardware key.
- **Separate privileged and nonprivileged accounts**: A general best practice for administrative users is the use of a dedicated admin account for sensitive functions, and a standard user account for normal functions such as daily web browsing or email. This approach offers two benefits: first, it reduces threats related to phishing by reducing the likelihood a user's admin account credentials will be stolen, and second, it allows the organization to implement more stringent controls on the admin account without creating undue overhead for the user's daily business account. In many cloud environments, this is implemented by default, as a user's credentials to log in to the cloud admin console are separate from their main email or computer login.
- **Enhanced logging and reviews**: This is also a general best practice not specific to cloud remote access. All admin accounts should be subject to additional logging and review of activity, as well as more frequent review of permissions.

- **Use of identity and access management tool**: Many CSPs offer identity and access management tools specific for their environments, and third party identity as a service (IDaaS) providers also exist, which offer the ability to manage logical access controls across cloud and on-prem applications. Many of these offer easy management of the previously discussed controls for administrator accounts, such as enhanced logging or more stringent password requirements. Examples of IDaaS tools include offerings from companies such as Okta, Microsoft Azure AD, Ping, and Auth0.
- **Single sign-on (SSO)**: Two of the major CSPs also offer productivity software: Google Workspace and Microsoft's 365 product line (previously known as Office 365). Both platforms enable users to log into other services using their company accounts, similar to the option in many consumer services to log in with a Facebook or Gmail account. This reduces the burden on users to remember passwords and simplifies administration of user access by reducing the number of accounts. Many IDaaS solutions also offer the ability to function as an SSO provider.

Operating System Baseline Compliance Monitoring and Remediation

A hardened OS implements many of the security controls required by an organization's risk tolerance and may also implement security configurations designed to meet the organization's compliance obligations. Once built, however, these systems do need to be monitored to ensure they stay hardened. This ongoing monitoring and remediation of any noncompliant systems must be part of the organization's configuration management processes, designed to ensure no unauthorized changes are made, any unauthorized changes are identified and rolled back, and changes are properly approved and applied through a change control process.

Monitoring and managing OS configuration against baselines can be achieved in a number of ways. Some are similar to legacy, on-prem techniques, while newer methods are also emerging, including the following:

- **Use of a configuration management database (CMDB) and CM audits**: The organization's CMDB should capture all configuration items (CIs) that have been placed under configuration management. This database can be used for manual audits as well as automated scanning to identify systems that have drifted out of their known secure state. For example, an auditor performing a manual audit could pull the registry file from a sample of Windows servers and compare the entries against the configuration baseline values. Any systems that deviate from the baseline should be reconfigured, unless a documented exception exists.

Automated configuration and vulnerability scanners can also perform this task on a more routine basis, and any findings from these scans should be treated using standard vulnerability management processes.

- **Organization-wide vulnerability scanning**: System-specific vulnerability and configuration scans should be complimented by the organization's broader vulnerability management policy. For example, insecure services such as FTP may be disallowed on all organization systems, and a vulnerability scanner can easily identify a server responding to FTP requests. This vulnerability indicates a system that does not conform to baseline configuration and that requires immediate remediation action.
- **Immutable architecture**: This is an evolving solution to the problem of systems that, over time, can drift away from baseline configurations. Immutable architecture is unchangeable but short lived. In cloud environments, it is possible to tear down all virtual infrastructure elements used by an old version of software and deploy new virtual infrastructure quite simply; in traditional architecture, this process would be much more difficult and time-consuming. Immutable architecture can address baseline monitoring and compliance by limiting the amount of time a host can exist in a noncompliant state; the next time the application is deployed, the old host is torn down, and a new VM is built from the standard baseline image.

Patch Management

Maintaining a known good state is not a static activity, unfortunately. Today's hardened system is tomorrow's highly vulnerable target for attack, as new vulnerabilities are discovered, reported, and weaponized by attackers. Patch management involves identifying vulnerabilities in your environment, applying appropriate patches or software updates, and validating the patch has remediated the vulnerability without breaking any functionality or creating additional vulnerabilities.

Information needed to perform patch management can come from a variety of sources, but primary sources include vendor-published patch notifications, such as Microsoft's Patch Tuesday, as well as vulnerability scanning of your environment. Patch management processes will vary depending on the cloud service model you are using.

- For SaaS environments, the consumer has almost no responsibilities, as applying patches is the CSP's purview. Verifying that patches are being applied according to established SLAs is a recommended practice for the consumer organization, and maintaining oversight on this metric is important. Additionally, some SaaS offerings provide staggered or on-your-own schedule patching, especially for custom software. In these models, the consumer may have the option to apply patches/updates immediately or apply to a small sample of users for testing

purposes. In this case, the consumer must understand the shared responsibility and take appropriate action.

- For IaaS and PaaS, it is usually the consumer's exclusive responsibility to apply patches to existing infrastructure. Hardware-based patches will likely be handled by the CSP, who is also likely to maintain up-to-date template, used for PaaS VMs. Once a consumer creates a VM from a PaaS template the CSP retains responsibility for patching the OS, while the consumer is responsible for patching any applications installed on the VM.

Patch management tools exist that can help to identify known software vulnerabilities and the state of patching across an organization's system, such as Windows Server Update Services (WSUS). Such tools can also be used to orchestrate and automate patch application, though in some cases automation may not be desirable if a patch has operational impacts. There are plenty of patches and updates from major companies that caused unknown issues when installed, including turning otherwise functional hardware into bricks! A generic patch management process ought to incorporate the following:

- **Vulnerability detection**: This may be done by security researchers, customers, or the vendor. A software flaw, bug, or other issue that could be exploited drives the need for a patch.
- **Publication of patch**: The vendor provides notice, either through a standard update mechanism or through ad hoc communication such as press releases, circulating details to information sharing and analysis centers (ISACs), or other means. Consuming organizations ought to have subscriptions to relevant information sources for their industry and infrastructure. In other words, if an organization uses Red Hat Linux exclusively, the IT and security teams ought to be subscribed to relevant data feeds from Red Hat to receive notice of vulnerabilities and patches.
- **Evaluation of patch applicability**: Not all users of software will need to apply all patches. In some cases, there may be features or functions that are not in use by an organization, meaning the patched vulnerability has no real impact. As a general best practice, all applicable patches should be applied unless there is a known functionality issue with the patch.
- **Test**: Most patches should be tested in a limited environment to identify any potential functionality issues before being broadly deployed; patches designed to close highly critical vulnerabilities may be treated as an exception, since the vulnerability they remediate is riskier than the potential for an outage.
- **Apply and track**: Assuming a patch does have negative functionality impacts, the organization must identify all systems that require the patch and then plan for and

track the deployment to ensure it is applied to all systems. In many organizations, there will be a key security metric worth tracking. Patches should have timeframes for application based on their criticality; for example, critical patches must be applied within seven days of release. This service level helps the organization measure and track risk and associated remediation efforts.

NOTE VMs present a particular challenge for patching efforts. As discussed, one of the features in many virtualization tools is the ability to power down a VM when it is not needed, but this can have the unintended consequence of leaving systems unpatched. The organization should have compensating controls, such as powering on all VMs when patches are deployed or performing checks when a VM is first powered on to detect and apply any missing patches.

- **Rollback if needed**: Not all patching goes smoothly, so a rollback plan is essential. Virtualization makes this incredibly easy, as a snapshot can be created before the patch is applied, but that step must be part of the deployment process.
- **Document**: Patched software represents a changed system baseline; the new, patched version of software is the new known good state. Systems used to track this, such as the CMDB need to be updated to record this official change and new expected configuration.

Evolving cloud architectures are introducing new ways of managing patching, which offers significant advantages if applied correctly.

- The use of infrastructure as code, where system architecture is documented in a definition file used by the CSP to spin up virtual infrastructure, offers a way for organizations to utilize a CSP's patched template files. These definition files are often used to create virtual infrastructure in PaaS environments and should always make use of the latest patched version of the platforms in question.
- Immutable architecture, which is created each time a system is deployed and then torn down when the next deployment occurs, can also help prevent the spread of old systems with unpatched software. This relies on infrastructure as code and should be configured to always make use of the latest, patched infrastructure elements.
- Software composition analysis (SCA) is a concern for applications built with open source software components, many of which provide highly reusable elements of modern applications such as data forms or code libraries. Since this type of software is included in many applications, SCA tools identify flaws/vulnerabilities in these included pieces of the application. This represents a merging of application

security and patch management and is a key automation point for security in application development. Identifying vulnerabilities in these included functions and ensuring the latest, patched versions are in use by your application is a critical part of both development processes and patch management.

✔ Equifax Data Breach

In 2017, credit monitoring company Equifax suffered a massive data breach that affected more than 140 million people, primarily in the United States but in other countries as well. The company maintains highly sensitive personal information used in financial decision-making, and this information was exposed due to an unpatched software flaw in a web application server. The vulnerability had been published and a patch provided by the software vendor, but it was not properly applied throughout the Equifax environment. This unpatched software vulnerability allowed attackers to break in and steal data; estimates of the costs associated with this breach top $1 billion, including fines, lawsuits, and an overhaul of the company's information security program!

Source: www.bankinfosecurity.com/equifaxs-data-breach-costs-hit-14-billion-a-12473

Performance and Capacity Monitoring

Monitoring is a critical concern for all parties in cloud computing. The CSP should implement monitoring to ensure they are able to meet customer demands and promised capacity, and consumers need to perform monitoring to ensure service providers are meeting their obligations and that the organization's systems remain available for users.

The majority of monitoring tasks will be in support of the availability objective, though indicators of an attack or misuse may be revealed as well, such as spikes in processor utilization that could be caused by cryptocurrency mining malware. Alerts should be generated based on established thresholds, and appropriate action plans initiated in the event of an event or disruption. Monitoring is not necessarily designed to detect incidents, however. It is also critical for CSPs to measure the amount of services being used by customers so they can be billed accurately.

Infrastructure elements that should be monitored include the following:

- **Network**: Cloud computing is, by design, accessible via networks, so it is essential to monitor their performance and availability. Bandwidth utilization, link state (up or down), and number of dropped packets are examples of metrics that might be captured.
- **Compute**: Though traditionally defined as CPU utilization, often measured in cores or number of operations, compute can also include other forms of processing such as GPUs, field-programmable gate arrays (FPGA), or even service calls to an API or microservices architecture, where the service is billed based on the number of requests made.

- **Storage and memory**: Data storage and memory are often measured in terms of total amount used, but other measures also exist such as number of reads and writes (input output operations, or IOPS), as well as speed of data access. Speed is a critical measure for many cloud services that are priced based on data retrieval, allowing cloud consumers to realize cost savings by storing infrequently needed data in a slower environment, while keeping essential data more quickly accessible in a more expensive environment.

Regardless of what is being monitored and who performs it, adequate staffing is critical to make monitoring effective. Just as reviews make log files impactful, appropriate users of performance data are also essential. If a metric is captured but the cloud consumer never reviews it, they run the risk of a service being unavailable with no forewarning or paying for services that were not actually usable. CSPs also face risks of customers complaining, customers refusing to pay, and loss of reputation if their services are not routinely available.

Hardware Monitoring

Although cloud computing relies heavily on virtualization, at some point physical hardware is necessary to provide all the services. Monitoring this physical hardware is essential, especially for availability as hardware failures can have an outsized impact in virtualization due to multiple VMs relying on a single set of hardware.

Various tools exist to perform physical monitoring, and the choice will depend on the hardware being used as well as organizational business needs. Similar to capacity monitoring, all hardware under monitoring should have alert thresholds and response actions if a metric goes outside expected values, such as automated migration of VMs from faulty hardware or an alert to investigate and replace a faulty component. Hardware targets for monitoring may include the following:

- **Compute hardware and supporting infrastructure**: The CPU, RAM, fans, disk drives, network gear, and other physical components of the infrastructure have operating tolerances related to heat and electricity. Monitoring these devices to ensure they are within tolerance is a critical aspect of ensuring availability; a CPU that overheats or a power supply that sends too much wattage can shorten the lifespan of a component or cause it to fail altogether. Fan speed can also be used as an indicator of workload; the harder a system is working, the more heat it produces, which causes fans to spin faster.

  Some devices also include built-in monitoring, such as hard drives that support self-monitoring, analysis, and reporting technology (SMART). SMART drives monitor a number of factors related to drive health and can identify when a failure is imminent. Many SSDs also provide reporting when a set number of sectors have failed, which means the drive is reaching the end of its useful life.

Especially in large environments, storage tools like storage area networks (SAN) will include their own health and diagnostic tools; the information from these should be integrated into the organization's continuous monitoring strategy.

- **Environmental**: All computing components generate heat and are generally not designed for use in very wet environments or in direct contact with water. Monitoring environmental conditions including heat, humidity, and the presence of water can be critical to detecting issues early.

The types of monitoring tools in use will depend on a number of factors. Many vendors of cloud-grade hardware such as SAN controllers or virtualization clusters include diagnostic and monitoring tools. The usefulness of these built-in tools may be limited if your organization's measurement needs require data that is not captured, in which case a third-party tool may be required.

In general, hardware monitoring will be the purview of the CSP and not the consumer, as the CSP is likely to retain physical control of hardware. Data center design and infrastructure management are entire fields of endeavor largely outside the scope of the CCSP, but it is obviously important for a CSP to have appropriately skilled team members. The Uptime Institute `(uptimeinstitute.com/tiers)` is one resource that provides guidance and education on designing and managing infrastructure; another is the International Data Center Authority `(www.idc-a.org)`.

Configuration of Host and Guest Operating System Backup and Restore Functions

There is a clear delineation of responsibility between the CSP and consumer when it comes to configuring, testing, and managing backup and restoration functions in cloud environments. In SaaS cloud models, the CSP retains full control over backup and restore and will often be governed by SLA commitments for restoration in the event of an incident or outage.

In the PaaS model, there will be backup and restoration responsibilities for both the consumer and the CSP, especially for VMs. The CSP is responsible for maintaining backup and restoration of the host OS and any hypervisor software and for ensuring the availability of the system in line with agreed-upon service levels.

Backup and recovery of individual VMs in IaaS are the responsibility of the consumers and may be done in a variety of ways. This might include full backups, snapshots, or definition files used for infrastructure as code deployments. Regardless of which backup method is utilized, a number of security considerations should be taken into account.

- Sensitive data may be stored in backups, particularly in snapshot functions that do not support the same access controls or encryption as the OS they are created from. In this case, access to snapshots and need-to-know principles must be considered when designing the backup plan.

- Snapshots and backups may be created on the same physical hardware as the running VMs, which violates a core principle of data backups: physical separation. As a best practice, these should be stored on different hardware (if possible) or in a different availability zone to ensure an incident affecting the main environment does not also impact the backup data.
- Integrity of all backups should be verified routinely to ensure they are usable. Snapshots or backups should also be updated as system changes occur, especially patches or major configuration changes.

Configuration of resiliency functions, such the use of automatic data replication, failover between availability zones offered by the CSP, or the use of network load balancing, will always be the responsibility of the consumer. The CSP is responsible to maintain the capabilities that enable these options, but the consumer must architect their cloud environment, infrastructure, and applications appropriately to meet their own resiliency objectives.

Network Security Controls

Cloud environments are inherently network accessible, so the security of data in transit between the consumer and the CSP is a critical concern, with both parties sharing responsibility for architecting secure networks. CSPs must ensure that they adequately support the networks they are providing in the cloud service environment, and consumers are responsible, in some cloud service models, for architecting their own secure networks using a combination of their own tools and those provided by the CSP.

One major concern related to network security is the ability of some tools to function in a cloud paradigm. The early days of virtualization brought challenges for many security tools that relied on capturing network traffic as it flowed across a switch; the devices were attached to a SPAN or mirror port on the switch and received a copy of all traffic for analysis. VMs running on the same physical host did not need to send traffic outside the host to communicate, rendering tools listening for traffic on a switch useless. Complex software-defined networks (SDNs) that can span multiple data centers around the world likely require more advanced solutions, and security practitioners must be aware of these challenges.

Firewalls

Firewalls are most broadly defined as a security tool designed to isolate and control access between segments of a network, whether it is an internal network and the public Internet or even between environments such as an application with highly sensitive data and other internal apps. Firewalls operate by inspecting traffic and making a decision whether to forward the traffic (allow) or drop it (deny) and are often used to isolate or *segment*

networks by controlling what network traffic is allowed to flow between segments. There are a variety of firewall types.

- **Static packet or stateless**: These are the original type of firewall, designed to inspect network packets and compare them against a ruleset. For example, the firewall might see TCP destination port 23 (Telnet) in a packet's headers and decide to drop the packet since Telnet is an insecure protocol. Static firewalls operate quickly but can struggle with complex situations like voice over IP (VoIP) and videoconferencing where a call is initiated on one port, but the actual call traffic is exchanged on a different port negotiated for that communication session.
- **Stateful**: These are an evolution of static firewalls and offer the ability for the firewall to understand some context regarding communication (known as a *state*). For example, the firewall might allow traffic on a random high-number port from a particular host if it had also seen previous traffic on port 20/21 (FTP). Since FTP clients often negotiate a custom port to be used for a specific file transfer, this traffic makes sense in the context of previous traffic; without the previous traffic on port 20, the firewall would block the traffic on the high-number port. In this case, the firewall has more intelligence and flexibility to make decisions, but with a higher processing overhead and cost.
- **Web application firewall (WAF) and API gateway**: These are a highly specialized form of network access control devices that are designed to handle specific types of traffic, unlike a generic firewall that can handle any network traffic. WAFs and API gateways allow for the analysis of traffic destined specifically for a web application or an application's API and can be useful in detecting more complex attacks such as SQL injection, which could not be identified by looking at raw network traffic. These devices apply a set of rules to HTTP conversations and look for anomalous interactions with the application.
- **Security groups**: In SDNs, it can be difficult or impossible to locate specific infrastructure elements of the network. For example, a globally load-balanced application may exist in several data centers all over the world, and it would be a headache to place firewalls at the edge of each data center's network. Security groups (also called *network security groups*, or NSGs) are an abstraction layer that allows a consumer to define protections required, and the CSP's infrastructure deploys appropriate virtualized resources as needed. In the previous example, a cloud consumer defines a set of allowed traffic for the application, and the CSP's hardware and software will be configured uniquely for each data center to implement those rules. This is a similar concept to infrastructure as code, where hardware variations between data centers are abstracted and presented to the consumer in a virtual, self-serve manner.

- **Next-generation firewalls (NGFW)**: Although more of a marketing term than a unique type of firewall, NGFWs combine multiple firewall functions into a single device, such as a stateful firewall and API gateway. Many NGFWs also include other network security protections such as intrusion detection or VPN services.

Firewalls may be hardware appliances that would traditionally be deployed by the CSP, or virtual appliances that can be deployed by a cloud consumer as a VM. Host-based firewalls, which are software-based, are also often considered a best practice in a layered defense model. In the event a main network firewall fails, each host still has some protection from malicious traffic, though all devices obviously need to be properly configured.

There are a number of cloud-specific considerations related to firewall deployment and configuration, such as the use of security groups for managing network-level traffic coupled with host-based firewalls to filter traffic to specific hosts. This approach is an example of *microsegmentation*, which amounts to controlling traffic on a granular basis—often at the level of a single host. In a cloud environment, an NSG might block traffic on specific ports from entering a DMZ, and then the host firewalls would further restrict traffic reaching a host based on ports or protocols. Traditional firewall rules may also be ineffective in a cloud environment, which necessitates these new approaches. In an auto-scaling environment, new hosts are brought online and dynamically assigned IP addresses. A traditional firewall would need its ruleset updated to allow traffic to these new hosts; otherwise, they will not be able to handle traffic at all. The newly created resources can be automatically placed into the proper security group with no additional configuration required.

Intrusion Detection/Intrusion Prevention Systems (IDS/IPS)

As the name implies, an intrusion detection system (IDS) is designed to detect a system intrusion when it occurs. An intrusion prevention system (IPS) is a bit of a misnomer, however—it acts to limit damage once an intrusion has been detected. The goal in both cases is to limit the impact of an intrusion, either by alerting personnel to an intrusion so they can take remedial action or by automatically shutting down an attempted attack.

An IDS is a passive device that analyzes traffic and generates an alert when traffic matching a pattern is detected, such as a large volume of unfinished TCP handshakes. IPS goes further by taking action to stop the attack, such as blocking traffic from the malicious host with a firewall rule, disabling a user account generating unwanted traffic, or even shutting down an application or server that has come under attack.

Both IDS and IPS can be deployed in two ways, and the choice of deployment as well as location are critical to ensure the devices can see all traffic they require to be effective.

A network-based intrusion detection system/ intrusion prevention system (NIDS/NIPS) sits on a network where it can observe all traffic and may often be deployed at a network's perimeter for optimum visibility. Similar to firewalls, however, NIDS/NIPSs may be challenged in a virtualized environment where network traffic between VMs never crosses a switch. A host-based intrusion detection system/ intrusion prevention system (HIDS/HIPS) is deployed on a specific host to monitor traffic. While this helps overcome problems associated with invisible network traffic, the agents required introduce processing overhead, may require licensing costs, and may not be available for all platforms an organization is using.

Honeypots and Honeynets

Honeypots and honeynets can be useful monitoring tools if used appropriately. They should be designed to detect or gather information about unauthorized attempts to gain access to data and information systems, often by appearing to be a valuable resource. In reality, they contain no sensitive data, but attackers attempting to access them may be distracted or deflected from high-value targets or give up information about themselves such as IP addresses.

In most jurisdictions, there are significant legal issues concerning the use of honeypots or honeynets, centered around the concept of *entrapment*. This legal concept describes an agent inducing a person to commit a crime, which may be used as a defense by the perpetrator and render any attempt to prosecute them ineffective. It is therefore imperative that these devices never be set up with an explicit purpose of being attractive targets or designed to "catch the bad guys."

Vulnerability Assessments

Vulnerability assessments should be part of a broader vulnerability management program, with the goal of detecting vulnerabilities before an attacker finds them. Many organizations will have a regulatory or compliance obligation to conduct vulnerability assessments, which will dictate not only the schedule but also the form of the assessment. An organization with an annual PCI assessment requirement should be checking for required configurations and vulnerabilities related to credit cardholder data, while a medical organization should be checking for required protected health information controls and vulnerabilities.

Vulnerability scanners are an often-used tool in conducting vulnerability assessments and can be configured to scan on a relatively frequent basis as a detective control. Human vulnerability assessments can also be utilized, such as an internal audit function or standard reviews like access and configuration management checks. Even a physical walkthrough of a facility to identify users who are not following clean desk or workstation

locking policies can uncover vulnerabilities, which should be treated as risks and remediated.

A more advanced form of assessments an organization might conduct is penetration or pen testing, which typically involves a human tester attempting to exploit any vulnerabilities identified. Vulnerability scanners typically identify and report on software or configuration vulnerabilities, but it can be difficult to determine if a particular software vulnerability could actually be exploited in a complex environment. The use of vulnerability scanners and pen testers may be limited by your CSP's terms of service, so a key concern for a CCSP is understanding the type and frequency of testing that is allowed.

Management Plane

The management plane is mostly used by the CSP and provides virtual management options analogous to the physical administration options a legacy data center would provide, such as powering VMs on and off or provisioning virtual infrastructure for VMs such as RAM and storage. The management plane will also be the tool used by administrators for tasks such as migrating running VMs to different physical hardware before performing hardware maintenance.

Because of the functionality it provides, the management plane requires appropriate logging, monitoring, and access controls, similar to the raised floor space in a data center or access to domain admin functions. Depending upon the virtualization toolset, the management plane may be used to perform patching and maintenance on the virtualization software itself. Functionality of the management plane is usually exposed through an API, which may be controlled by an administrator from a command line or via a graphical interface.

A key concept related to the management plane is *orchestration*, or the automated configuration and management of resources. Rather than requiring an administrator to individually migrate VMs off a cluster before applying patches, the management plane can automate this process. The admin schedules a patch for deployment, and the software comprising the management plane coordinates moving all VMs off the cluster, preventing new VMs from being started, and then enters maintenance mode to apply the patches.

The cloud management *console* is often confused with the cloud management *plane*, and in reality, they perform similar functions and may be closely related. The *management console* is usually a web-based console for use by the cloud consumer to provision and manage their cloud services, though it may also be exposed as an API that customers can utilize from other programs or a command line. It may utilize the management plane's API for starting/stopping VMs or configuring VM resources such as RAM and network access, but it should not give a cloud consumer total control over

the entire CSP infrastructure. The management plane's access controls must enforce minimum necessary authorization to ensure each consumer is able to manage their own infrastructure and not that of another customer.

IMPLEMENT OPERATIONAL CONTROLS AND STANDARDS

IT service management (ITSM) frameworks consist of operational controls designed to help organizations design, implement, and improve IT operations in a consistent manner. They can be useful in speeding up IT delivery tasks, providing more consistent oversight, and are also critical to processes where elements of security risk management are implemented. Change management is one example; it helps the organization to maintain a consistent IT environment that meets user needs and also implements security controls such as a change control board where the security impact of changes can be adequately researched and addressed.

The two standards that a CCSP should be familiar with include ISO 20000-1 (not to be confused with ISO 27001) and ITIL (formerly an acronym meaning Information Technology Infrastructure Library). Both frameworks focus on the process-driven aspects of delivering IT services to an organization, such as remote collaboration services, rather than focusing on just delivering IT systems like an Exchange server. In ITIL, the set of services available is called a *service catalog*, which includes all the services available to the organization.

Both frameworks start with the need for policies to govern the ITSM processes, which should be documented, well understood by relevant members of the organization, and kept up-to-date to reflect changing needs and requirements. ISO 20000-1 and ITIL emphasize the need to deeply understand user needs and also focus on gathering feedback to deliver continuous service improvement. Stated another way, those in charge of IT services should have a close connection to the users of the IT system and strive to make continual improvements; in this regard, it is similar to the Agile development methodology.

Change Management

Change management is concerned with keeping the organization operating effectively, even when changes are needed such as the modification of existing services, addition of new services, or retirement of old services. To do this, the organization must implement a proactive set of formal activities and processes to request, review, implement, and document all changes.

Many organizations utilize a ticketing system to document all steps required for a change. The first step is initiation in the form of a change request, which should capture details such as the purpose of the proposed change, the owner, resources required, and any impacts that have been identified, such as downtime required to implement the change or impacts to the organization's risk posture.

The change then goes for a review, often by a change control or change advisory board (CCB or CAB). This review is designed to verify if the proposed change offers business benefits/value appropriate to its associated costs, understand the impact of the change and ensure it does not introduce unacceptable levels of risk, and, ideally, confirm that the change has been properly planned and can be reversed or rolled back in the event it is unsuccessful. This step may involve testing and additional processes such as decision analysis, and it may be iterated if the change board needs additional information from the requestor.

Once a change has been approved, it is ready for the owner to execute the appropriate plan to implement it. Since many changes will result in the acquisition of new hardware, software, or IT services, there will be a number of security concerns that operate concurrently with the change, including acquisition security management, security testing, and the use of the organization's certification and accreditation process if the change is large enough. In the event a change is not successful, fallback, rollback, or other restoration actions need to be planned to prevent a loss of availability.

Not all changes will be treated the same, and many organizations will implement different procedures based on categories of changes.

- **Low-risk**: These are changes that are considered unlikely to have a negative impact and are therefore pre-authorized to reduce operational overhead. Examples include application of standard patches, addition of standard assets to address capacity (for example, deploying a standard server build to provide additional processing capability), or installation of approved software that is not part of a standard baseline but is required for a particular job function.
- **Normal changes**: These changes require the full change management request-review-implement process. They will typically follow a routine schedule based on the meeting schedule of the change board.

✔ Automating Change Management

In the case of a continuous integration/continuous deployment (CI/CD) software development environment, change reviews may be automated when new code is ready for deployment, particularly security testing such as code analysis. The goal of this automation is to reduce operational overhead while still adequately managing the risk associated with the change (in this case, new software).

- **Emergency changes**: Things happen unexpectedly, and the process of requesting, testing, and receiving approval for a change may require too much time. For changes required to resolve an incident or critical security concern, formal procedures should be utilized to implement the change as needed to address the incident and document all details related to the change. Depending on the organization, a faster or less cumbersome change control decision process may be utilized; for example, only one CCB member approval is required, or the change may be reviewed and approved retroactively.

Continuity Management

Continuity is concerned with the availability aspect of the CIA triad and is a critical consideration for both cloud customers and providers. Continuity management addresses the reality that, despite best efforts and mitigating activities, sometimes adverse events happen. How the organization responds should be planned and adequate resources identified prior to an incident, and the business continuity policy, plan(s), and other documentation should be readily available to support the organization's members during an interruption.

It is essential for both cloud customers and providers to do the following:

- **Identify critical business functions and resources.** This is usually accomplished by conducting a business impact assessment (BIA), which assists the organization to understand its essential assets and processes. For a customer, this may be business-critical applications, while for the provider, it will be the infrastructure and other resources required to deliver the cloud services. The BIA is a structured method of identifying what impact a disruption of critical business functions poses to the organization, as well as the resources necessary to recover a minimum level of functionality.
- **Prioritize recovery.** Not all systems or assets can be recovered all at once, so it is essential that the organization develop a prioritization of critical processes that are essential to the continued functioning of its operations and identify which assets are essential to those processes. The set of ordered steps should also be documented to ensure dependencies are documented and restored in the correct order; for example, that power to a facility is restored before any information systems can be operated in it.
- **Plan continuity.** This will entail identifying continuity capabilities such as automated failovers, as well as understanding relevant cloud offerings and how they are to be used. In some cases, the cloud will function as a backup for an organization's on-prem infrastructure, while in other cases, the cloud's availability features will be utilized, such as different availability regions around the world,

automatic duplication of data to multiple sites, etc. The cloud customer is responsible for understanding the availability features of their chosen cloud provider and properly architecting their cloud applications to meet continuity requirements.

- **Document plans and procedures.** Issues that cause a loss of availability are often high-stress situations, such as a natural disaster. In these instances, it is preferable to have employees working from a previously prepared set of instructions rather than trying to think and respond on-the-fly. Decision-makers and standard processes/tools may be unavailable, so appropriate alternatives should be documented.

NOTE Training is also critical in high-stress continuity situations. Members with key responsibilities should receive training to ensure they are aware of what they need to do in an emergency, rather than trying to make crucial decisions under duress.

There are a variety of standards related to continuity management; these may be useful to the organization in planning, testing, and preparing for contingency circumstances. Many legal and regulatory frameworks mandate the use of a particular standard depending on an organization's industry or location. The CCSP should be aware of the relevant framework for their industry; the following are some key frameworks:

- **NIST Risk Management Framework and ISO 27000**: Since both frameworks focus on information security concerns, both deal with business continuity and disaster recovery (BCDR), terms that fall under the larger category of *continuity management*. In the NIST framework, a family of controls called contingency planning are specified depending on the system's risk profile, while the ISO 27002 framework specifies information security aspects of business continuity management.
- **Health Insurance Portability and Accountability Act (HIPAA)**: Healthcare data in the United States is governed by this standard, which mandates adequate data backups, disaster recovery planning, and emergency access to healthcare data in the event of a system interruption.
- **ISO 22301:2019 Security and resilience — Business continuity management systems**: This specifies the requirements needed for an organization to plan, implement and operate, and continually improve the continuity capability. This includes adequate support from leadership within the organization, planning resources for managing continuity, and steps to implement/operate the program such as conducting a BIA, exercising contingency plans, and monitoring the capability's effectiveness.

Information Security Management

The goal of an information security management system (ISMS) is to ensure a coherent organizational approach to managing information security risks; stated another way, it is the overarching approach an organization takes to preserving the confidentiality, integrity, and availability (the CIA triad) of systems and data in use. The operational aspects of an ISMS include standard security risk management activities in the form of security controls such as encryption, as well as supporting business functions required for the organization to achieve risk management goals like formal support and buy-in from management, skills and training, and adequate oversight and performance evaluation.

Various standards and frameworks exist to help organizations implement, manage, and, in some cases, audit or certify their ISMS. Most contain requirements to be met in order to support goals of the CIA triad, as well as best practices for implementation and guidance on proper use and operation of the framework. While many security control frameworks exist, not all are focused on the larger operational task of implementing an ISMS. Payment Card Industry Data Security Standard (PCI DSS), for example, focuses specifically on securing cardholder data that an organization is processing. Frameworks that focus on both security controls as well as the overall ISMS functions include the following:

- **ISO 27000 series**: The ISO 27001 standard provides a set of requirements for an organization's ISMS. It is often confused with ISO 27002, which is the code of practice for information security controls, due to the inclusion of the 27002 set of controls as an appendix to 27001. The two are interrelated, because 27001 sets out the high-level requirements an organization must meet to provide leadership for, plan, implement, and operate an ISMS.
  - ISO 27002 provides a set of controls and implementation guidance, broken down by domains such as Asset Management and Cryptography. 27001 and 27002 provide a road map for an organization to understand its security risk posture and implement appropriate security controls to mitigate and manage those risks.
  - There are additional ISO standards that provide guidance on implementing and managing security controls in cloud-specific environments. ISO 27017 is a "Code of practice for information security controls based on ISO/IEC 27002 for cloud services," and ISO 27018 is a "Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors." Both documents enhance/extend the guidance offered in ISO 27002 controls to deal with particular security risk challenges in cloud implementation.

- ISO 27701 is also relevant to cloud security environments where personally identifiable information (PII) is being handled. 27701 extends the ISMS guidance in 27001 to manage risks related to privacy, by implementing and managing a privacy information management system (PIMS). As many privacy regulations require security controls and risk management, this standard will be relevant to a CCSP whose organization acts as a data owner or processor.

- **NIST RMF, SP 800-53, and CSF**: Although it uses different terminology, the NIST Risk Management Framework (RMF) specified in Special Publication (SP) 800-37, *Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy*, has the same objective as the ISO-defined ISMS: identifying information security risks and applying adequate risk mitigations in the form of security controls. SP 800-37 provides the guidance for creating the organization's risk management framework and points to NIST SP 800-53, *Security and Privacy Controls for Federal Information Systems and Organizations*, for the control requirements and implementation guidance.

  While the NIST RMF and SP 800-53 standards are mandated for use in many parts of the U.S. federal government, they are free to use for any organization. The NIST Cybersecurity Framework (CSF) was originally designed to help private-sector critical infrastructure providers design and implement information security programs; however, its free-to-use and relatively lightweight approach have made it a popular ISMS tool for many nongovernment organizations.

- **AICPA SOC 2**: The Service Organization Controls (SOC 2) framework has seen wide adoption among cloud service providers for a variety of reasons, primarily the relatively lightweight approach it provides as well as the use of a third party to perform audits, which provides increased assurance for business partners and customers. While not as robust as the ISO and NIST frameworks, SOC 2 contains Trust Services Criteria (TSC), which cover organizational aspects including Security, Availability, Processing Integrity, Confidentiality, and Privacy. The Common Criteria apply to all organizations and contain similar requirements to other frameworks: executive support and buy-in, assessment and treatment of risks, monitoring of controls, and implementation guidance. The other TSCs such as Availability, may be implemented at the discretion of the organization; typically, an organization's service offering will drive which TSCs are chosen.

Continual Service Improvement Management

Most ITSM models include some form of monitoring capability utilizing functions such as internal audit, external audit and reporting, or the generation of security metrics and

management oversight of processes via these metrics. The organization's IT services, including the ISMS and all related processes, should be monitored for effectiveness and placed into a cycle of continuous improvements. The goals of this continuous improvement program should be twofold: first to ensure the IT services (including security services) are meeting the organization's business objectives, and second to ensure that the organization's security risks remain adequately mitigated.

One critical element of continual service improvement includes elements of monitoring and measurement, which often take the form of security metrics. Metrics can be tricky to gather, particularly if they need to be presented to a variety of audiences. It may be the case that business leaders will be less interested in deeply technical topics, which means the metrics should be used to aggregate information and present it in an easily understood, actionable way.

For instance, rather than reporting a long list of patches and Common Vulnerabilities and Exposures (CVEs) addressed (undoubtedly an important aspect of security risk management), a more appropriate metric might be the percentage of machines patched within the defined timeframe for the criticality of the patch; for example, 90 percent of machines were patched within seven days of release. Acceptable values should also be defined, which allows for key performance indicators (KPIs) to be reported. In this example, the KPI might be red (bad), if the organization's target is 99 percent patch deployment within seven days of release—a clear indicator to management that something needs their attention.

There are other sources of improvement opportunity information as well, including audits and actual incidents. Audits may be conducted internally or externally, and findings from those audits can be viewed as improvement opportunities. Actual incidents, such as a business interruption or widespread malware outbreak, should be concluded with a lessons learned or postmortem analysis, which provides another source of improvement opportunities. The root cause of the incident and any observations made during the recovery can be used to improve the organization's IT security services.

Incident Management

It is important to understand the formal distinction between *events* and *incidents* as the foundation for incident management.

- **Events** are any observable item, including routine actions such as a user successfully logging into a system, a file being accessed, or a system being unavailable during a scheduled maintenance window. Many routine events will be logged but do not require any additional actions.
- **Incidents**, by contrast, are events that are both unplanned and have an adverse impact on the organization. Incidents typically require investigation and remedial action by some combination of IT, operations, and security personnel. Examples

of incidents include unexpected restart of a system, ransomware preventing users from accessing a system, or a loss of network connectivity to a cloud service.

All incidents should be investigated and remediated as appropriate to restore the organization's normal operations as quickly as possible and to minimize adverse impact to the organization such as lost productivity or revenue. This resumption of normal service is the primary goal of incident management.

Not all incidents will require participation by the security team. For example, a spike in new user traffic to an application after a marketing campaign goes live, which leads to a partial loss of availability, is an operational issue and not a security one. A coordinated denial-of-service attack by a foreign nation-state, however, is an incident that requires participation by both IT and security personnel to successfully remediate.

All organizations require some level of incident management capability; that is, the tools and resources needed to identify, categorize, and remediate the impacts of incidents. This capability will revolve around an incident management plan, which should document the following:

- Definitions of incident types, such as internal operational incidents, security incidents, and cloud provider incidents.
- The incident response team (IRT) personnel. Note the composition of this team will be dependent upon the type of incident, but an incident response coordinator should always be appointed to assess the situation and identify the requisite team members on a per-incident basis.
- Roles and responsibilities for the IRT personnel in each incident type. This should include roles internal to the organization, as well as responsibilities of external stakeholders such as law enforcement, business partners or customers, and the cloud provider if the incident may affect them.
- Resources required such as operations or security management tools to facilitate detection and response, such as a security information and event management (SIEM) or IDS, and required personnel.
- Incident management processes following a logical lifecycle from detection of the incident to response to restoration of normal service. The response coordinator should determine relevant response requirements, including the following:
  - Communications appropriate to the specific incident, both internal and external, with stakeholders including customers, employees, executive management, law enforcement, regulatory bodies, and possibly the media
  - Any required breach or privacy law notifications, if the incident involved PII or other regulated data

A variety of standards exist to support organizations developing an incident response capability, including the ITIL framework, NIST Special Publication 800-61 *Computer Security Incident Handling Guide*, and ISO 27035, *Security incident management*. All standards implement a lifecycle approach to managing incidents, starting with planning before an incident occurs, activating the response, and following the documented steps, and ending with reporting on the results and any lessons learned to help the organization better mitigate or respond to such incidents in the future. Figure 5.1 shows an example of the NIST SP 800-61 lifecycle and a description of the activities.

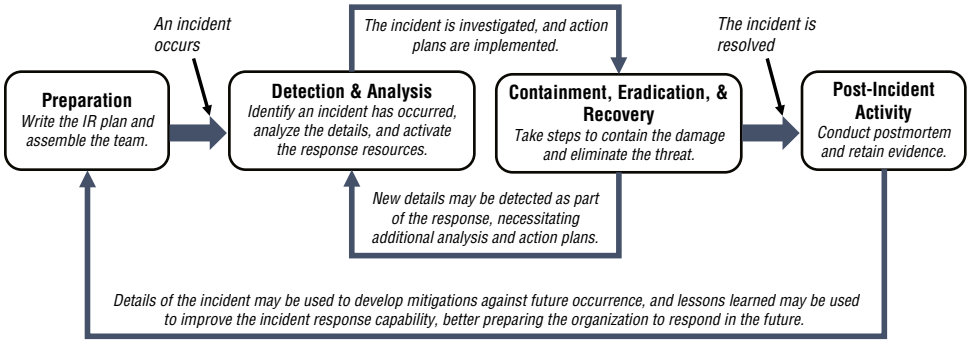

FIGURE 5.1 **NIST incident response lifecycle phases**

The CCSP's role in developing the capability is, obviously, the responses required for security incidents, while other stakeholders from IT or operations will provide input relevant to their responsibilities. All incident response frameworks emphasize the importance of planning ahead for incidents by identifying likely scenarios and developing response strategies before an incident occurs, as incidents can be a high-stress situation and ad hoc responses are less preferable than pre-planned, rehearsed responses.

As with all aspects of cloud service use, there are shared responsibilities between the CSP and the consumer when responding to incidents. For many incidents, the CSP will not be involved; for example, an internal user at a consumer organization breaches policies to misuse company data. This does not impact the CSP; however, some incidents will require coordination, such as a denial-of-service attack against one consumer, which could impact other consumers by exhausting resources. The CSP could also suffer an incident, like an outage of a facility or theft of hardware resources, which must be reported to consumer organizations and may trigger their incident management procedures. The major CSPs have dedicated incident management teams to coordinate incident responses, including resources designed to provide notice to consumers such as a status page or dedicated account managers. Your incident planning must include

coordination with your CSPs, and you should be aware of what capabilities are and are not available to you. Many CSPs forbid physical access to their resources during an incident response unless valid law enforcement procedures, such as obtaining a warrant, have been followed.

Another important aspect of the organization's incident management capability is the proper categorization and prioritization of incidents based on their impact and criticality. Incident management seeks to restore normal operations as quickly as possible, so prioritizing incidents and recovery steps is critical. This is similar to the risk assessment process where risks are analyzed according to their impact and likelihood; however, since incidents have already occurred, they are measured by the following:

- **Criticality/impact**: The effect the incident will have on the organization, often measured as low/moderate/high/critical.
- **Urgency**: The timeframe in which the incident must be resolved to avoid unwanted impact. For example, unavailable systems that do not impact life, health, or safety will always be less urgent than systems that do impact these factors; therefore, they should be prioritized lower for resolution.

Many organizations utilize a P score (P0–P5) to categorize incidents. Members of the incident response team use this score to prioritize the work required to resolve an incident. For example, a P5 or low-priority item may be addressed as time permits, while a P0, which equates to a complete disruption of operations, requires that all other work be suspended. In many organizations, a certain priority rating may also be a trigger for other organizational capabilities, such as the invocation of a business continuity or disaster recovery plan if the incident is sufficiently disruptive to normal operations.

Problem Management

In the ITIL framework, problems are the causes of incidents or adverse events, and the practice of problem management seeks to improve the organization's handling of these incidents. Problems are, in essence, the root cause of incidents, so problem management utilizes root-cause analysis to identify the underlying problem or problems that lead to an incident and seeks to minimize the likelihood or impact of incidents in the future; it is therefore a form of risk management.

Identified problems are tracked as a form of common knowledge, often in a known issues or known errors database. These document an identified root cause that the organization is aware of, as well as any shared knowledge regarding how to fix or avoid them. Examples might include a set of procedural steps to follow when troubleshooting a particular system, or a *workaround*, which is a temporary fix for an incident. Workarounds do not mitigate the likelihood of a problem occurring but do provide a quick fix, which

supports the incident management goal of restoring normal service as quickly as possible. Problems are risks to the organization, and if the workarounds do not provide sufficient risk mitigation, then the organization should investigate a more permanent solution to resolve the underlying cause.

Release Management

The last few years have seen an enormous shift from traditional release management practices due to widespread adoption of Agile development methodologies. The primary change is the frequency of releases due to the increased speed of development activities in continuous integration/continuous delivery, often referred to as a CI/CD pipeline. Under this model, developers work on small units of code and merge them back into the main branch of the application's code as soon as they are finished.

Inherent in a CI/CD pipeline is the concept of automated testing, which is designed to more quickly identify problems in an easier-to-solve way. By running tests on only small units of code being integrated, it's easier for the developers to identify where the problem is and how to fix it. From the user's perspective, they get access to new features more quickly than waiting for a monolithic release and, ideally, get fewer bugs due to the faster feedback to developers that automated testing offers.

Release management activities typically comprise the logistics needed to release the changed software or service and may include identifying the relevant components of the service, scheduling the release, and post-implementation reviewing to ensure the change was implemented as intended, i.e., that the new software or service is functioning as intended. The process has obvious overlap with change management processes.

In the Agile methodology, the process of release management may not involve manual scheduling, instead relying on the organization's standard weekly, monthly, or quarterly release schedule. For organizations using other development methodologies, the process for scheduling the deployment might require coordination between the service provider and consumers to mitigate risks associated with downtime, or the deployment may be scheduled for a previously reserved maintenance window during which customer access will be unavailable.

Once scheduled, the release manager must perform a variety of tasks, including identifying whether all changes to be released have successfully passed required automated tests as well as any manual testing requirements. Other manual processes such as updating documentation and writing release notes may also be part of the organization's release management activities. Once all steps have been completed, the release can be deployed and tested and is ready for users.

Deployment Management

In more mature organizations, the CD in CI/CD stands for continuous deployment, which automates the process of release management to deliver a truly automatic CI/CD pipeline. Once a developer has written their code and checked it in, an automated process is triggered to test the code, and if all tests pass, it is integrated and deployed automatically to users. This has the advantage of getting updated software and services deployed to users quickly and offers security benefits as well. For example, automated testing is typically less expensive than manual testing, making it feasible to conduct more tests and increase the frequency in complement to the organization's manual testing plans.

Even organizations with continuous deployment may require some deployment management processes to deal with deployments that cannot be automated, such as new hardware or software. In this case, the release management process should develop a set of deployment steps including all required assets, dependencies, and deployment order to ensure the deployment process is successful.

One recent technology development supporting this trend of more frequent deployment is *containerization*, which packages application code and non-OS software the application requires into a container. Containers can be run on any computing platform regardless of underlying hardware or operating system, so long as container software such as the Docker Engine is available. The container software makes the resources of the computing environment available in response to the requirements of the containerized applications when they run, similar to the way virtualization makes hardware resources available to a virtualized guest OS.

Containers offer advantages of portability. In other words, they can be run on any OS and hardware platform with container software, as well as availability advantages over traditional infrastructure due to requiring fewer pieces of software to run. Continuous deployment pipelines often make use of containers, as they provide more flexibility and can speed development.

Deployment scheduling for noncontinuous environments may follow a set schedule such as a routine maintenance window or be deployed in a phased approach where a small subset of users receive the new deployment. This phased approach can offer advantages for riskier deployments such as large operating system updates, where unexpected bugs or issues may be encountered. By rolling out to a subset of the user pool, the impact of these bugs is reduced, and the organization has more opportunity to find and correct them. Organizations may also choose not to push deployments but instead allow users to pull the updated software/service on their own schedule. This is the model many consumer OSs follow, where an update is made available and users are free to accept or

delay the update at their discretion. This is advantageous for uptime as a user may not want to restart their machine in the middle of an important task, though it does lead to the problem of software never being updated, which means vulnerabilities fixed by that update are also not addressed.

✔ Immutable Infrastructure

Cloud services have significantly shifted the way many organizations build their information system environments. In traditional models, a physical location with utilities had to be built, appropriate equipment like servers and routers installed, and then operating systems and application software installed on top of that hardware. Once installed, software was placed into an operations and maintenance cycle where new patches or features were installed, but often the basic infrastructure was untouched for years, due to the time and cost involved. You wouldn't go out and build a whole new data center just to apply a minor patch! This led to sometimes "stale" assets such as hardware or software that hadn't been updated or that had drifted away from a secure configuration, which can introduce major vulnerabilities.

In cloud environments where all resources exist as virtual pools, there are fewer limitations on building virtual infrastructure, tearing it down, and rebuilding it. This gives rise to the idea of immutable architecture, which is built as needed, remains in a consistent state in line with a validated image during its (rather short) useful life, and then is destroyed and replaced by a new version when the system is next deployed. This helps overcome the problem of assets becoming stale by ensuring systems with up-to-date patches are always deployed and by blocking the infrastructure from being changed, which is why it's called *immutable*, which means unchangeable.

Configuration Management

Configuration management (CM, not to be confused with change management, which is also abbreviated CM) comprises practices, activities, and processes designed to maintain a known good configuration of something. *Configuration items* (CIs) are the things that are placed under configuration control and may be assets such as source code, operating systems, documentation, or even entire information systems.

Changes to CIs are usually required to go through a formal change process designed to ensure the change does not create an unacceptable risk situation, such as introducing

a vulnerable application or architecture. Part of the change management process must include updating the CMDB to reflect the new state of the services or components after the change is executed. For example, if a particular host is running Service Pack 1 (SP1) of an OS and a major upgrade to SP2 is performed, the CMDB must be updated once the change is completed successfully.

IT service CM may include hardware, software, or the cloud services and configurations in use by a consumer organization, while the CSP would also need to include configurations of the service infrastructure as well as the supply chain used to provide the services.

In many organizations, a formal CMDB will be used to track all CIs, acting as a system of record against which current configurations may be compared in order to detect systems that have gone out of line with expected configurations. The CMDB can also be useful for identifying vulnerabilities. As a source of truth for systems and software versions running in the organization, it is possible to query the CMDB to identify software running in the organization that contains a newly disclosed vulnerability. Furthermore, the CMDB can be useful to support audits by acting as a source of population data to allow auditors to choose a subset of systems to review. If the CMDB is not updated after changes are made, the organization is likely to face audit findings stemming from failure to properly follow processes.

Due to the type of information contained in a CMDB, such as version numbers, vendor information, hardware components, etc., it can also be used as the organization's asset inventory. Many tools that provide CMDB functionality can also be used to automatically detect and inventory systems, such as by monitoring network records to identify when a new system joins or by integrating with cloud administrative tools and adding any new cloud services that are invoked to the CMDB.

Checklists or baseline are often mentioned in discussions of CM, primarily as starting points or guidance on the desired secure configuration of particular system types. Configuration checklists are often published by industry or regional groups with specific guidance for hardening of operating systems like Windows, macOS, and various Linux distributions, as well as the hardening of popular applications such as Microsoft Office and collaboration tools. In many cases, the vendors themselves publish security checklists indicating how their products' various security settings can be configured. In all cases, these checklists are usually an input to an organization's CM process and should be tailored to meet the organization's unique needs.

✔ Infrastructure as Code

Similar to immutable architecture, which supports configuration management security goals for cloud consumers by preventing unwanted changes, infrastructure as code can help organizations ensure known good versions of infrastructure are deployed. This might include specially hardened OS or applications, network configurations, or other elements of infrastructure.

Infrastructure as code is a form of virtualization whereby system configuration information is written as a definition file that can be used by the cloud service to create machines with particular settings, rather than manual hardware and software configuration. In this way, developers can specify their application's needed infrastructure, and all instances will be configured according to that definition, ensuring that properly patched and configured systems are deployed automatically without the need for human intervention. While it's possible human error might occur in writing a definition file, it's less likely than an error in a mundane repetitive task like building and deploying servers.

Service Level Management

In the ITSM view of IT as a set of services, including information security, there is a function for defining, measuring, and correcting issues related to delivery of the services. In other words, performance management is a critical part of ITSM. This is closely related to the process of continual service improvement, and in fact, the same metrics are likely to be used by the organization to determine if services are meeting their defined goals.

Service level management rests on the organization's defined requirements for a service. The most common service level many cloud organizations encounter is availability, often expressed as a percentage of time that a system can be reached and utilized, such as 99.9 percent. A service with a 99.9 percent availability level must be reachable approximately 364.64 days per year; put another way, the system can only be down for less than 24 hours each year. Examples of other service levels that may be managed include number of concurrent users supported by a system, durability of data, response times to customer support requests, recovery time in the event of an interruption, and timeframes for deployment of patches based on criticality.

A key tool in managing service levels is the service level agreement (SLA), which is a formal agreement similar to a contract, but focused on measurable outcomes of the service being provided. This measurement aspect is what makes SLAs critical elements of security risk mitigation programs for cloud consumers, as it allows them to define, measure, and hold the cloud provider accountable for the services being consumed.

SLAs require routine monitoring for enforcement, and this typically relies on metrics designed to indicate whether the service level is being met. Availability metrics are often measured with tools that check to see if a service can be reached. For example, a script may run that checks to see if a website loads at a particular address. The script may run once an hour and log its results; if the SLA is 99.9 percent, then the service should not be down for more than nine hours in a given year. If the service level is not met, the SLA should define penalties, usually in the form of a refund or no obligation for the consumer to pay for the time the service was unavailable.

Defining the levels of service is usually up to the cloud provider in public cloud environments, though there is obviously a need to meet customer demands in order to win business. Requirements should be gathered for cloud service offerings regardless of the deployment model, and customer feedback should also be gathered and used as input to the continual service improvement. The metrics reported in SLAs are a convenient source of input to understand if the services are meeting customer's needs.

Availability Management

In cloud environments, the ability for users to reach and make use of the service are incredibly important, so the provider must ensure that adequate measures are in place to preserve the availability aspect of the relevant services. Availability and uptime are often used synonymously, but there is an important distinction. A service may be "up"—that is, reachable but not available—meaning it cannot be used. This could be the case if a dependency like the access control system is not available, so users can get to a login page for the cloud service but no further.

Due to the expansive nature of availability management, it is critical to view this as a holistic process. Factors that could negatively impact availability include many of the same concerns that an organization would consider in business continuity and disaster recovery, including loss of power, natural disasters, or loss of network connectivity.

There are additional concerns for providing a service that meets the agreed-upon service levels, including the issue of maintenance. Some cloud service providers exclude periods of scheduled maintenance from their availability guarantees. For example, a system will be available 99 percent of the time with the exception of the third Saturday of each month. This gives defined timeframes to make changes that require a loss of availability, as well as some flexibility for unexpected events or emergency maintenance outside the normal schedule.

Many of the tools that make cloud computing possible provide integral high availability options. For example, many virtualization tools support automatic moving of guest machines from a failed host in the event of an outage, or provide for load balancing so that sudden increases in demand can be distributed to prevent a denial of service. Many cloud services are also designed to be highly resilient, particularly PaaS and SaaS

offerings that can offer features like automatic data replication to multiple data centers around the world, or concurrent hosting of applications in multiple data centers so that an outage at one does not render the service unreachable.

Cloud consumers have a role to play in availability management as well. Consumers of IaaS will, obviously, have the most responsibility with regard to availability of their cloud environment, since they are responsible for virtually everything except the physical facility. PaaS and SaaS users will need to properly architect their cloud solutions to take advantage of their provider's availability options. For example, some cloud providers offer automatic data replication for cloud-hosted databases, but there may be configuration changes required to enable this functionality. There may be other concerns as well, such as data residency or the use of encryption, which can complicate availability; it is up to the cloud consumer to gather and understand these requirements and to configure their cloud services appropriately.

Capacity Management

In ITSM, one of the core concerns of availability is the amount of service capacity available compared with the amount being subscribed to. In a simple example, if a service has 100 active users but only 50 licenses available, that means the service is over capacity and 50 users will face a denial-of-service condition. In this simple example the service is oversubscribed, meaning there are more users than capacity. The service provider must be able to predict, measure, and plan for adequate capacity to meet its obligations; failure to do so could result in financial penalties in the form of SLA enforcement.

As users, we are aware of the negative impacts resulting from a lack of system resources—irritating situations such as a spinning hourglass or beach ball when our desktop computer's RAM capacity is exceeded. While a minor irritant for individual users, this situation could prove quite costly for a business relying on a cloud service provider's infrastructure. Any service that is being consumed should be measurable, whether it is network bandwidth, storage space, processing capability, or availability of an application. Measured service is one of the core elements of cloud computing, so metrics that illustrate demand for the service are relatively easy to identify.

Cloud service providers must take appropriate measures to identify the service capacity they need to provision. These measures might include analysis of past growth trends to predict future capacity, identifying capacity agreed to in SLAs, or even analysis of external factors such as knowing that a holiday season will cause a spike in demand at certain customers like online retailers. Monitoring of current services, including utilization and demand, should also be part of the analysis and forecasting model.

In some cases, cloud service providers and their customers may be willing to accept a certain amount of oversubscription, especially as it could offer cost savings. To extend the

previous example, assume the service provider offers 50 licenses and the business has 100 users split between the United States and India. Given the time zone difference between the two countries, it is unlikely that all 100 users will try to access the system simultaneously, so oversubscription does not present an issue.

If the consumer does require concurrent accessibility for all 100 users, then they must specify that as an SLA requirement. The provider should then utilize their capacity management processes to ensure adequate capacity is provisioned to meet the anticipated demand.

SUPPORT DIGITAL FORENSICS

Digital forensics, broadly, is the application of scientific techniques to the collection, examination, and interpretation of digital data. The primary concern in forensics is the integrity of data, as demonstrated by the chain of custody. Digital forensics is a field that requires very particular skills and is often outsourced to highly trained professionals, but a CCSP must be aware of digital forensic needs when architecting systems to support forensics and how to acquire appropriate skills as needed to respond to a security incident.

Digital forensics in cloud environments is complicated by a number of factors; some of the very advantages of cloud services are also major disadvantages when it comes to forensics. For example, high availability and data replication mean that data is stored in multiple locations around the world simultaneously, which complicates the identification of a single crime scene. Multitenant models of most cloud services also present a challenge, as there are simply more people in the environment who must be ruled out as suspects. The shared responsibility model also impacts digital forensics in the cloud. As mentioned previously, most CSPs do not allow consumers physical access to hardware or facilities, and even with court orders like a warrant, the CSPs may have procedures in place that make investigation, collection, and preservation of information more difficult. This is not to frustrate law enforcement, but is a predicament caused by the multitenant model; allowing investigation of one consumer's data might inadvertently expose data belonging to other consumers. Investigation of one security incident should, as a rule, not be the cause of other security breaches!

Forensic Data Collection Methodologies

In legal terminology, *discovery* means the examination of information pertinent to a legal action. E-discovery is a digital equivalent comprising steps including identification, collection, preservation, analysis, and review of electronic information. There are two important standards a CCSP should be familiar with related to e-discovery.

- **ISO 27050**: ISO 27050 is a four-part standard within the broader ISO 27000 family of information security standards.
  - Part 1, *Overview and concepts*, defines terms and requirements for organizations to consider when planning for and implementing digital forensics to support e-discovery.
  - Part 2, *Guidance for governance and management of electronic discovery*, offers a framework for directing and maintaining e-discovery programs, with correlation to other elements of the 27000 framework for managing information security.
  - Part 3, *Code of practice for electronic discovery*, provides detailed requirements for achieving e-discovery objectives in alignment with the standard, including evidence management and analysis.
  - Part 4, *Technical readiness*, is under development as of 2020 and is designed to provide more discrete guidance on enabling various systems and architectures to support digital forensics.
- **Cloud Security Alliance (CSA) Security Guidance Domain 3: Legal Issues: Contracts and Electronic Discovery**: This standard is part of the CSA's freely available guidance related to cloud security and covers legal issues, contract requirements, and special issues raised by e-discovery.
  - Legal Issues details concerns related to privacy and data protection and how moving data to the cloud can complicate an organization's legal obligations. Examples include data residency, where laws may restrict the geographic locations that data may be stored in, as well as liability issues when a CSP is acting as a data processor.
  - Contract Considerations lists concerns and recommendations for dealing with common contract issues related to security and privacy. These include performing adequate due diligence on any CSP vendors and their practices, ensuring contractual obligations are properly documented, and performing ongoing monitoring of the CSP and services to ensure they do not exceed the organization's changing risk tolerance.
  - Special Issues Raised by E-Discovery details a number of critical concerns both CSPs and consumers must consider when choosing and architecting cloud solutions. These include possession, custody, and control of data; in short, outsourcing processing of data to a CSP does not absolve the cloud consumer of legal responsibility for the security of that data. Other issues

include challenges related to discovery itself, as data may exist in multiple locations or multiple pieces (data dispersion), as well as issues related to privacy in a multitenant environment if one tenant is subject to legal action. In addition, there may be tools like bit-by-bit analysis or data replication that may be impossible in cloud environments, as these tools make use of hardware features that are not present in a virtualized cloud environment.

- The full document may be accessed here: `cloudsecurityalliance.org/artifacts/csa-security-guidance-domain-3-legal-issues-contracts-and-electronic-discovery`.

When legal action is undertaken, it is often necessary to suspend some normal operations such as routine destruction of data or records according to a defined schedule. In this case, a process known as *legal hold* will be utilized, whereby data is preserved until the legal action is completed. Provisions for legal hold, such as extra storage availability and proper handling procedures, must be part of contracts and SLAs, and during legal proceedings, the cloud consumer should have easy access to appropriate points of contact at the CSP to facilitate e-discovery or other law enforcement requirements.

The process of collecting evidence is generally a specialized activity performed by experts, but security practitioners should be aware of some steps, especially those performed at the beginning before a forensics expert is brought in.

- Logs are essential. All activities should be logged including time, person performing the activity, tool(s) used, system or data inspected, and results.
- Document everything, including physical or logical system states, apps running, and any physical configurations of hardware as appropriate.
- Some data is volatile and requires special handling. In a traditional computer system, RAM is often a particular concern for forensics, because it requires constant power to retain data. In cloud architectures where VMs may be spun up on demand or microservices run only as long as needed to perform a particular task, identifying any ephemeral data or services and preserving them may be critical.
- Whenever possible, work on copies of data or images of systems, as simple actions like opening a folder or file can overwrite or change critical elements of the evidence, leading to a loss of integrity and possible destruction of evidence.
- Verify integrity often and follow standard procedures. Incident response plans will often be the first set of steps leading into an investigation where evidence is required, and they should incorporate checks for integrity of evidence and handling such as hashing and verifying physical custody records.

Evidence Management

When handling evidence, the *chain of custody* documents the integrity of data, including details of time, manner, and person responsible for various actions such as collecting, making copies, performing analysis, and presenting the evidence. Chain of custody does not mean that the data has not been altered in any way, as it is often necessary to make changes such as physically collecting and moving a piece of hardware from a crime scene to a lab. Instead, chain of custody provides a documented, reliable history of how the data has been handled, so if it is submitted as evidence, it may be relied upon. Adequate policies and procedures should exist, and it may be appropriate to utilize the skills of trained forensic experts for evidence handling.

The scope of evidence collection describes what is relevant when collecting data. In a multitenant cloud environment, this may be particularly relevant, as collecting data from a storage cluster could inadvertently expose data that does not belong to the requesting party. Imagine two competing companies both utilize a CSP, and Company A makes a request for data relevant to legal proceedings. If the CSP is not careful about the evidence collected and provided to Company A, they may expose sensitive data about Company B to one of their competitors!

Evidence presented to different audiences will follow different rules. For example, an industry regulator may have a lower integrity threshold for evidence, as they are not able to assess criminal penalties for wrongdoing. A court of law, however, will have higher constraints as the stakes are higher—they have the power to levy fines or possibly imprison individuals. Evidence should possess these five attributes in order to be useful:

- **Authentic**: The information should be genuine and clearly correlated to the incident or crime.
- **Accurate**: The truthfulness and integrity of the evidence should not be questionable.
- **Complete**: All evidence should be presented in its entirety, even if might negatively impact the case being made (it is illegal in most jurisdictions to hide evidence that disproves a case).
- **Convincing**: The evidence should be understandable and clearly support an assertion being made; for example, this particular person accessed this particular system and copied this particular file for which they were not authorized.
- **Admissible**: Evidence must meet the rules of the body judging it, such as a court, which may rule out evidence such as hearsay (indirect knowledge of an action) or evidence that has been tampered with.

Collect, Acquire, and Preserve Digital Evidence

There are four general phases of digital evidence handling: collection, examination, analysis, and reporting. There are a number of concerns in the first phase, collection, which are essential for a CCSP. Evidence may be acquired as part of standard incident response processes before the need for forensic investigation and criminal prosecution have been identified, so the incident response team needs to handle evidence appropriately in case a chain of custody needs to be demonstrated. Proper evidence handling and decision-making should be a part of the incident response procedures and training for team members performing response activities.

There are a number of challenges associated with evidence collection in a cloud environment, including the following:

- **Control**: Using a cloud service involves loss of some control, and different service models offer varying levels of access. SaaS models typically offer little to no visibility outside a consumer's data or app instance, so investigating network security failures may be impossible as the CSP does not disclose them. On the other hand, IaaS gives an organization complete control over their virtualized network security, but may stop short of any physical network data that might be pertinent to an investigation. Evidence that is inaccessible is obviously a challenge to an investigation; proper contract terms should be in place with any CSP to support an organization's likely investigation needs.
- **Multitenancy and shared resources**: Evidence collected while investigating a security incident may inadvertently become another data breach in a multitenant environment, if the evidence collected includes information that does not belong to the investigating organization. This can also cause problems of attribution—was a particular incident caused by a malicious attacker or by the accidental action of another tenant?
- **Data volatility and dispersion**: Cloud environments support high availability of data, which often requires novel data storage techniques like *sharding*, which is breaking data into smaller pieces and storing multiple copies of each piece across different data centers. Reconstructing that data could be an error-prone process, lowering the believability when evidence is presented, and may involve multiple legal frameworks and jurisdiction issues when data is stored across multiple countries. Other positive features of cloud computing such as live VM migration can complicate evidence gathering as well. When a VM can move to anywhere in the world virtually instantaneously, it may be hard to prove that a specific set of actions carried out by a specific person occurred in a specific place, leading to evidence that is worthless for prosecution.

Preparing for Evidence Collection

There are a number of important steps the organization should take prior to an incident that can support investigations and forensics. These can be built into several processes the organization is likely to perform for other security objectives, including the following:

- **Logging and monitoring**: All apps, systems, and infrastructure should generate audit trails, which should be forwarded to a secure, centralized logging tool such as a syslog server or a SIEM platform. These tools should have unique access controls so that insiders cannot easily cover their tracks by deleting or altering logs, which also increases the work factor for an attacker by requiring multiple sets of compromised credentials to pull off an attack and cover the evidence. Regular reviews of logs for suspicious activity or automated correlation and alerting should be performed to identify potentially suspicious activity.
- **Backup and storage**: Evidence often needs to be compared against a baseline to show how malicious activity caused changes. Adequate backups are useful to show how a system was configured before an incident to prove that a particular activity was the cause of an incident.
- **Baselines and file integrity monitoring**: The known good state of a system is useful for comparison to an existing system when an investigator is trying to determine if malicious activity has occurred. Deviations from the baseline or files that have been changed can be indicators of an attack, or they could be the sign of poorly implemented change management and configuration management practices. If all intentional changes made by the organization are properly documented, it makes spotting unintended or malicious changes much easier.
- **Data and records retention**: Most organizations have some requirement to retain records for a specific period of time dictated by business needs or legal/regulatory requirements. These records may be a useful source of information in an investigation. At the end of that retention period, data is generally destroyed, but the retention policy should also have clear requirements and procedures for placing a legal hold on records that may be pertinent, thereby preventing their destruction until the investigation is complete.

Evidence Collection Best Practices

Although forensics experts may perform significant amounts of evidence collection, security practitioners must be aware of some best practices, including the following:

- When collecting evidence, it is best to utilize original physical media whenever possible, as copies may have unintended loss of integrity. Note that this applies only to collection; the best practice for analysis is to always use verified copies to preserve the original evidence. As mentioned, collecting physical evidence

in the cloud may not be possible, though cyber-forensic tools such as Netscout are emerging that ease collection of digital forensic evidence, and some CSPs offer digital forensics services using their own tools and staff. In exceptional circumstances where physical evidence must be collected, it is likely the CSP will require that law enforcement be involved, particularly an agency with international jurisdiction. Even then, there is no guarantee; in a well-publicized case, *Microsoft Corp. v. United States*, Microsoft challenged a U.S. Department of Justice warrant for data stored in an Irish data center, which Microsoft claimed was outside U.S. jurisdiction. The case was ultimately rendered moot by passage of the Clarifying Lawful Overseas Use of Data Act (US Cloud Act), but new laws being written for specific circumstances like this are exceedingly rare.

- Verify integrity at multiple steps by using hashing, especially when performing operations such as copying files. Calculate original and copy hashes and compare them to ensure they match.
- Follow all documented procedures, such as the use of a dedicated evidence custodian for collection, logging of activities performed, leaving systems powered on to preserve volatile data, etc. These procedures should be documented in the organization's incident response plan.
- Establish and maintain communications with relevant parties such as the CSP, internal legal counsel, and law enforcement for guidance and requirements.

Evidence Preservation Best Practices

Once evidence has been collected, it must be adequately preserved to support a variety of goals: maintain the chain of custody, ensure admissibility, and be available for analysis to support the investigation. Preservation activities and concerns should cover the following:

- **Adequate physical security**: Most of us are familiar with TV police shows where an evidence locker is used. In the best-case scenario, the detectives must check out evidence following a documented procedure; in the worst-case scenario, the evidence has been removed with no documentation, leading to a criminal getting away with their actions. Digital evidence must be stored on physical media, which requires adequate physical protections, similar to a data center, and should also be subject to integrity tracking similar to a library book: documenting who checked out the evidence at what time and additionally documenting the actions taken such as copying data.
- **Physical and environmental maintenance**: Since evidence is stored on physical media, it may require environmental maintenance such as temperature and humidity controls, as well as available power to preserve data. Battery replacement or charging will be especially important for mobile devices.

- **Blocking interference**: Computing systems have access to a wide variety of wireless communications media including Bluetooth and WiFi and may also try to communicate with the outside world via wired connections. Forensic analysts and evidence handlers need to adequately shield evidence-containing devices from this interference. To block wireless communications during analysis, the use of a Faraday cage, which blocks electromagnetic signals, is a best practice. For transporting mobile devices, a Faraday bag is recommended; in both cases, this can prevent these devices from being remotely wiped or reset. Workstations and other devices should be analyzed using a physically air-gapped network to prevent similar activities.
- **Working from copies**: Unlike evidence collection where use of originals is preferred, examination and analysis activities should be performed on copies of data and devices wherever possible, with frequent integrity comparisons to ensure the copy being analyzed matches the original. Tools such as virtualization are useful here, as they can create an exact copy (image) of a system for analysis. Where copies are not available, tools such as write blockers should be used; these devices allow for read-only access to devices and prevent writing to them. Even the simple act of connecting a drive to a modern OS causes files such as a search index to be written, which could destroy or damage data that the investigator needs.
- **Document everything**: Remember the chain of custody does not mean data or a system has not been changed at all, but defensibly documents the who, what, how, why, and when of changes. Checking evidence out for analysis, calculating hashes for comparison, and making copies for analysis are examples of actions that should be documented.

MANAGE COMMUNICATION WITH RELEVANT PARTIES

Adequate coordination with a variety of stakeholders is critical in any IT operation, and the move to utilize cloud computing resources coupled with an increasingly regulated and dispersed supply chain elevates the priority of managing these relationships. Communication is a cornerstone of this management; providing adequate and timely information is critical. While this may be a skillset fundamental to project managers rather than security practitioners, it is worth understanding the importance of effective communication and supporting it whenever possible.

Effective communication should possess a number of qualities. The contents, nature, and delivery of communications will drive many decisions, which can be elicited using a series of questions about the information to be conveyed.

- **Who?** The intended audience will determine the contents of a communication, such as the level of technical detail included or amount of information. In security incidents, communications to the general public may be reviewed by the organization's legal department to avoid any admission of culpability or exposing sensitive information.
- **What?** The goal of the communication must be met by the contents. For example, reports to investors and business partners would contain different sets of details, as partners are likely to be under an NDA but the general public is not. If the message does not clearly answer the question, "So what?", it is likely the audience will not find it terribly useful.
- **Why?** The purpose of the communication should be clear, and the intended audience should be able to make immediate use of it. We are all familiar with technical error messages that provide less-than-helpful details of an internal memory error at a specific register address, but that does not help the average user to carry on with their work.
- **When?** Is the communication timely? If a data breach happened two years ago and the organization knew but did not report it, the information is likely to be useless as users will already have been targeted for identity theft scams. Furthermore, the communication is likely to cause reputational harm, as the organization could be accused of covering up the incident rather than reporting it in a timely fashion.

There are a number of stakeholders or constituents with whom an organization is likely to communicate regarding IT services, organizational news and happenings, and emergency information. Establishing clear methods, channels, and formats for this communication is critical.

Vendors

Few organizations exist in a vacuum; the modern supply chain spans the globe, and regulatory oversight has begun to enforce more stringent oversight of this supply chain. It is therefore essential that an organization and its security practitioners understand the supply chain and establish adequate communications.

The first step in establishing communications with vendors is an inventory of critical third parties on which the organization depends. This inventory will drive third-party or vendor risk management activities in two key ways. First, some vendors may be critical

to the organization's ongoing functioning, such as a CSP whose architecture has been adopted by the organization. Second, some vendors of goods and services may provide critical inputs to an organization like a payment card processor whose service supports the organization's ability to collect money for its goods or services.

Communication with critical vendors should be similar to internal communications due to the critical role these vendors play in the business. If a vendor incident is likely to impact an organization's operations, the organization ought to have well-established communications protocols to receive as much advance notice as possible. If a consumer notices an incident such as loss of availability of a vendor's service, there should be adequate reporting mechanisms to raise the issue and resolve it as quickly as possible.

Many vendor communications will be governed by contract and SLA terms. When a CSP is a critical vendor, there should be adequate means for bidirectional communication of any issues related to the service, such as customer notices of any planned outages or downtime, emergency notifications of unplanned downtime, and customer reporting for service downtime or enhancements. In many cases, this will be done through a customer support channel with dedicated personnel as well as through ticketing systems, which creates a trackable notice of the issue, allowing all parties to monitor its progress.

Customers

As cloud consumers, most organizations will be the recipients of communications from their chosen CSPs. While this might seem to imply there are no responsibilities other than passively receiving information from and reporting issues to the CSP, consumers do have a critical accountability: defining SLA terms. Levels of communication service from the CSP should all be defined and agreed-upon by both parties, such as speed of acknowledging and triaging incidents, required schedule for notification of planned downtime or maintenance, days/times support resources are available, and even the timeframe and benchmarks for reporting on the service performance. SLAs may be generic and standardized for all customers of a CSP or may be highly specific and negotiated per customer, which offers more flexibility but usually at greater cost.

Shared Responsibility Model

A key source of information to be communicated between CSPs and their customers is the responsibility for various security elements of the service. The CSP is solely responsible for operational concerns like environmental controls within the data center, as well as security concerns like physical access controls. Customers using the cloud service are responsible for implementing data security controls, like encryption, that are appropriate to the type of data they are storing and processing in the cloud. Some areas require action by both the provider and customer, so it is crucial for a CCSP to understand which cloud service models are in use by the organization and which areas of security must be addressed by each party. This is commonly referred to as the *shared*

responsibility model, which defines who is responsible for different aspects of security across the different cloud service models. The generic model in Table 5.1 identifies key areas of responsibility and ownership in various cloud service models.

TABLE 5.1 Cloud Shared Responsibility Model

RESPONSIBILITY	IAAS	PAAS	SAAS
Data classification	C	C	C
Identity and access management	C	C/P	C/P
Application security	C	C/P	C/P
Network security	C/P	P	P
Host infrastructure	C/P	P	P
Physical security	P	P	P

C=Customer, P=Provider

A variety of CSP-specific documentation exists to define shared responsibility in that CSP's offerings, and a CCSP should obviously be familiar with the particulars of the CSP their organization is utilizing. The following is a brief description of the shared responsibility model for several major CSPs and links to further resources:

- **Amazon Web Services (AWS)**: Amazon identifies key differences for responsibility "in" the cloud versus security "of" the cloud. Customers are responsible for data and configuration in their cloud apps and architecture, while Amazon is responsible for shared elements of the cloud infrastructure including hardware, virtualization software, environmental controls, and physical security.

  More information can be found here: `aws.amazon.com/compliance/shared-responsibility-model`.

- **Microsoft Azure**: Microsoft makes key distinctions by the service model and specific areas such as information and data and OS configuration. Customers always retain responsibility for managing their users, devices, and data security, while Microsoft is exclusively responsible for physical security. Some areas vary by service model, such as OS configuration, which is a customer responsibility in IaaS but a Microsoft responsibility in SaaS.

  More information can be found here: `docs.microsoft.com/en-us/azure/security/fundamentals/shared-responsibility`.

- **Google Cloud Platform (GCP)**: Google takes a different approach with a variety of shared responsibility documentation specific to different compliance

frameworks such as ISO 27001, SOC 2, and PCI DSS. The same general rules apply, however: customer data security is always the customer's responsibility, physical security is always Google's responsibility, and some items are shared depending on what service offerings are utilized.

More information can be found here: `cloud.google.com/security`.

Partners

Partners will often have a level of access to an organization's systems similar to the organization's own employees but are not directly under the organization's control. Communication with partners will be similar to communication with employees, with initial steps required when a new relationship is established, ongoing maintenance activities throughout the partnership, and termination activities when the partnership is wound down. Each of these phases should deliver clear expectations regarding security requirements.

- **Onboarding**: During onboarding, a partner is introduced to the organization, and two important processes must be executed: third-party or vendor due diligence activities to assess the partner prior to sharing data or granting access, and communication of security requirements through contracts, SLAs, or other documentation. Elements of the organization's security program may be shared with the partner such as security policies and training.
- **Management**: Once a partner has been onboarded, the organization's ongoing security efforts must include any partner activities. This should include managing access and permissions, auditing and oversight activities, and processes such as incident response and disaster recovery testing.
- **Offboarding**: When a partnership ends, the organization must follow procedures to terminate access and communicate any security requirements relevant to the termination, such as reminders of nondisclosure requirements and the return of any assets the partner might have.

Regulators

There are a vast array of regulatory bodies governing information security, and most of them have developed cloud-specific guidance for compliant use of cloud services. In the early days of cloud computing, security practitioners often faced significant unknowns when moving to the cloud, but as the cloud became ubiquitous, regulators delivered clear guidance, and CSPs moved quickly to deliver cloud solutions tailored to those compliance requirements. A CCSP is still responsible for ensuring their cloud environment is in compliance with all regulatory obligations applicable to their organization.

The main component of regulatory communication regarding the cloud is monitoring incoming information regarding regulatory requirements in the cloud. For example, the recent implementation of GDPR in the European Union (EU) has caused many organizations to make architectural decisions regarding their cloud applications. GDPR's restrictions on data leaving the geographic boundaries of the EU mean many organizations need to implement additional privacy controls to allow for that data transfer or host applications in EU-based data center. The CCSP should subscribe to feeds and be aware of regulatory changes that impact their organization's use of the cloud.

Similar to monitoring incoming cloud requirements, a CCSP may also be required to report information to regulatory bodies regarding their organization's state of compliance. For example, U.S.-based companies with GDPR privacy requirements may be required to communicate to the U.S. Department of Commerce on the status of privacy compliance under the U.S. Privacy Shield framework. CCSPs should be aware of regulatory reporting requirements specific to their organization, and ensure required documentation, artifacts, audits, etc., are communicated in a timely fashion.

Other Stakeholders

Communications may not be a primary job responsibility of a CCSP, but important details of security risk management work may need to be shared. Working with appropriate personnel in the organization will be crucial to ensure information is communicated in a timely manner and with relevant details for each audience, such as the following:

- **Public**: Details of security incidents are the most obvious category, especially if the breach could affect security or privacy. In this case, adequately trained public relations (PR) professionals as well as legal personnel should be consulted and may handle communications with input from the security team.
- **Security researchers**: Increasingly, security experts are performing research on specific organization's apps and services, with the goal of responsibly disclosing their findings before an attacker can exploit them. Organizations should have a method for receiving and responding to this type of communication, often via a publicly documented security email address. Getting details from a researcher, following up if more information is required, and in some cases paying a bounty for the discovered bug may all be ongoing communications that need to be handled.
- **Investors**: In many countries, formal communication with investors is required on a regular basis, such as quarterly business filings, annual financial statements, or unusual business activities such as a merger. Security posture, incidents, and

compliance may all be relevant to some investors, and due to legal requirements, certain security details may be required.

- **Crisis communication**: An incident that causes significant disruption or disturbance such as a natural disaster may require communications to multiple stakeholder groups at once. Internal employees, business partners, investors, and the general public may all be impacted by a business disruption; in this case, the organization's communications personnel should have documented procedures, and security practitioners should look to these personnel for guidance on what information may be required.

MANAGE SECURITY OPERATIONS

Security operations represent all activities undertaken by an organization in monitoring, maintaining, and generally running their security program. This may include continuous oversight of security systems to identify anomalies or incidents, as well as an organizational unit to house concerns related to security processes like incident response and business continuity. A large percentage of security process and procedure documentation will be attached to this function, and these processes should also be the target of continuous improvement efforts.

As in many security topics, an ISO standard exists that may be useful for security practitioners conducting security operations: ISO 18788, *Management system for private security operations — Requirements with guidance for use*. A word of caution: this document contains a great deal of material not applicable to cybersecurity or infosec concerns, as the numbering might imply (it is not part of the 27000 series). One extraneous topic is implementing and managing a private security force with authorization to use force; however, it also provides a business management framework for understanding the organization's needs, guidance on designing strategic and tactical plans for operating security programs, and applying a risk management approach. It also suggests a standard plan, do, check, act framework for implementing and improving security operations.

Security Operations Center

The security operations center (SOC) is an organizational unit designed to centralize a variety of security tasks and personnel at the tactical (mid-term) and operational (day-to-day) levels of the organization. While security strategy may rely on input from top leaders such as a board of directors, department secretary or minister, or the C-suite executives, SOC personnel are responsible for implementing steps required to achieve that strategy and maintain daily operations. Building and running a SOC in a traditional,

all on-prem environment is basically building a monitoring and response function for IT infrastructure. Extending these concepts to the cloud may require some trade-offs, as the CSP will not offer the same level of access for monitoring that an on-prem environment does.

It is also important to note that there will be at least two SOCs involved in cloud environments. The CSP should run and manage their own SOC focused on the elements they control under the shared responsibility model such as infrastructure and physical security, while consumers should run their own SOC for their responsibility areas, chiefly data security when using the cloud.

The CISO Mind Map published by security author Rafeeq Rehman, found at `rafeeqrehman.com/?s=mindmap`, provides a more information-security-centric view of security operations than ISO 18788. Updated each year, the Mind Map details the items that a CISO's role should cover; the largest element of the job responsibilities is security operations, which is broken down into three main categories. This provides a strong framework for responsibilities the SOC should undertake, including the following:

- **Threat prevention**: This subcategory includes preventative controls and risk mitigations designed to reduce the likelihood of incidents occurring. These include adequate network security, vulnerability and patch management programs, application security (appsec), information system hardening, and maintenance activities for common security tools such as PKI.
  - Threat prevention in the cloud will involve similar activities such as keeping an adequate asset inventory and utilizing it to detect vulnerabilities and fix them via patching, particularly when using IaaS. In some PaaS and virtually all SaaS, vulnerability and patch management will be the responsibility of the CSP. Implementing network security in the cloud requires adequate selection and deployment of cloud-appropriate tools like firewalls and intrusion detection tools; many traditional tools do not work in virtualized networks or rely on capabilities like mirroring network traffic from a SPAN port on a switch to inspect traffic, which is not available in a virtual cloud network. SOC operations need to be able to access that data and integrate it with monitoring tools, often through the use of API.
- **Threat detection**: Detecting threats requires tools, processes, and procedures, such as log capture/analysis/correlation, which is often achieved with a SIEM tool. Other capabilities include real-time monitoring tools such as data loss/leak prevention (DLP) tools and network security solutions like IDS/IPS, antimalware, and firewalls. Some security testing processes also fall in this category such as red (offensive) and blue (defensive) team exercises.

- The general process of running the SOC falls into this category, covering topics such as resource management and training. Adequately trained personnel and robust procedures are essential to the success of a security program. Due to the tools in use, there will be architectural management concerns as well, chief among them is ensuring the tools in use adequately monitor the infrastructure. As many organizations migrate to the cloud, legacy tools designed to monitor and defend a single network become obsolete since they do not scale to a globally accessible cloud environment.
- Building out a SOC for cloud operations may require the use of different monitoring tools, as many legacy tools are designed to be deployed on a physical network to monitor traffic and events. That capability is not available in a virtualized, cloud network, and given the use of a CSP's resources, many organizations implement additional encryption to protect data. This has the downside of often rendering data invisible to monitoring tools. CSPs offer a number of built-in security monitoring and alerting tools, such as Azure Sentinel and AWS Amazon GuardDuty, which can be used as standalone programs or integrated with SOC monitoring tools using an API.
- No security program is foolproof, so detecting threats and incidents is critical. Ideally, threats should be proactively identified by the organization via threat hunting *before* an attacker exploits them. If an attacker finds and exploits an unknown vulnerability, this function also provides incident detection capabilities that can reactively identify an incident that would be handled by the third function.

- **Incident management**: Once an incident has occurred, the SOC will often serve as the main point of coordination for the incident response team (IRT). This subcategory will house many of the functions previously discussed, such as developing incident response capabilities, handling communications like regulator and customer notifications, and handling forensics.
  - Two particular threats stand out for extra attention in this category. The first is data breach preparation, which encompasses the majority of incident response planning activities, business continuity planning, logging and monitoring functions, and the recently included cyber risk insurance. Insurance represents a form of risk transfer, which helps to shift the impact of a risk from the organization to another party, in this case the organization's insurer. Many cyber risk insurance plans also offer access to useful incident handling specialties such as data breach and privacy lawyers, forensic experts, and recovery services.

- The other threat explicitly called out is one that has grown in recent years: ransomware. Attackers are exploiting encryption to lock users out of information systems until a ransom is paid, so preparing for or preventing one of these attacks is critical. Preventative steps can include file integrity monitoring, designed to detect unwanted changes such as ransomware encrypting files or even the malware being installed. Planning for recovery is also crucial beforehand, with adequate business continuity, disaster recovery, and backup plans being the primary implementations for ransomware recovery.
- As discussed in the incident management section of this chapter, coordination between the CSP and consumer will be essential for incident response and handling and is a key difference from on-prem environments.

A SOC is typically made up of security analysts, whose job involves taking incoming data and extracting useful information, and security engineers who can keep operations running smoothly. There may be overlap with operations personnel, and in some organizations, the SOC may be combined with other operational functions. Common functions that may be performed in the SOC are the following:

- **Continuous monitoring and reporting**: All the security tools implemented by an organization require at least some level of ongoing oversight by a human being. The SOC will typically centralize relevant data into a dashboard, which may be projected on large screens in the SOC physical office or available as a web page so all members can see vital information. In addition to SOC members, certain information may be reported to other members of the organization in the form of metrics or incident reports.
- **Data security**: The SOC should have the ability to perform monitoring across any environments where sensitive data is being stored, processed, or transmitted. This can be used to ensure data is being adequately protected in various states across all stages of its lifecycle. Threat hunting and vulnerability management functions in the SOC should look for risks to data, such as improperly configured cloud storage environments or insecure data transmission when users are connecting with SaaS applications.
- **Alert prioritization**: Not all alerts are critical, require immediate attention, or represent imminent harm to the organization. SOC functions related to log management should include definitions to assist in prioritizing alerts received from various sources such as monitoring tools, as well as defined procedures for taking action on alerts.

Loss of commercial power at a facility is an alert worth monitoring, but may not be a major incident if backup power is available and the power is likely to be restored quickly. If the organization is experiencing exceptional operations, such as retail during a major holiday, then the organization may choose to preemptively declare a business interruption and shift processing to an alternate facility. Detecting and triaging incidents, up to and including declaration of an interruption or disaster, is a logical function for the SOC to perform due to the type of data they work with.

- **Incident response**: SOC personnel are uniquely situated to detect and respond to anomalous activity such as an interruption or incident and as such are often the core constituent of an incident response team. Skills and expertise such as digital forensics, whether internal to the team or via third-party services, may be a logical fit on this team as well. At the conclusion of an incident response, this team is also well suited to perform root-cause analysis and recommend remedial actions.
- **Compliance management**: The SOC is likely to have information that is quite crucial to managing the organization's compliance posture. Although compliance functions may be implemented elsewhere in the organization to avoid conflict of interest, critical information monitored by the SOC can be useful, such as server configurations, results of routine vulnerability scans, and crucial risk management activities like business continuity.

As always, there are two key perspectives for the SOC. CSPs will likely need a robust SOC function with 24×7×365 monitoring of the environment. While such a capability will be expensive, the cost is likely justified by requirements of the cloud consumers and can be shared among this large group. Cloud consumers may operate a SOC for their own operations, which will include any on-prem IT services as well as their cloud services and environments. This may require the use of some legacy tools deployed to more traditional cloud services such as IaaS or PaaS, newer tools designed to monitor services such as SaaS (for example, access management or encryption tools), or even the use of CSP-provided monitoring capabilities.

As an example, the major CSPs offer security incident reporting services that customers can log in to if an incident affects them. They also offer the following public status pages that list operational information for public-facing services:

- **AWS Service Health Dashboard**: `status.aws.amazon.com`
- **Microsoft Azure status**: `status.azure.com/en-us/status`
- **Google Cloud Status Dashboard**: `status.cloud.google.com`

One crucial decision to be made when designing a SOC is the use of internal resources or outsourcing the function (build versus buy). As previously mentioned, the

CSPs can likely justify the cost of robust SOC resources due to cost sharing among customers and the requirements those customers will impose. A small organization with only cloud-based architecture may decide to outsource their security operations and monitoring, as many services provide dedicated support for specific cloud platforms at a lower cost than building the same function internally. These are known as managed security services providers (MSSPs). Like most business decisions, this will be a trade-off between control and cost and should be made by business leaders with input from security practitioners. A CCSP should understand the cloud architecture and communicate any risks the organization might assume by utilizing a third-party SOC.

Monitoring of Security Controls

Monitoring of security controls used to be an activity closely related to formal audits that occur relatively infrequently, sometimes once a year or even once every three years. A newer concept is known as *continuous monitoring*, which is described in the NIST SP 800-37 Risk Management Framework (RMF) as "Maintaining ongoing awareness to support organizational risk decisions." Information that comes from an audit conducted more than a year ago is not ongoing awareness. Instead, the RMF specifies the creation of a continuous monitoring strategy for getting near real-time risk information.

Real-time or near real-time information regarding security controls comprises two key elements: the status of the controls and any alerts or actionable information they have created. Network resources are at risk of attacks, so network security controls like IDS are deployed. Continuous monitoring of the IDS's uptime is critical to ensure that risk is being adequately mitigated. A facility to view any alerts generated by the device, as well as personnel and processes to respond to them, is also crucial to the organization's goal of mitigating security risks; if the IDS identifies malicious network activity but no action is taken to stop it, the control is not effective.

A longer-term concern for monitoring security controls and risk management is the suitability of the current set of tools. As organizations evolve, their infrastructure will likely change, which can render existing tools ineffective. The SOC should be charged with ensuring it can monitor the organization's current technology stack, and a representative should be part of change management or change control processes. Migrating a business system from on-prem hosting to a SaaS model will likely have a security impact with regard to the tools needed to monitor it, and the change board should ensure this risk is planned for as part of the change.

In general, the SOC should have some monitoring capabilities across all physical and logical infrastructure, though detailed monitoring of some systems may be performed by another group. For example, a physical access control system dashboard may be best monitored by security guards who can perform appropriate investigation if an alarm is

triggered. Some organizations run a network operations center (NOC) to monitor network health, and NOC engineers would be best suited to manage telecommunications equipment and ISP vendors. However, an operational incident in either of these two systems, such as a break-in or loss of ISP connectivity, could be an input to the SOC's incident management function. Several controls that might be particularly important for SOC monitoring include the following:

- **Network security controls**: This includes traditional devices (or their virtual equivalents) such as network firewalls, web app firewalls (WAF), and IDS/IPS. The SOC should be able to see if devices are functioning, alerts such as suspicious activity or problems, and trends such as volume of dropped packets on a firewall or activity on a honeypot system. These could be indicators of a potential attack and are best dealt with proactively. Services may be included in the SOC monitoring as well, especially if used to achieve critical risk mitigation objectives like identity management (ID as a service, or IDaaS) or cloud application orchestration via platforms like Kubernetes.
- **Performance and capacity**: All IT services have performance requirements and capacity limitations. While the cloud can theoretically offer benefits with both, the process may not be entirely automated. Some cloud service offerings are simply cloud-based, virtualized versions of legacy services. In these cases, there can be capacity constraints if the organization's usage grows; for example, a PaaS database may run out of storage as usage grows. Core services like storage, network bandwidth, and compute capacity should be monitored. Many of these will also be SLA metrics that the organization should monitor, so this function can achieve two objectives.
- **Vulnerability assessments**: Vulnerability scanners can be configured to run on a defined frequency, either time- or trigger-based. Unlike audit frameworks in the past, which included a vulnerability scan once a year during audit, organizations implementing continuous monitoring should seek to define more frequent scan schedules. Time-based schedules should be made to balance any operational overhead, such as system performance slowdown, with user needs—often achieved by running scans outside of normal business hours. Trigger-based scans can be set up to conduct scans when certain activities occur, such as a new version of an application being deployed to the production environment. These can often be useful in achieving real-time risk visibility, as the scan is conducted at the same time a new attack surface is created. Vulnerability assessments may be conducted from an internal or external perspective to simulate what vulnerabilities a trusted insider or malicious outsider might be able to find and exploit.

Log Capture and Analysis

NIST SP 800-92, *Guide to Computer Security Log Management*, defines a log as "a record of the events occurring within an organization's systems and networks" and further identifies that "Many logs within an organization contain records related to computer security . . . including security software, such as antivirus software, firewalls, and intrusion detection and prevention systems; operating systems on servers, workstations, and networking equipment; and applications." These logs of internal activity can unfortunately be overwhelming for humans to attempt to meaningfully review due to the sheer number of events generated by modern information systems. Security information and event management tools offer assistance.

SIEM

SIEM tools provide a number of functions useful to security; namely, the following:

- **Centralization**: It would be impossible for a human to navigate all the log files of individual hosts in even a moderately complex information system. There may be database, middleware, and web servers each with a corresponding log file, not to mention application logs for the software running on each of those systems. A SIEM provides the ability to centralize all log data. Logs are forwarded or sent to the SIEM tool from the system they originate on; this may be done as a native function of the system or rely on an agent installed that copies log files and sends them to the SIEM. This can also be used to enforce a key access control for log data. Users with credentials to monitored systems may be able to delete or change logs on those systems but can be easily denied permission to the SIEM tool.
- **Normalization**: Log files generated by disparate systems may contain data that is similar but not exactly the same. One system may generate log files with a user's ID, like jsmith, while another utilizes the user's email address, like john.smith@company.com. This makes it harder to analyze the data, so SIEM platforms can transform data to a common format, such as the use of UTC for timestamps to avoid issues with time zones, or the use of consistent field names like Timestamp instead of Event Time.
- **Correlation and detection**: Once data has been centralized and normalized, the SIEM can better support detection of suspicious events. This is often done by comparing activities in the log with a set of rules, such as a user's location. If a user typically accesses a system every day from their desk on the corporate network, then that user suddenly logging on from another country halfway around the world is suspicious. Some SIEM platforms incorporate more complex detection methods like artificial intelligence, which uses learning models to identify potentially suspicious activity or weed out false positives.

Correlation refers to discovering relationships between two or more events; in this example, if the user has suddenly logged in from a new location, accessed email, and started downloading files, it could indicate compromised credentials being used to steal data. It could also indicate that the user is attending a conference and is getting some work done in between sessions, but the organization should still perform checks to verify. If travel is a common situation, the organization might also integrate data from an HR or travel reservation system to correlate travel records for a certain user with their activity accessing data from a particular country. If the user is not known to be traveling in that country, the activity is highly suspicious.

Other sources of information could include external information like threat intelligence or CSP status feeds, which can be relevant in investigating anomalies and categorizing them as incidents. Major CSPs offer monitoring capabilities in their platforms as well, and this data may be critical for investigating anomalies.

- **Alerting**: Once suspicious activity is detected, the SIEM should generate an alert, and a SOC analyst should follow a documented procedure to review and take action on the alert. Many SOCs will utilize a system for handling this process, such as an IT ticketing system that generates work tickets in a queue for analysts to take action on. The activities related to each ticket are captured along with details such as who is working on it, steps taken, and artifacts from the process like investigation logs.

PaaS and SaaS in particular can pose issues for logging and monitoring, as the logs themselves may not be visible to the consumer. CSPs typically don't share internal logs with consumers due to the risk of inadvertently exposing customer details, so these services may be a black hole in the organization's monitoring strategy. Solutions have been designed that can address this, such as the cloud access security broker (CASB), which is designed to log and monitor user access to cloud services. These can be deployed inline to a user's connection to cloud services or collect information via API with cloud services. The CASB can monitor access and interaction with applications; for example, user Alice Doe logged into Dropbox at 12:30, and uploaded file `Super Secret Marketing Data .xlsx` at 12:32. Dropbox as a CSP may not share this data with consumers, but the CASB can help overcome that blind spot.

Log Management

NIST SP 800-92 details critical requirements for securely managing log data, such as defining standard processes and managing the systems used to store and analyze logs.

Because of their critical nature in supporting incident investigations, logs are often a highly critical data asset and worthy of robust security mechanisms. These may include the following:

- High-integrity storage media such as write once, read many (WORM), which allows a definitive copy of the data to be written just once. This prevents tampering with log files after they are written to disk, but allows for analysis.
- Restricted access to only SOC personnel, as logs may contain highly sensitive information. This could include details of internal system configuration, as well as sensitive information being processed if an application or database error contains pieces of the sensitive data.
- Capacity monitoring and rotation to ensure adequate storage space. Logs grow quite quickly, and allocating sufficient space for logs is essential. Maintaining all data readily available may be cost-prohibitive, so some log files may be stored offline if appropriate or rotated (written over) after a set period of time.
- Retention and secure deletion when logs are no longer needed. Log files should be part of the organization's retention schedule and should be in compliance with applicable legal and regulatory obligations and are also critical to reconstruction of historical events as needed for incident investigations. Because of the possibility of highly sensitive data, secure destruction methods should be defined for log data.
- Proper configuration of logging functions on all systems. Many systems and apps provide configuration options for what data is written to log files based on the value of the information. Informational alerts, such as a server synchronizing its clock with a timeserver, may not be of much value. Events like a user successfully logging in, entering the wrong credentials, or changing their password would be of more value, as they provide evidence of the organization's access controls at work. The *clipping level* defines which categories of events are and are not written to logs, such as user authentication events, informational system notices, or system restarts.
- Testing and validation of log data to ensure integrity is not compromised. Log data collected, especially on a centralized tool like a SIEM, should be periodically validated against the originals to ensure it is being transmitted without alternation. It should not be possible to alter data being forwarded to the centralized platform, and any attempts to tamper with or disable the forwarding functionality should result in an alert so corrective action can be taken.

Incident Management

An *incident* is any unplanned event that actually does or has the ability to reduce the quality of an IT service. In security terms, reducing the quality is synonymous with impacting any element of the CIA triad. As an example, an event could be a loss of commercial power at a data center. If the organization has been notified that the power company is performing maintenance and is able to continue running with backup generators, then this is merely an event. The IT services provided by the data center can continue uninterrupted. If instead, the power is cut unexpectedly and the facility must switch to backup power, this could cause systems to be unresponsive or lose data during the transition to backup power. This is an obvious negative impact to data integrity and system availability.

Incident management or incident response (IR) exists to help an organization plan for incidents, identify them when they occur, and restore normal operations as quickly as possible with minimal adverse impact to business operations. This is referred to as a capability, or the combination of procedures and resources needed to respond to incidents, and generally comprises three key elements:

- **Incident response plan (IRP)**: The IRP (sometimes known as an *incident management plan*) is a proactive control designed to reduce the impact of an incident. It defines structure for the processes to be used when responding to an incident, which allow team members to respond in a quick and orderly fashion. People are often in high-stress situations when an outage or interruption occurs, and having a plan with detailed scenarios and response steps can provide better decision-making and response outcomes. The IRP should include detailed, scenario-based response procedures for team members to follow, which are based on incidents that the organization is likely to face. For example, an IT consulting firm is unlikely to face a major attack on an industrial control system (ICS) but is likely to fall victim to phishing attacks. The IRP should include detailed instructions on how to identify a phishing campaign, prevent the message from propagating, and provide cleanup or remediation steps for users who have fallen victim, such as immediately locking user accounts until investigation can be conducted.
- **Incident response team (IRT)**: The IRP should detail the personnel needed to respond to incidents. This team is likely to be dynamic based on the type of incident; for example, a small malware infection is likely to require help desk technicians to assist, while a data breach of PII will require legal counsel to assist with data breach notifications. All members of the IRT should be trained on their responsibilities *before* an incident occurs, and a designated coordinator must be appointed to lead the incident response effort. This coordinator should be empowered to dynamically assemble a team based on the incident, up to and including overriding conventional job duties for the duration of the incident. IRT members

should also have dedicated communication protocols in place, such as a phone tree or instant messaging group, allowing them to receive information on incidents in a timely manner.

- **Root-cause analysis**: Once an incident has been resolved, the IRT should perform a root-cause analysis, document these findings, and offer suggestions for preventing the incident in the future. Some incidents, such as natural disasters, may be unavoidable, but the organization may be able to implement proactive measures to reduce their impact, such as relocating staff ahead of a forecastable natural disaster. Incident response plans and procedures should also be updated as appropriate to help the organization better respond in the future to incidents of a similar nature.

Incident Classification

To ensure an incident is dealt with correctly, it is important to determine how critical it is and prioritize the response appropriately. Each organization may classify incidents differently, but a generic scheme plots Urgency against Impact. These are assigned values from Low, Medium, or High, and incidents that are High priority are handled first. The following are descriptions and examples of these criteria:

- **Impact**: How much or how significantly does the incident degrade IT services? An issue with identity management that forces users to try logging in multiple times before successfully accessing a system is irritating but has a minor effect on operations. A data center being completely unreachable due to cut network cables means a complete loss of the business function and should be dealt with before other issues.
- **Urgency**: How soon does the organization need resolution? An outage on a system for internal staff social information like team sporting events is unlikely to be critical and can wait for restoration if a mission-critical system is also suffering an outage.

Incident classification criteria and examples should be documented for easy reference. The IRP should contain this information, and it is also advisable to include it in any supporting systems like incident trackers or ticketing systems. These ratings are subjective and may change as the incident is investigated, so the IR coordinator should ensure that critical information like prioritization is communicated to the team.

Incident Response Phases

The organization's IRP should include detailed steps broken down by phases. At a high level, there are activities to be conducted prior to an incident and after an incident occurs; namely, planning the IR capability and the actual execution of a response when

an incident is detected. There are a number of IR models that contain slightly different definitions for each phase, but in general they all contain the following:

- **Prepare**: Preparation is the phase where the IR capability's foundation is established. During this phase, the IRP and IRT should be documented, training given to IRT members, and adequate detection abilities implemented.
- **Detect**: To be effective, the IR capability requires the ability to detect and draw attention to events that could negatively impact the organization's operations. This is most often in the form of the organization's continuous monitoring tools, which can identify anomalous activity and alert SOC personnel trained to analyze, prioritize, and initiate response procedures. Other methods of detection can include noncontinuous monitoring activities like routine audits, user-reported issues such as unexpected application or system behavior, and even external entities like security researchers.

  Security researchers or malicious actors may draw attention to a vulnerability they have discovered or, worse, exploited, in which case the organization must take steps to investigate if the claim is true and take appropriate action. Organizations are also beginning to subscribe to external intelligence feeds from third-party services that can provide advanced alert of an incident, such as compromised user credentials showing up on the dark web, or adjacent domains being registered that might be used in a phishing attack.

  As soon as an incident is detected, it must be documented, and all actions from the point of detection through to resolution should be documented as well. Initial analysis or triage of incidents, prioritization, members of the IRT called upon to deal with the incident, and plans for implementing recovery strategies should be documented. As discussed in the section on digital forensics, it may be the case that a seemingly simple incident evolves into a malicious act requiring criminal charges. In this case, as much evidence as possible, handled correctly, will be crucial and cannot be created after the fact.

  Investigation will begin as the IRT starts to gather information about the incident. This can be as simple as attempting to reproduce a user-reported app issue to determine if it is only affecting that user or is a system-wide issue. This is a key integration point to the practice of digital forensics. As soon as it appears, the incident may require prosecution or escalation to law enforcement, and appropriately trained digital forensics experts must be brought in.

  Notification may also occur during this stage, once the incident is properly categorized. In many cases, this will be done to satisfy legal or compliance obligations, such as U.S. state privacy laws and the EU GDPR, which require

notification to authorities within a certain timeframe after an incident is detected. Appropriate communication resources like legal counsel or public relations personnel may be required on the IRT to handle these incidents.

- **Respond**: Once the organization has been alerted to an incident, its IR capability is activated, and an appropriate remediation must be developed. In most incidents, the overriding concern during the response phase should be containing the incident; that is to say, preventing it from causing further damage. In some cases, gathering information about the attack may be required, during which time the incident continues to occur; in this case, the extra information gained must be more valuable than the damage caused by the ongoing attack.

  Once sufficient information is gathered, a containment strategy must be formulated and implemented. This should follow the documented scenario-based responses in the IRP whenever possible; for example, responding to a ransomware attack by isolating any affected machines and working to establish how the malware was installed. Once this is ascertained, deploying techniques for preventing other machines from being affected may be more important than recovering data on machines that have already been compromised. These types of decisions should be made by qualified personnel with as much information as possible at their disposal.

  Notification is typically handled as part of incident detection, and formal reporting is an ongoing task that starts after the initial notification. In simple cases, the report may comprise a single document at the resolution of an incident with the particulars, while in other cases, ongoing reporting of a dynamic situation will be required. As you may have seen during high-profile data breaches at public organizations, initial reporting is made on the incident with information available at that moment, even if it is estimated. As the investigation proceeds, updated reports are delivered to convey new information. The IR coordinator, along with appropriate communication resources on the IRT, is responsible for creating and disseminating all required reporting to appropriate stakeholders.

  During the response phase, additional information should be gathered and documented regarding the incident. This may include details of any external attacker or malicious insider who might be responsible and should be conducted in accordance with the digital evidence handling previously discussed.

- **Recover**: This phase may actually be started as soon as the detection phase and continue until the cause of the incident is completely eradicated and normal operations have resumed. Actions like changing a password for a user who responded to a phishing email or following documented procedures like restarting a server for unexpected app behavior can be recovery actions.

Eradication of the underlying cause is also a critical element of the recovery. This may include actions such as replacing faulty hardware, blocking a specific IP address or user from accessing a resource, or rebuilding compromised systems from known good images. Containment prevents the incident from spreading further, while eradication removes the immediate cause of the incident. It may be the case that neither prevents the problem from reoccurring in the future, which is a part of post-incident activities.

Recovery and eradication end when the organization returns to normal operations. This is defined as the pre-incident service level delivered using standard procedures.

- **Post-incident**: Once normal operations have resumed, the IRT should ensure all documentation from the incident is properly created and stored. Lessons learned from conducting the response should be gathered and used to improve the incident response process itself. Finally, the organization should conduct a root-cause analysis to identify any underlying causes and appropriate steps to prevent the incident from recurring in the future.

Incident response is designed to help the organization restore normal operations quickly, but there are situations when this will be impossible. In these cases, the incident may need to be upgraded to an *interruption*, which is an event whose impact is significant enough to disrupt the organization's ability to achieve its goals or mission. A few users with malware infections on their workstations is an incident that can likely be handled by normal IT resources, but an outbreak affecting all critical systems and a large percentage of users is likely to require more resources.

In such cases, the IR coordinator may be empowered to declare an interruption or disaster, which invokes processes like BCDR. Similar to IR, there should be clear plans in place to guide the organization's response, such as emergency authorization to buy new equipment outside of normal purchasing processes or invocation of alternate procedures like using a third party to process information. The IR coordinator should be aware of their role and responsibilities in these plans, including the process for declaring and notifying appropriate members of the BCDR team to take over.

Cloud-Specific Incident Management

As discussed, the migration to a third-party CSP can introduce additional complexity to an organization. When planning for incident management, the CSP must be considered as a critical stakeholder. Appropriate points of contact should be documented and reachable in the event of an incident, such as a service delivery or account manager who can

support the organization's incident response and recovery. An incident at the CSP should be communicated to all the CSP's consumers, and the CSP may be required to provide specific information in the case of regulated data or negligence. Even if a data breach is the fault of the CSP, the consumer who is the data controller is still legally liable on several counts, including notifying affected individuals and possible fines.

Communication from a consumer to the CSP may be critical, especially if the incident has the ability to affect other CSP customers. Additionally, some options for performing incident management, such as rebuilding compromised architecture, will be different in the cloud environment. It is possible to rapidly redeploy completely virtual architecture to a known good state in the cloud; the same task in a traditional data center environment could take significant time.

Incident Management Standards

There are three standards that can guide a security practitioner in designing and operating an incident management capability.

- **Carnegie Mellon University Software Engineering Institute (SEI) – Incident Management Capability Assessment**: SEI publishes a variety of capability maturity models, which can be useful for organizations assessing how robust their procedures are currently and identifying opportunities for future improvement. This model is freely available and utilizes categories to break down essential activities, including Prepare, Protect, Detect, Respond, and Sustain. Within each category are subcategory activities designed to help the organization proactively build the IR capability, respond when incidents occur, and continuously improve the capability.

  You can find the technical report document at `resources.sei.cmu.edu/asset_files/TechnicalReport/2018_005_001_538866.pdf`.

- **NIST SP 800-61, Computer Security Incident Handling Guide**: This NIST standard is also freely available and breaks incident handling down into four high-level phases: Preparation, Detection and Analysis, Containment Eradication and Recovery, and Post-Incident Activity. It focuses on creating incident handling checklists designed to speed response times and is a useful guide for organizations in the U.S. federal government or those following other NIST standards such as the RMF or using NIST SP 800-53 as a security control framework.

  You can find NIST SP 800-61 here: `csrc.nist.gov/publications/detail/sp/800-61/rev-2/final`.

- **ISO 27035**: As with all ISO standards. there are several documents in this standard, including *Part 1: Principles of Incident Management*, *Part 2: Guidelines to Plan and Prepare for Incident Response*, and *Part 3: Guidelines for ICT Incident Response Operations*. The standard is similar to other frameworks including a phased approach broken down into pre-, during-, and post-incident steps. Unlike other frameworks, these documents are not freely available. The steps are most closely aligned with the security controls framework and implementation approach outlined in the rest of the ISO 27000 standard.

SUMMARY

Cloud security operations, like all other security practices, must be anchored by two key principles: operations must be driven by the organization's business objectives or mission and must preserve the confidentiality, integrity, and availability of data and systems in the cloud. Operations is a far-reaching topic covering the selection, implementation, and monitoring of physical and logical infrastructure, as well as security controls designed to address the risks posed in cloud computing.

There are a variety of standards that can assist the CCSP in implementing or managing security controls for cloud environments. These cover major objectives such as access control, securing network activity, designing operational control programs, and handling communications. Choosing the correct standard will be driven by each organization's location, industry, and possibly costs associated with the various standards. All programs implemented should have feedback mechanisms designed to continuously improve security as risks evolve.

Also key is an understanding of the shared responsibility model. CSPs will perform the majority of work related to physical infrastructure, though cloud consumers may need physical security for infrastructure that connects them to cloud computing. Logical infrastructure is a more equally shared responsibility: in non-SaaS models, the CSP runs the underlying infrastructure, but consumers have key responsibilities in securing the logical infrastructure in their virtual slice of the cloud.

DOMAIN 6

CCSP®

Legal, Risk, and Compliance

THE CLOUD OFFERS COMPANIES and individuals access to vast amounts of computing power at economies of scale only made possible by distributed architectures. However, those same distributed architectures can bring a unique set of risks and legal challenges to companies due to the geographic distribution of services the cloud provides. The cloud, powered by the Internet, flows freely across the borders of countries as information is stored in data centers worldwide and travels the globe. Determining what laws apply to cloud computing environments is an ongoing challenge, as in many situations, the company may be based in one country, host services in another, and be serving customers across the globe. In addition, the convenience and scalability offered by the cloud necessitates the acceptance of additional risks introduced by the reliance on a third-party provider.

ARTICULATING LEGAL REQUIREMENTS AND UNIQUE RISKS WITHIN THE CLOUD ENVIRONMENT

One indisputable fact that has arisen with cloud computing is that the legal and compliance efforts of computing have become more complex and time-consuming. With data and compute power spread across countries and continents, international disputes have dramatically increased. Hardly a day goes by without a news story about a conflict relating

to violations of intellectual property, copyright infringements, and data breaches. These types of disputes existed well before the ascendence of cloud computing, but they have definitely become more commonplace and costly. It is in many ways the simplicity of cloud computing to have services span jurisdictions that has led to the marked increase in complexity of compliance. To prepare for these challenges, a cloud security professional must be aware of the legal requirements and unique risks presented by cloud computing architectures.

Conflicting International Legislation

With cloud technologies comes the inherent power to distribute computing across countries and continents. Though it is certainly possible to use the cloud on a local or regional scale through configuration or policy, many cloud customers use cloud services across regions and jurisdictions to provide redundancy and closer service delivery across the crowded Internet. So how do companies address the fact that their servers, customers, and physical locations may in fact be governed by completely separate sets of laws? It is a constant challenge that requires cloud security practitioners to be familiar with multiple sets of laws and when they might apply. Take, for instance, the law of the land in the European Union (the EU, a group of 27 European countries that share a common set of laws and economic policies). The EU is governed by data privacy laws known as the General Data Protection Regulations (GDPR) that we will cover later in the chapter. Do these laws apply to U.S.-based companies doing business in the EU or with EU citizens? What about in situations where EU laws conflict with state and federal laws in the United States? Because of the international nature of cloud offerings and customers, cloud practitioners must be aware of multiple sets of laws and regulations and the risks introduced by conflicting legislation across jurisdictions. These conflicts may include the following:

- Copyright and intellectual property law, particularly the jurisdictions that companies need to deal with (local versus international) to protect and enforce their IP protections
- Safeguards and security controls required for privacy compliance, particularly details of data residency or the ability to move data between countries, as well as varying requirements of due care in different jurisdictions
- Data breaches and their aftermath, particularly breach notification
- International import/export laws, particularly technologies that may be sensitive or illegal under various international agreements

Craig Mundie, the former chief of Microsoft's research and strategy divisions, explained it in these terms:

> *People still talk about the geopolitics of oil. But now we have to talk about the geopolitics of technology. Technology is creating a new type of interaction of a geopolitical scale and importance. . . . We are trying to retrofit a governance structure which was derived from geographic borders. But we live in a borderless world.*

In simple terms, a cloud security practitioner must be familiar with a number of legal arenas when evaluating risks associated with a cloud computing environment. The simplicity and inexpensive effort of configuring computing and storage capabilities across multiple countries and legal domains has led to a vast increase in the complexity of legal and compliance issues.

Evaluation of Legal Risks Specific to Cloud Computing

The cloud offers delivery and disaster recovery possibilities unheard of a few decades ago. Cloud service providers can offer content delivery options to host data within a few hundred miles of almost any human being on Earth. Customers are not limited by political borders when accessing services from cloud providers. However, the ability to scale across the Internet to hundreds of locations brings a new set of risks to companies taking advantage of these services. Storing data in multiple foreign countries introduces legal and regulatory challenges. Cloud computing customers may be impacted by one or more of the following legislative items:

- Differing legal requirements; for example, state laws in the United States requiring data breach notifications, which have varying timeframes and state-level reporting entities versus international with country-level reporting entities.
- Different legal systems and frameworks in different countries, such as legislation versus case law/precedent. A Certified Cloud Security Professional (CCSP) shouldn't be a master of all but should acquire advice from competent legal counsel.
- Challenges of conflicting law, such as EU GDPR and the U.S. CLOUD Act, one of which requires privacy and the other that mandates disclosure. A CCSP needs to be aware that these competing interests exist and may need to be used to drive business decisions (for example, corporate structure to avoid the conflict.)

Legal Frameworks and Guidelines That Affect Cloud Computing

Cloud security practitioners should be keenly aware of a number of legal frameworks that may affect the cloud computing environments they maintain. The following frameworks are the products of multinational organizations working together to identify key priorities in the security of information systems and data privacy.

The Organization for Economic Cooperation and Development

The Organization for Economic Cooperation and Development (OECD) guidelines lay out privacy and security guidelines. (See `www.oecd.org/sti/ieconomy/privacy-guidelines.htm`.) The OECD guidelines are echoed in European privacy law in many instances. The basic principles of privacy in the OECD include the following:

- **Collection limitation principle**: There should be limits on the collection of personal data as well as consent from the data subject.
- **Data quality principle**: Personal data should be accurate, complete, and kept up-to-date.
- **Purpose specification principle**: The purpose of data collection should be specified, and data use should be limited to these stated purposes.
- **Use limitation principle**: Data should not be used or disclosed without the consent of the data subject or by the authority of law.
- **Security safeguards principle**: Personal data must be protected by reasonable security safeguards against unauthorized access, destruction, use, or disclosure.
- **Openness principle**: Policies and practices about personal data should be freely disclosed, including the identity of data controllers.
- **Individual participation principle**: Individuals have the right to know if data is collected on them, access any personal data that might be collected, and obtain or destroy personal data if desired.
- **Accountability principle**: A data controller should be accountable for compliance with all measures and principles.

In addition to the basic principles of privacy, there are two overarching themes reflected in the OECD: first, a focus on using risk management to approach privacy protection, and second, the concept that privacy has a global dimension that must be addressed by international cooperation and interoperability. The OECD council adopted guidelines in September 2015, which provide guidance in the following:

- **National privacy strategies**: Privacy requires a national strategy coordinated at the highest levels of government. CCSPs should be aware of the common elements of national privacy strategies based on the OECD suggestions and work to help their organizations design privacy programs that can be used across multiple jurisdictions. Additionally, a CCSP should be prepared to help their organization design a privacy strategy using business requirements that deliver a good cost-benefit outcome.
- **Data security breach notification**: Both the relevant authorities and the affected individual(s) must be notified of a data breach. CCSPs should be aware that

multiple authorities may be involved in notifications and that the obligation to notify authorities and individuals rests with the data controller.

- **Privacy management programs**: These are operational mechanisms to implement privacy protection. A CCSP should be familiar with the programs defined in the OECD guidelines.

Asia Pacific Economic Cooperation Privacy Framework

The Asia Pacific Economic Cooperation Privacy Framework (APEC) is an intergovernmental forum consisting of 21 member economies in the Pacific Rim. (See `www.aicpa.org/content/dam/aicpa/interestareas/informationtechnology/resources/privacy/downloadabledocuments/10252-346-records-management-pro.pdf`.) The goal of this framework is to promote a consistency of approach to information privacy protection. This framework is based on nine principles.

- **Preventing harm**: An individual has a legitimate expectation of privacy, and information protection should be designed to prevent the misuse of personal information.
- **Collection limitation**: Collection of personal data should be limited to the intended purposes of collection and should be obtained by lawful and fair means with notice and consent of the individual. As an example, an organization running a marketing operation should not collect Social Security or national identity numbers, as they are not required for sending marketing materials.
- **Notice**: Information controllers should provide clear and obvious statements about the personal data that they are collecting and the policies around use of the data. This notice should be provided at the time of collection. You are undoubtedly familiar with the banners and pop-ups in use at many websites to notify users of what data is being collected.
- **Use of personal information**: Personal information collected should be used only to fulfill the purposes of collection and other compatible or related purposes except: a) with the consent of the individual whose personal information is collected; b) when necessary to provide a service or product requested by the individual; or, c) by the authority of law and other legal instruments, proclamations, and pronouncements of legal effect.
- **Integrity of personal information**: Personal information should be accurate, complete, and kept up-to-date to the extent necessary for the purposes of use.
- **Choice and consent**: Where appropriate, individuals should be provided with clear, prominent, easily understandable, accessible, and affordable mechanisms to exercise choice in relation to the collection, use, and disclosure of their personal information.

- **Security safeguards**: Personal information controllers should protect personal information that they hold with appropriate safeguards against risks, such as loss or unauthorized access to personal information or unauthorized destruction, use, modification, or disclosure of information or other misuses.
- **Access and correction**: Individuals should be able to obtain from the personal information controller confirmation of whether the personal information controller holds personal information about them, and have access to information held about them, challenge the accuracy of information relating to them, and have the information rectified, completed, amended, or deleted.
- **Accountability:** A personal information controller should be accountable for complying with measures. Companies must identify who is responsible for complying with these privacy principles.

General Data Protect Regulation

The General Data Protect Regulation (GDPR) is perhaps the most far-reaching and comprehensive set of laws ever written to protect data privacy. (See `gdpr.eu/what-is-gdpr`.) In the EU, the GDPR mandates privacy for individuals, defines companies' duties to protect personal data, and prescribes punishments for companies violating these laws. GDRP fines for violating personal privacy can be massive at 20 million Euros or 4 percent of global revenue (whichever is greater). For this reason alone, a CCSP should be familiar with these laws and the effects that they have on any company operating within, housing data in, or doing business with citizens of these countries in the 27-nation bloc. GDPR became the law of the land in May 2018, and incorporated many of the same principles of the former EU Data Protection Directive.

GDPR identifies formally the role of the data subject, controller, and processor. The data subject is defined as an "identified or identifiable natural person," or, more simply, a person. The GDPR draws a distinction between a data controller and a data processor to formally recognize that not all organizations involved in the use and processing of personal data have the same degree of responsibility.

- A data controller under GDPR determines the purposes and means of processing personal data.
- A data processor is the body responsible for processing the data on behalf of the controller.

In cloud environments, this is often separate companies, where a company may be providing services (the data controller) on a cloud provider (the data processor).

NOTE Keep in mind that privacy does not equal data security, and sometimes the two ideals are at odds.

The GDPR is a massive set of regulations that covers almost 90 pages of details and is a daunting challenge to many organizations. A CCSP should be familiar with the broad areas of the law, but ultimately organizations should consult an attorney to ensure that operations are compliant. GDRP encompasses the following main areas:

- **Data protection principles**: If a company processes data, it must do so according to seven protection and accountability principles.
  - **Lawfulness, fairness, and transparency**: Processing must be lawful, fair, and transparent to the data subject.
  - **Purpose limitation**: The organization must process data for the legitimate purposes specified explicitly to the data subject when it was collected.
  - **Data minimization**: The organization should collect and process only as much data as absolutely necessary for the purposes specified.
  - **Accuracy**: The organization must keep personal data accurate and up-to-date.
  - **Storage limitation**: The organization may store personally identifying data for only as long as necessary for the specified purpose.
  - **Integrity and confidentiality**: Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (for example, by using encryption).
  - **Accountability**: The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.
- **Data security**: Companies are required to handle data securely by implementing appropriate technical measures and to consider data protection as part of any new product or activity.
- **Data processing**: GDPR limits when it is legal to actually process user data. There are very specific instances when this is allowed under the law.
- **Consent**: Strict rules are in place for how a user is to be notified that data is being collected.
- **Personal privacy**: The GDPR implements a litany of privacy rights for data subjects, which gives individuals far greater rights over their data that companies collect. This includes rights to be informed, access, rectify, restrict, obtain, and the much touted "right to be forgotten," which allows the end user to request deletion of all personal data.

Additional Legal Controls

The cloud service provider has also significantly complicated the legal landscape when it comes to legal controls. The provider of the infrastructure, platform, or software as a service must be evaluated as an additional party when considering compliance. If a cloud provider is responsible for a violation of a legal control, the company can still be responsible for the legal ramifications. This applies to legal and contractual controls such as the following:

- **Heath Insurance Portability and Accountability Act (HIPAA)**: This 1996 U.S. law regulates the privacy and control of health information data.
- **Payment Card Industry Data Security Standard (PCI DSS)**: The current standard version 3.2 affects companies that accept, process, or receive electronic payments.
- **Privacy Shield**: This provides a voluntary privacy compliance framework through which companies can comply with portions of the GPDR in order to facilitate the movement of EU citizens' data into the United States by companies based in the United States. Note that this agreement is facing legal challenges in the EU.
- **Sarbanes–Oxley Act (SOX)**: This law was enacted in 2002 and sets requirements for U.S. public companies to protect financial data when stored and used. It is intended to protect shareholders of the company as well as the general public from accounting errors or fraud within enterprises. This act specifically applies to publicly traded companies and is enforced by the Securities and Exchange Commission (SEC). It is applicable to cloud practitioners in particular because it specifies what records must be stored and for how long, internal control reports, and formal data security policies.
- **Gramm–Leach–Bliley Act (GBLA):** U.S. federal law that requires financial institutions to explain how they share and protect their customers' private information.

Laws and Regulations

The cloud represents a dynamic and changing environment, and monitoring/reviewing legal requirements is essential to staying compliant. All contractual obligations and acceptance of requirements by contractors, partners, legal teams, and third parties should be subject to periodic review. When it comes to compliance, words have real meaning. You (and Microsoft Word's thesaurus) might think that the terms *law* and *regulation* are interchangeable, but when it comes to data privacy issues, they can have starkly different outcomes if they are not satisfied. Understanding the type of compliance that your company is subject to is vital to a CCSP. The outcome of noncompliance can vary greatly, including imprisonment, fines, litigation, loss of a contract, or a combination of

these. To understand this landscape, a CCSP must first recognize the difference between data protected by legal regulations and data protected by contract.

- **Statutory requirements** are required by law.
  - U.S. federal laws, such as HIPAA, GLBA, and SOX outlined in the previous section
  - Additional federal laws such as the Family Education Rights and Privacy Act (FERPA), which deals with privacy rights for students, and the Federal Information Security Management Act (FISMA), which protects federal data privacy
  - State data privacy laws, which now exist in all 50 states as well as U.S. territories
  - International laws, which we will discuss in a later section covering country-specific laws
- **Regulatory requirements** may also be required by law, but refer to rules issued by a regulatory body that is appointed by a government entity.
  - In the United States, many regulatory examples exist, mostly dealing with contractor compliance to do business with the government. An example would be security requirements outlined by NIST that are necessary to handle government data (such as NIST 800-171).
  - At the state level, several states use regulatory bodies to implement their cybersecurity requirements. One example is the New York State Department of Financial Services (NY DFS) cybersecurity framework for the financial industry (23 NYCRR 500).
  - There are numerous international requirements dealing with cloud legal requirements (including APEC and GDPR).
- **Contractual requirements** are required by a legal contract between private parties. These are not laws, and breaching these requirements cannot result in imprisonment (but can result in financial penalties, litigation, and terminations of contracts). Common contractual compliance requirements include the following:
  - Payment Card Industry Data Security Standard (PCI DSS)
  - Financial Industry Regulatory Authority (FINRA)
  - Service Organization Controls (SOC)
  - Generally Accepted Privacy Principles (GAPP)
  - Center for Internet Security (CIS) Critical Security Controls (CSC)
  - Cloud Security Alliance (CSA) Cloud Controls Matrix (CCM)

It is vital to consider the legal and contractual issues that apply to how a company collects, stores, processes, and ultimately deletes data. Few companies or entities have the luxury of not considering the important national and international laws that must be complied with to do business. Perhaps not surprisingly, company officers have a vested interest in complying with these laws. Laws and regulations are specific in who is responsible for the protection of information. Federal laws like HIPAA spell out that senior officers within a company are responsible for (and liable for) the protection of data. International regulations like GDPR and state regulations such as NYCRR 500 identify the role of a data protection officer or chief information security officer and outline their culpability for any data loss incident.

If you are using cloud infrastructure from a cloud provider, you the customer must impose all legal, regulatory, and contractual obligations that are imposed on you by relevant legislation onto your cloud provider. No matter who is hosting your data or services, you the customer are accountable for compliance of effective security controls and data security.

Forensics and eDiscovery in the Cloud

When a crime is committed, law enforcement or other agencies may perform eDiscovery using forensic practices to gather evidence about the crime that has been committed and to prosecute the guilty parties. eDiscovery is defined as any process in which electronic data is pursued, located, secured, and searched with the intent of using it as evidence in a civil or criminal legal case. In a typical eDiscovery case, computing data might be reviewed offline (with the equipment powered off or viewed with a static image) or online (with the equipment on and accessible). In the cloud environment, almost all eDiscovery cases will be done online due to the nature of distributed computing and the difficulty in taking those systems offline.

Forensics, especially cloud forensics, is a highly specialized field that relies on expert technicians to perform the detailed investigations required to adequately perform discovery while not compromising the potential evidence and chain of custody. In any forensic investigation, a CCSP can typically expect to work with an expert outside consultant or law enforcement personnel.

eDiscovery Challenges in the Cloud

For a security professional, the challenges of eDiscovery are greatly increased by the use of cloud technologies. If you received a call from an outside law enforcement agency (never a phone call that anyone likes to receive) about criminal activities on one of your servers in your on-premise data centers, you might immediately take that server offline in order to image it, always having the data firmly in your possession. Now imagine if the same service was hosted in a SaaS model in the cloud distributed across 100 countries using your content delivery network. Could you get the same data easily? Could you maintain possession, custody, or control of the electronically stored information? Who has jurisdiction in

any investigation? Can a law enforcement agency compel you to provide data housed in another country? All of these questions make cloud eDiscovery a distinct challenge.

eDiscovery Considerations

When you're considering a cloud vendor, eDiscovery should be considered during any contract negotiation phase. If you think about your cloud providers today, can you recall how many conversations you might have had with their support teams? Would you know who to contact if an investigation was needed?

To pose another question, do you actually know where the data is physically housed? Many companies make use of the incredible replication and fail-over technologies of the modern cloud, but that means data may be in different time zones, countries, or continents when an investigation is launched. As we have outlined, laws in one country or state may in fact clash or break with laws of another. It is the duty of the cloud security profession to exercise due care and due diligence in documenting and understanding to the best of their ability the relevant laws and statutes pertaining to an investigation prior to it commencing.

In some cases, proactive architectural decisions can simplify issues down the road. For example, if a company chooses to house customer data within their legally defined jurisdiction (for example, EU citizen data is always housed on cloud servers within the EU), it can simplify any potential investigations or jurisdictional arguments that might arise.

Conducting eDiscovery Investigations

Where there is a problem in technology, there is usually a solution, and the marketplace for eDiscovery tools has grown commensurately with the explosion in the use of cloud service providers. There are a healthy number of vendors that specialize in this area. For a CCSP, it is important to learn about eDiscovery actors when evaluating contracts and service level agreements (SLAs) to determine what forms of eDiscovery are permissible. These are the main types of investigatory methods:

- **SaaS-based eDiscovery**: Why not use the cloud to perform eDiscovery on the cloud? These tools from a variety of vendors are used by investigators and law firms to collect, preserve, and review data.
- **Hosted eDiscovery**: In this method, your hosting provider includes eDiscovery services in the contractual obligations that can be exercised when and if necessary. This may limit a customer to a preselected list of forensic solutions, as CSPs are often wary of customer-provided tools due to the potential for affecting other cloud customers during the process.
- **Third-party eDiscovery**: When there are no obligations in a contract, a third party may be contracted to perform eDiscovery operations. In this model, the CCSP must work especially closely with the third party to provide appropriate access to necessary resources in the cloud.

Cloud Forensics and Standards

As you can see, the forensic and eDiscovery requirements for many legal controls are greatly complicated by the cloud environment. Unlike on-premises systems, it can be difficult or impossible to perform physical search and seizure of cloud resources such as storage or hard drives. Standards from bodies such as the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) and the Cloud Security Alliance (CSA) provide guidance to cloud security practitioners on best practices for collecting digital evidence and conducting forensics investigations in cloud environments. CCSPs should be familiar with the following standards:

- **Cloud Security Alliance**: The CSA Security Guidance Domain 3: Legal Issues: Contracts and Electronic Discovery highlights some of the legal aspects raised by cloud computing. Of particular importance to CCSPs are the guidance provided on negotiating contracts with cloud service providers in regard to eDiscovery, searchability, and preservation of data.
- **ISO/IEC 27037:2012**: This provides guidelines for the handling of digital evidence, which include the identification, collection, acquisition, and preservation of data related to a specific case.
- **ISO/IEC 27041:2014-01**: This provides guidance on mechanisms for ensuring that methods and processes used in the investigation of information security incidents are "fit for purpose." CCSPs should pay close attention to sections on how vendor and third-party testing can be used to assist with assurance processes.
- **ISO-IEC 27042:2014-01**: This standard is a guideline for the analysis and interpretation of digital evidence. A CCSP can use these methods to demonstrate proficiency and competence with an investigative team.
- **ISO/IEC 27043**: The security techniques document covers incident investigation principles and processes. This can help a CCSP as a "playbook" for many types of incidents, including unauthorized access, data corruption, system crashes, information security breaches, and other digital investigations.
- **ISO/IEC 27050-1**: This standard covers electronic discovery, the process of discovering pertinent electronically stored information involved in an investigation.

UNDERSTANDING PRIVACY ISSUES

The Internet age has brought with it an unprecedented amount of information flow. Knowledge (and cat videos) is shared by billions of Internet-connected humans at the speed of light around the world every day. The flow of information is two-way, however,

and corporations and governments work hard to analyze how you interact with the Internet, both from your web activity as well as from your phone habits. Billions of consumers have installed listening devices in their homes (think Google Home and Amazon Alexa) and carry a sophisticated tracking device willingly on their person each and every day (the smart phone). All of this data allows better marketing investments as well as more customized recommendations for consumers.

With the advent of cloud computing, the facilities and servers used to store and process all of that user data can be located anywhere on the globe. To ensure availability and recoverability, it is good practice to ensure that data is available or at least backed up in multiple geographic zones to prevent loss due to widespread natural disasters such as an earthquake or hurricane. Indeed, in certain countries (like the tiny nation of Singapore at under 300 square miles), it would be impossible to protect from natural disasters without sending data to additional countries (and therefore legal jurisdictions).

Difference between Contractual and Regulated Private Data

In any cloud computing environment, the legal responsibility for data privacy and protection rests with the cloud consumer, who may enlist the services of a cloud service provider (CSP) to gather, store, and process that data. If your company uses Amazon Web Services (AWS) to host your website that gathers customer data, the responsibility for that data rests with you. You as the data controller are responsible for ensuring that the requirements for protection and compliance are met and that your contracts with the cloud service provider stipulate what requirements must be met by the CSP.

There are a number of terms that describe data that might be regulated or contractually protected. Personally identifiable information (PII) is a widely recognized classification of data that is almost universally regulated. PII is defined by the NIST standard 800-122 as "any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information." (See `nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf`.)

Protected health information (PHI) was codified under the HIPAA statutes in 1996. Any data that might relate to a patient's health, treatment, or billing that could identify a patient is PHI. When this data is electronically stored, it must be secured using best practices, such as unique user accounts for every user, strong passwords and MFA, principles of least privilege for accessing PHI data, and auditing all access and changes to a patient's PHI data.

Table 6.1 outlines types of regulated private data and how they differ.

TABLE 6.1 **Types of Regulated Data**

DATA TYPE	DEFINITION
Personally identifiable information (PII)	Personally identifiable information is information that, when used alone or with other relevant data, can identify an individual. PII may contain direct identifiers (such as SSN numbers) that can identify a person uniquely, or quasi-identifiers (such as race) that can be combined with other quasi-identifiers (such as date of birth) to successfully recognize an individual.
Protected health information (PHI)	Protected health information, also referred to as personal health information, generally refers to demographic information, medical histories, test and laboratory results, mental health conditions, insurance information, and other data that a healthcare professional collects to identify an individual and determine appropriate care.
Payment data	Bank and credit card data used by payment processors in order to conduct point-of-sale transactions.

There are two main types of regulated private data that we will discuss in further detail.

Contractual Private Data

When your organization collects, processes, transmits, or stores private data as part of normal business processes, the information must be protected according to relevant laws and regulations. If an organization outsources any of those functions to a third party such as a cloud service provider, the contracts must specify what the applicable rules and requirements are and what laws that the cloud service provider must adhere to.

For example, if a company is handling PHI that must be processed within the United States, any contracts with a CSP must clearly specify that fact. It is the obligation of the data owner to make all terms and conditions clear and transparent for the outsourcing contract.

Regulated Private Data

The biggest differentiator between contractual and regulated data is that regulated data must be protected under legal and statutory requirements, while contractual data is protected by legal agreements between private parties. Both PII and PHI data are subject to regulation, and the disclosure, loss, or altering of these data can subject a company (and individuals) to statutory penalties including fines and imprisonment. In the United States, protection of PHI data is required by HIPAA, which provides requirements and penalties for failing to meet obligations.

Regulations are put into place by governments and government-empowered agencies to protect entities and individuals from risks. In addition, they force providers and processors to take appropriate measures to ensure protections are in place while identifying penalties for lapses in procedures and processes.

One of the major differentiators between contracted and regulated privacy data is in breach reporting. A data breach of regulated data (or unintentional loss of confidentiality of data through theft or negligence) is covered by myriad state and federal laws in the United States. There are now data privacy laws covering PII in all 50 states, with legislative changes happening every year. Some states have financial penalties of as little as $500 for data privacy violations, while others have fines in the hundreds of thousands of dollars as well as prison time for company officers. The largest state fine for a data privacy breach was handed to Uber for a 2016 breach and totaled over $148 million dollars in fines to the state of California.

Components of a Contract

As we have detailed, the outsourcing of services to a cloud service provider in no way outsources the risk to the company as the data owner. A CCSP must be familiar with key concepts when contracting with cloud service providers to ensure that both parties understand the scope and provisions when private data is concerned. Key contract areas include the following:

- **Scope of data processing**: The provider must have a clear definition of the permitted forms of data processing. For example, data collected on user interactions for a cloud customer should not be used by the CSP to inform new interface designs.
- **Subcontractors**: It is important to know exactly where all processing, transmission, storage, and use of data will take place and whether any part of the process might be undertaken by a subcontractor to the cloud service provider. If a subcontractor will be used in any phase of data handling, it is vital that the customer is contractually informed of this development.
- **Deletion of data**: How data is to be removed or deleted from a cloud service provider is important and helps protect against unintentional data disclosure. In the old days, data was often physically destroyed (with a screwdriver through a hard drive). In the cloud environment, that is next to impossible, so contracts must spell out the methods that data is to be securely removed from any environment (and subcontractor environment). One example is cryptoshredding, or rendering data inaccessible by cryptographic methods.
- **Data security controls**: If data requires a level of security controls when it is processed, stored, or transmitted, that same level of security control must be ensured

in a contract with a cloud service provider. Ideally, the level of data security controls would exceed what is required (as is often the case for cloud service providers). Typical security controls include encryption of data while in transit and while at rest, access control and auditing, layered security approaches, and defense-in-depth measures.

- **Physical location of data**: Knowledge of the physical location of where data is stored and processed (and, in some cases, transferred) is vital to ensure compliance with legal and regulatory requirements. Contractually indicating what geographic locations are acceptable for data storage and processing (and failover/backup) can protect a company from falling victim to unintended legal consequences. In addition to the location of the data, it is important to consider the location of the cloud service provider's offices and the offices of any subcontractors in use.
- **Return or surrender of data**: How does one break up with a cloud service provider? They know a lot about you, and they have your data. It is important to include in any contract the termination agreements, which should specify how and how quickly data is to be returned to the data owner. In addition, the contract should outline how returned data is ensured to be removed from the cloud service provider's infrastructure. Cloud data destruction standards should be identified when entering into a contract to ensure that methods are standardized and documented when data is returned or destroyed.
- **Audits**: Right to audit clauses should be included to allow a data owner or an independent consultant to audit compliance to any of the areas identified earlier. In practice, most CSPs will agree to be bound by standardized audit reports such as SOC 2 or ISO 27001, and most contracts specify the frequency that these reports must be produced.

Country-Specific Legislation Related to Private Data

If by now you're sensing a theme that data privacy is a complicated problem, it's about to get reinforced. An individual's right to privacy (and therefore their right to have their data handled in a secure and confidential way) varies widely by country and culture. As a result, the statutory and regulatory obligations around data privacy are vastly different depending on the citizenship of a data subject, the location of data being stored, the jurisdiction of the data processor, and even multinational agreements in place that may require data disclosure to government entities upon request.

There are many different attitudes and expectations of privacy in countries around the world. In some more authoritarian regimes, the data privacy rights of the individual are almost nonexistent. In other societies, data privacy is considered a fundamental right. Let's take a look around the world to examine some of the differences in data privacy around the globe.

The European Union (GDPR)

The 27-member EU serves as the bellwether for personal privacy in the developed world. The right to personal and data privacy is relatively heavily regulated and actively enforced in Europe, and it is enshrined into European law in many ways. In Article 8 of the European Convention on Human Rights (ECHR), a person has a right to a "private and family life, his home and his correspondence," with some exceptions. Some additional types of private data under the GDPR include information such as race or ethnic origin, political affiliations or opinions, religious or philosophical beliefs, and information regarding a person's sex life or sexual orientation.

In the European Union, PII covers both facts and opinions about an individual. Individuals are guaranteed certain privacy rights as data subjects. Significant areas of Chapter 3 of the GDPR (see `gdpr.eu/tag/chapter-3`) include the following on data subject privacy rights:

- The right to be informed
- The right of access
- The right to rectification
- The right to erasure (the so-called "right to be forgotten")
- The right to restrict processing
- The right to data portability
- The right to object
- Rights in relation to automated decision-making and profiling

Australia

Australian privacy law was revamped in 2014, with an increased focus on the international transfer of PII. (See `www.oaic.gov.au/privacy/australian-privacy-principles`.)

Current law in Australia, under the Australian Privacy Act, allows for the offshoring of data but requires the transferring entity (the data processor) to ensure that the receiver of the data holds and processes it in accordance with the principles of Australian privacy law. As discussed in the previous section, this is commonly achieved through contracts that require recipients to maintain or exceed privacy standards. An important consequence under Australian privacy law is that the entity transferring the data out of Australia remains responsible for any data breaches by or on behalf of the recipient entities, meaning significant potential liability for any company doing business in Australia under current rules.

The United States

Data privacy laws in the United States generally date back to fair information practice guidelines that were developed by the precursor to the Department of Health & Human Services (HHS). (See Ware, Willis H. (1973, August). "Records, Computers and the Rights of Citizens," Rand. Retrieved from `www.rand.org/content/dam/rand/pubs/papers/2008/P5077.pdf`.) These principles include the following concepts:

- For all data collected, there should be a stated purpose.
- Information collected from an individual cannot be disclosed to other organizations or individuals unless specifically authorized by law or by consent of the individual.
- Records kept on an individual should be accurate and up-to-date.
- There should be mechanisms for individuals to review data about them, to ensure accuracy. This may include periodic reporting.
- Data should be deleted when it is no longer needed for the stated purpose.
- Transmission of personal information to locations where "equivalent" personal data protection cannot be assured is prohibited.
- Some data is too sensitive to be collected unless there are extreme circumstances (such as sexual orientation or religion).

Perhaps the defining feature of U.S. data privacy law is its fragmentation. There is no overarching law regulating data protection in the United States. In fact, the word *privacy* is not included in the U.S. Constitution. However, there are now data privacy laws in each of the 50 states as well as U.S. territories.

There are few restrictions on the transfer of PII or PHI out of the United States, a fact that makes it relatively easy for companies to engage cloud providers and store data in other countries. The Federal Trade Commission (FTC) and other regulatory bodies do hold companies accountable to U.S. laws and regulations for data after it leaves the physical jurisdiction of the United States. U.S.-regulated companies are liable for the following:

- Personal data exported out of the United States
- Processing of personal data by subcontractors based overseas
- Projections of data by subcontractors when it leaves the United States

Several important international agreements and U.S. federal statutes deal with PII. The Privacy Shield agreement is a framework that regulates the transatlantic movement of PII for commercial purposes between the United States and the European Union. Federal laws worth review include HIPAA, GLBA, SOX, and the Stored Communications Act, all of which impact how the United States regulates privacy and data. At the state level, it is worth reviewing the California Consumer Protection Act (CCPA), the strongest state privacy law in the nation.

Privacy Shield

Unlike the GDPR, which is a set of regulations that affect companies doing business in the EU or with citizens of the EU, Privacy Shield is an international agreement between the United States and the European Union that allows the transfer of personal data from the European Economic Area (EEA) to the United States by US-based companies. (See `www.impact-advisors.com/security/eu-us-privacy-shield-framework/` and `iapp.org/media/pdf/resource_center/Comparison-of-Privacy-Shield-and-the-Controller-Processor-Model.pdf`.) This agreement replaced the previous safe harbor agreements, which were invalidated by the European court of justice in October 2015. The Privacy Shield agreement is currently facing legal challenge in European Union courts as of July 2020. Adherence to Privacy Shield does not make U.S. companies GDPR-compliant. Rather, it allows the company to transfer personal data out of the EEA into infrastructure hosted in the United States.

To transfer data from the EEA, U.S. organizations can self-certify to the Department of Commerce and publicly commit to comply with the seven principles of the agreement. Those seven principles are as follows:

- **Notice**: Organizations must publish privacy notices containing specific information about their participation in the Privacy Shield Framework; their privacy practices; and EU residents' data use, collection, and sharing with third parties.
- **Choice**: Organizations must provide a mechanism for individuals to opt out of having personal information disclosed to a third party or used for a different purpose than that for which it was provided. Opt-in consent is required for sharing sensitive information with a third party or its use for a new purpose.
- **Accountability for onward transfer**: Organizations must enter into contracts with third parties or agents who will process personal data for and on behalf of the organization, which require them to process or transfer personal data in a manner consistent with the Privacy Shield principles.
- **Security**: Organizations must take reasonable and appropriate measures to protect personal data from loss, misuse, unauthorized access, disclosure, alteration, and destruction, while accounting for risks involved and nature of the personal data.
- **Data integrity and purpose limitation**: Organizations must take reasonable steps to limit processing to the purposes for which it was collected and ensure that personal data is accurate, complete, and current.
- **Access**: Organizations must provide a method by which the data subjects can request access to and correct, amend, or delete information the organization holds about them.

- **Recourse, enforcement, and liability**: This principle addresses the recourse for individuals affected by noncompliance, consequences to organizations for noncompliance, and compliance verification.

The Health Insurance Portability and Accountability Act of 1996

The HIPAA legislation of 1996 defined what comprises personal health information, mandated national standards for electronic health record keeping, and established national identifiers for providers, insurers, and employers. Under HIPAA, PHI may be stored by cloud service providers provided that the data is protected in adequate ways.

The Gramm–Leach–Bliley Act (GLBA) of 1999

This U.S. federal law requires financial institutions to explain how they share and protect their customers' private information. GLBA is widely considered one of the most robust federal information privacy and security laws. This act consists of three main sections:

- The Financial Privacy Rule, which regulates the collection and disclosure of private financial information
- The Safeguards Rule, which stipulates that financial institutions must implement security programs to protect such information
- The Pretexting provisions, which prohibit the practice of pretexting (accessing private information using false pretenses)

The act also requires financial institutions to give customers written privacy notices that explain their information-sharing practices. GLBA explicitly identifies security measures such as access controls, encryption, segmentation of duties, monitoring, training, and testing of security controls.

The Stored Communication Act of 1986

The Stored Communication Act (SCA), as enacted as Title II of the Electronic Communication Privacy Act, created privacy protection for electronic communications (such as email or other digital communications) stored on the Internet. In many ways, this act extends the Fourth Amendment of the U.S. Constitution—the people's right, to be "secure in their persons, houses, papers, and effects, against unreasonable searches and seizures"—to the electronic landscape. It outlines that private data is protected from unauthorized access or interception (by private parties or the government).

The California Consumer Privacy Act of 2018

It may seem odd to talk about a state privacy law in the context of international privacy protection, but the California Consumer Privacy Act (CCPA) is at present the strongest

data privacy law in the United States. The act is intended to protect within state law the rights of a California citizen to the following:

- Know what personal data is being collected about them
- Know whether their personal data is sold or disclosed and to whom
- Opt out of the sale of personal data
- Access their personal data
- Request a business to delete any personal information about a consumer collected (the right to be forgotten)
- Be protected from discrimination (by service or price) for exercising their privacy rights

The CCPA outlines what entities need to comply with the law, the responsibility and accountability of the law's provisions, and the sanctions and fines that may be levied for violations of the law. Those fines can vary between $100 to $750 per consumer per data incident (or actual damages, whichever is greater), which makes the CCPA the most stringent privacy law in the United States.

Jurisdictional Differences in Data Privacy

As you have seen, this small review of different locations, jurisdictions, and legal requirements for the protection of private data demonstrates the challenge of maintaining compliance while using global cloud computing resources. Different laws and regulations may apply depending on the location of the data subject, the data collector, the cloud service provider, subcontractors processing data, and the company headquarters of any of the entities involved. A CCSP should always be aware of the issues in these areas and consult with legal professionals during the construction of any cloud-based services contract.

Legal concerns can prevent the utilization of a cloud services provider, add to costs and time to market, and significantly complicate the technical architectures required to deliver services. Nevertheless, it is vital to never replace compliance with convenience when evaluating services. In 2020, the video conferencing service Zoom was found to be engaged in the practice of routing video calls through servers in China in instances when no call participants were based there. This revelation caused an uproar throughout the user community and led to the abandonment of the platform by many customers out of privacy concerns. (See `www.theguardian.com/uk-news/2020/apr/24/uk-government-told-not-to-use-zoom-because-of-china-fears`.) In this example, customers abandoned the platform in large numbers because their data would certainly not enjoy the same privacy protections within China as would be expected in most liberal democracies.

Standard Privacy Requirements

With so many concerns and potential harm from privacy violations, entrusting data to any cloud provider can be daunting for many companies. Fortunately, there are industry standards in place that address the privacy aspects of cloud computing for customers. International organizations such as ISO/IEC have codified privacy controls for the cloud. ISO 27018 was published in July 2014, as a component of the ISO 27001 standard and was most recently updated in 2019. A CCSP can use the certification of ISO 27000 compliance as assurance of adherence to key privacy principles. Major cloud service providers such as Microsoft, Google, and Amazon maintain ISO 27000 compliance, which include the following key principles:

- **Consent**: Personal data obtained by a CSP may not be used for marketing purposes unless expressly permitted by the data subject. A customer should be permitted to use a service without requiring this consent.
- **Control**: Customers shall have explicit control of their own data and how that data is used by the CSP.
- **Transparency**: CSPs must inform customers of where their data resides and any subcontractors that might process personal data.
- **Communication**: Auditing should be in place, and any incidents should be communicated to customers.
- **Audit**: Companies must subject themselves to an independent audit on an annual basis.

Privacy requirements outlined by ISO 27018 enable cloud customers to trust their providers. A CCSP should look for ISO 27018 compliance in any potential provider of cloud services.

Generally Accepted Privacy Principles

Generally Accepted Privacy Principles (GAPP) consists of 10 principles for privacy from the American Institute of Certified Public Accountants (AICPA). (See `www.aicpa.org/content/dam/aicpa/interestareas/informationtechnology/resources/privacy/downloadabledocuments/10252-346-records-management-pro.pdf`.) The GAPP is a strong set of standards for the appropriate protection and management of personal data. The 10 main privacy principle groups consist of the following:

- **Management**: The organization defines, documents, communicates, and assigns accountability for its privacy policies and procedures.
- **Notice**: The organization provides notice of its privacy policies and procedures. The organization identifies the purposes for which personal information is collected, used, and retained.

- **Choice and consent**: The organization describes the choices available to the individual. The organization secures implicit or explicit consent regarding the collection, use, and disclosure of the personal data.
- **Collection**: Personal information is collected only for the purposes identified in the notice (see "Notice," earlier in the list).
- **Use, retention, and disposal**: The personal information is limited to the purposes identified in the notice the individual consented to. The organization retains the personal information only for as long as needed to fulfill the purposes or as required by law. After this period, the information is disposed of appropriately and permanently.
- **Access**: The organization provides individuals with access to their personal information for review or update.
- **Disclosure to third parties**: Personal information is disclosed to third parties only for the identified purposes and with implicit or explicit consent of the individual.
- **Security for privacy**: Personal information is protected against both physical and logical unauthorized access.
- **Quality**: The organization maintains accurate, complete, and relevant personal information that is necessary for the purposes identified.
- **Monitoring and enforcement**: The organization monitors compliance with its privacy policies and procedures. It also has procedures in place to address privacy-related complaints and disputes.

Standard Privacy Rights Under GDPR

In the EU, the GDPR codifies specific rights of the data subject that must be adhered to by any collector or processor of data. These rights are outlined in Chapter 3 of the GDPR (Rights of the Data Subject) and consist of 12 articles detailing those rights:

- Transparent information, communication, and modalities for the exercise of the rights of the data subject
- Information to be provided where personal data are collected from the data subject
- Information to be provided where personal data have not been obtained from the data subject
- Right of access by the data subject
- Right to rectification
- Right to erasure ("right to be forgotten")

- Right to restriction of processing
- Notification obligation regarding rectification or erasure of personal data or restriction of processing
- Right to data portability
- Right to object
- Automated individual decision-making, including profiling
- Restrictions

The complete language for the GDPR data subject rights can be found at `gdpr.eu/tag/chapter-3`.

UNDERSTANDING AUDIT PROCESS, METHODOLOGIES, AND REQUIRED ADAPTATIONS FOR A CLOUD ENVIRONMENT

The word *audit* can strike a nerve of terror in many IT professionals (and in the context of the IRS, most taxpayers as well). Even in simple IT architectures, the audit process can consist of rigorous, time-consuming efforts that must be followed exactly. Auditing in a cloud environment presents some additional challenges when compared to traditional on-premises requirements. This section will detail some of the controls, impacts, reports, and planning processes for a cloud environment.

It is important for cloud professionals to work in concert with other key areas of the business to successfully navigate the journey to and in cloud computing. Since the cloud can be utilized and affect so many departments within an organization, it is vital to coordinate efforts with legal counsel, compliance, finance, and executive leadership.

One key resource for navigating the challenges of cloud compliance that all CCSPs should be familiar with is the Cloud Controls Matrix (CCM) provided by the Cloud Security Alliance (CSA). As of this writing, version 4.0 of the CCM (updated August 2019) provides some of the latest guidance in the form of a "Rosetta Stone" mapping control domains to dozens of relevant compliance specifications, including HIPAA, GAPP, FERPA, FedRAMP, ISO 27001, ITAR, NIST, and many others. This free resource gives cloud practitioners a comprehensive list of control domains, best-practice guidance in control specifications, relevant architectures, delivery models, and supplier relationships, as well as direct mappings to each of the compliance specifications. CCSPs should make use of this resource whenever they're conducting an assessment of a cloud service provider or designing a compliant architecture in the cloud.

Internal and External Audit Controls

Internal audit and compliance have a key role in helping to manage and assess risk for both a cloud customer and provider. An organization may choose to keep internal audit functions within the company or outsource this function to a consultant or outside party.

Internal audit is an important line of defense for an organization. Many times, audits uncover issues with security or governance practices that can be corrected before more serious issues arise.

An internal auditor acts as a "trusted advisor" as an organization takes on new risks. In general, this role works with IT to offer a proactive approach with a balance of consultative and assurance services. An internal auditor can engage with relevant stakeholders to educate the customer to cloud computing risks (such as security, privacy, contractual clarity, Business Continuity Planning (BCP) and Disaster Recovery Planning (DRP) compliance with legal and jurisdictional issues). An internal audit also helps mitigate risk by examining cloud architectures to provide insights into an organization's cloud governance, data classifications, identity and access management effectiveness, regulatory compliance, privacy compliance, and cyber threats to a cloud environment.

It is desirable for an internal auditor to maintain independence from both the cloud customer and the cloud provider. The auditor is not "part of the team" but rather an independent entity who can provide facts without fear of reprisals or political blowback. An addition to the advisory role, an internal auditor will generally perform audits in the traditional sense, making sure the cloud systems meet contractual and regulatory requirements and that security and privacy efforts are effective.

Security controls may also be evaluated by external auditors. An external auditor definition does not work for the firm being audited. An external audit can be required by various pieces of financial legislation as well as information security standards such as ISO 27001. An external auditor is focused on ensuring compliance and therefore does not take on the role of a "trusted advisor" but rather more of regulator with punitive capability.

Impact of Audit Requirements

The requirement to conduct audits can have a large procedural and financial impact on a company. In a cloud computing context, the types of audit required are impacted largely by a company's business sector, the types of data being collected and processed, and the variety of laws and regulations that these business activities subject a company to. Some entities operate in heavily regulated industries subject to numerous auditing requirements (such as banks or chemical companies). Others may be brokers of data from international data subjects (such as big tech companies like Facebook, Google, and Microsoft).

Due to the changing nature of the cloud environment, auditors must rethink some traditional methods that have been used to collect evidence to support an audit. Consider the problems of data storage, virtualization, and dynamic failover.

- Is a data set representative of the entire user population? Cloud computing allows for the relatively easy distribution of data to take advantage of geographic proximity to end users, so obtaining a holistic representative sample may be difficult.
- Physical servers are relatively easy to identify and locate. Virtualization adds a layer of challenge in ensuring that an audited server is, in fact, the same system over time.
- Dynamic failover presents additional challenge to auditing operations for compliance to a specific jurisdiction.

Identity Assurance Challenges of Virtualization and Cloud

The cloud is fundamentally based in virtualization technologies. Abstracting the physical servers that power the cloud from the virtual servers that provide cloud services allows for the necessary dynamic environments that make cloud computing so powerful. Furthermore, the underlying virtualization technologies that power the cloud are changing rapidly. Even a seasoned systems administrator that has worked with VMware or Microsoft's Hyper-V may struggle with understanding the inherent complexity of mass scalable platforms such as AWS, Google, or Azure cloud.

Identity assurance in a computing context is the ability to determine with a level of certainty that the electronic credential representing an entity (whether that be a human or a server) can be trusted to actually belong to that person. As Peter Steiner famously captured in *The New Yorker*, "On the internet, nobody knows that you're a dog." Bots and cyber criminals make identity assurance for end users much more difficult. The cloud greatly increases the difficulty of assurance for machines.

Depending on the cloud architecture employed, a cloud security professional must now perform multiple layers of auditing (of both the hypervisor and the virtual machines themselves) to obtain assurance. It is vital for any cloud security to understand the architecture that a cloud provider is using for virtualization and ensure that both hypervisors and virtual host systems are hardened and up-to-date. Change logs are especially important in a cloud environment to create an audit trail as well as an alerting mechanism for identifying when systems may have been altered in inappropriate ways by accidental or intentional manners.

Types of Audit Reports

Any audit, whether internal or external, will produce a report focused either on the organization or on the organization's relationship with an outside entity or entities. In a

cloud relationship, oftentimes the ownership of security controls designed to reduce risk resides with a cloud service provider. An audit of the cloud service provider can identify if there are any gaps between what is contractually specified and what controls the provider has in place.

Service Organization Controls

The American Institute of CPAs (AICPA) provides a suite of offerings that may be used to report on services provided by a service organization that customers can use to address risks associated with an outsourced service. See Table 6.2.

TABLE 6.2 AICPA Service Organization Control Reports

REPORT	USERS	CONCERNS	DETAILS REQUIRED
SOC 1	User entities and the CPAs that audit their financial statements	Effect of the controls at the service organization on the user entities' financial statements	Systems, controls, and tests performed by the service auditor and results of tests.
SOC 2	Broad range of users that need detailed information and assurance about controls	Security, availability, and processing integrity of the systems the service organization uses to process users' data and the confidentiality and privacy of the information processed by these systems	Systems, controls, and tests performed by the service auditor and results of tests.
SOC 3	Broad range of users that need detailed information and assurance about controls but do not have the need for or knowledge to make effective use of SOC 2 reports	Security, availability, and processing integrity of the systems the service organization uses to process users' data and the confidentiality and privacy of the information processed by these systems	Referred to as a "Trust Services Report," SOC 3 reports are general used and can be freely distributed.

Table source: www.aicpa.org

The differences between the SOC reports are as follows:

- **Service Organizations Controls 1 (SOC 1)**: These reports deal mainly with financial controls and are intended to be used primarily by CPAs that audit an entity's financial statements.

- **SOC for Service Organizations: Trust Services Criteria (SOC 2)**: Report on Controls at a Service Organization Relevant to Security, Availability, Processing Integrity, Confidentiality, or Privacy. These reports are intended to meet the needs of a broad range of users that need detailed information and assurance about the controls at a service organization relevant to security, availability, and processing integrity of the systems the service organization uses to process users' data and the confidentiality and privacy of the information processed by these systems. These reports can play an important role in the following:
  - Organizational oversight
  - Vendor management programs
  - Internal corporate governance and risk management processes
  - Regulatory oversight

  There are two types of reports for these engagements:
  - **Type 1**: Report on the fairness of the presentation of management's description of the service organization's system and the suitability of the design of the controls to achieve the related control objectives included in the description as of a specified date.
  - **Type 2**: Report on the fairness of the presentation of management's description of the service organization's system and the suitability of the design and operating effectiveness of the controls to achieve the related control objectives included in the description throughout a specified period.
- **Service Organization Controls 3 (SOC 3)**: Similar to SOC 2, the SOC 3 report is designed to meet the needs of users who need the same types of knowledge offered by the SOC 2, but lack the detailed background required to make effective use of the SOC 2 report. In addition, SOC 3 reports are considered general use and can be freely distributed, as sensitive details have been removed and only general opinions and nonsensitive data are contained.

AICPA definitions of SOC controls can be found at the following locations:

- **SOC 1**: `www.aicpa.org/interestareas/frc/assuranceadvisoryservices/aicpasoc1report.html`
- **SOC 2**: `www.aicpa.org/interestareas/frc/assuranceadvisoryservices/aicpasoc2report.html`
- **SOC 3**: `www.aicpa.org/interestareas/frc/assuranceadvisoryservices/aicpasoc3report.html`

The Statement on Standards for Attestation Engagements (SSAE) is a set of standards defined by the AICPA to be used when creating SOC reports. The most current version

(SSAE 18) was made effective in May 2017, and added additional sections and controls to further enhance the content and quality of SOC reports.

The international counterpart of the AICPA, the International Auditing and Assurance Standards Board, issues the International Standard on Assurance Engagements (ISAE). There are a number of differences between the two standards, and a security professional should always consult the relevant business departments to determine which audit report(s) will be used when assessing cloud systems.

Cloud Security Alliance

The Cloud Security Alliance (CSA) offers a Security Trust Assurance and Risk (STAR) certification that can be used by cloud service providers, cloud customers, or auditors and consultants to ensure compliance to a desired level of assurance. (See `downloads.cloudsecurityalliance.org/assets/research/security-guidance/security-guidance-v4-FINAL.pdf.`) STAR consists of three levels of certification:

- **Level 1**: Self-Assessment is a complimentary offering that documents the security controls provided by various cloud computing offerings, helping users assess the security of cloud providers they currently use or are considering using.
- **Level 2**: Attestation is a collaboration between CSA and the AICPA to provide guidelines for CPAs to conduct SOC 2 engagements using criteria from the AICPA (Trust Service Principles, AT 101) and the CSA Cloud Controls Matrix. STAR Attestation provides for rigorous third-party independent assessments of cloud providers.
- **Level 3**: Continuous monitoring through automated processes that ensure security controls are monitored and validated at all times.

Restrictions of Audit Scope Statements

Audits are a fact of life to ensure compliance and prevent sometimes catastrophic effects of security failures. Though no employee relishes the thought of undergoing an audit, they are somewhat of a necessary evil to ensure that risks are adequately controlled and compliance is met. Any audit must be appropriately scoped to ensure that both the client and the auditor understand what is in the audit space and what is not. This is almost always driven by the type of report being prepared. An audit scope statement would generally include the following:

- Statement of purpose and objectives
- Scope of audit and explicit exclusions
- Type of audit
- Security assessment requirements

- Assessment criteria and rating scales
- Criteria for acceptance
- Expected deliverables
- Classification (for example, secret, top secret, public, etc.)

Any audit must have parameters set to ensure the efforts are focused on relevant areas that can be effectively audited. Setting these parameters for an audit is commonly known as the *audit scope restrictions*. Why limit the scope of an audit? Audits are expensive endeavors that can engage highly trained (and highly paid) content experts. In addition, the auditing of systems can affect system performance and, in some cases, require the downtime of production systems.

Scope restrictions can be of particular importance to security professionals. They can spell out the operational components of an audit, such as the acceptable times and time periods (for example, days and hours of the week) as well as the types of testing that can be conducted against which systems. Carefully crafting scope restrictions can ensure that production systems are not impacted adversely by an auditor's activity. Additionally, audit scope statements can be used to clearly define what systems are subject to audit. For example, an audit on HIPAA compliance would only include systems that process and store PHI.

Gap Analysis

As a precursor to a formal audit process, a company may engage in a gap analysis to help identify potential trouble areas to focus on. A gap analysis is the comparison of a company's practices against a specified framework and identifies any "gaps" between the two. These types of analysis are almost always done by external parties or third parties to ensure that they are performed objectively without potential bias. Some industry compliance standards (such as HIPAA or PCI) may require an outside entity to perform such an analysis. The reasoning is that an outside consultant can often catch gaps that would not be obvious to personnel working in the area every day.

If a gap analysis is being performed against a business function, the first step is to identify a relevant industry-standard framework to compare business activities against. In information security, this usually means comparison against a standard such as ISO/IEC 27002 (best-practice recommendations for information security management). Another common comparison framework used as a cybersecurity benchmark is the NIST cybersecurity framework.

A gap analysis can be conducted against almost any business function, from strategy and staffing to information security. The common steps generally consist of the following:

- Establishing the need for the analysis and gaining management support for the efforts.
- Defining scope, objectives, and relevant frameworks.

- Identifying the current state of the department or area (which involves the evaluation of the department to understand current state, generally involving research and interviews of employees).
- Reviewing evidence and supporting documentation, including the verification of statements and data.
- Identifying the "gaps" between the framework and reality. This highlights the risks to the organization.
- Preparing a report detailing the findings and getting sign-off from the appropriate company leaders.

Since a gap analysis provides measurable deficiencies and, in some cases, needs to be signed off by senior leadership, it can be a powerful tool for an organization to identify weaknesses in their efforts for compliance.

Audit Planning

Any audit, whether related to financial reporting, compliance, or cloud computing, must be carefully planned and organized to ensure that the results of the audit are relevant to the objectives. Audits genevrally consist of four phases, outlined in Figure 6.1.

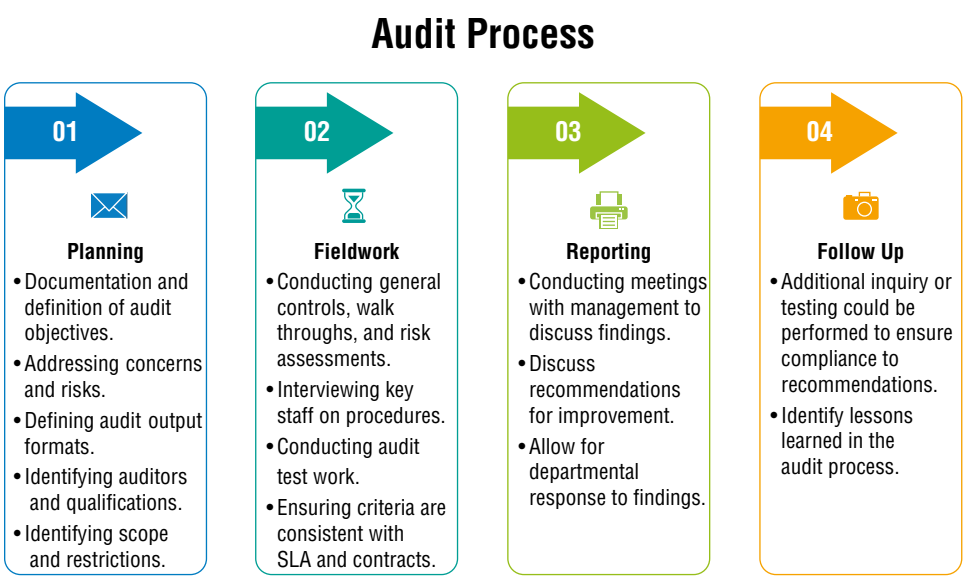

FIGURE 6.1 **Four phases of an audit**

The four phases of an audit consist of the following:

- **Audit planning**
  - Documentation and definition of audit objectives. This process is a collaborative effort to define what standards are being measured.

- Addressing concerns and risks.
- Defining audit output formats.
- Identifying auditors and qualifications.
- Identifying scope and restrictions.

- **Audit fieldwork**
  - Conducting general controls walk-throughs and risk assessments.
  - Interviewing key staff on procedures.
  - Conducting audit test work, which may include physical and architectural assessments with software tools.
  - Ensuring criteria are consistent with SLA and contracts.
- **Audit reporting**
  - Conducting meetings with management to discuss findings.
  - Discussing recommendations for improvement.
  - Allowing for departmental response to findings.
- **Audit follow-up**
  - Additional inquiry or testing could be performed to ensure compliance to recommendations.
  - Identify lessons learned in the audit process.

In many organizations, audit is a continuous process. This is often structured into business activities to provide an ongoing view into how an organization is meeting compliance and regulatory goals. As a cloud security professional, audit is a powerful tool to ensure that your organization is not in the news for the wrong reasons.

Internal Information Security Management Systems

An information security management system (ISMS) is a systematic approach to information security consisting of processes, technology, and people designed to help protect and manage an organization's information. These systems are designed to strengthen the three aspects of the CIA triad: confidentiality, availability, and integrity. An ISMS is a powerful risk management tool in place at most medium and large organizations, and having one gives stakeholders additional confidence in the security measures in place at the company.

The most effective ISMSs are aligned with standards, most importantly ISO 27001, the international standard that provides the specifications for a best-practice ISMS and details the compliance requirements.

Benefits of an ISMS

Though the function of an ISMS can vary from industry to industry, there are a number of benefits to implementation that hold true across all industries.

- **Security of data in multiple forms**: An ISMS will help protect data in all forms, whether paper, digital, or cloud based.
- **Cyber attacks**: Having an ISMS will make an organization more resilient to cyber attacks by having best practices in security in place throughout the operation.
- **Central information management**: An ISMS will put in place centralized frameworks for managing information, reducing shadow systems, and easing the burden of data protection.
- **Risk management**: Having a codified set of processes and procedures in place will reduce operational risks in a number of areas, including information security, business continuity, and adapting to evolving security threats.

As with any major organizational element, an ISMS requires buy-in from company leadership to be effective. In the cloud computing environment, an ISMS is extremely valuable in requiring detailed documentation on cloud provider environments and change logging.

Internal Information Security Controls System

As we have detailed a number of times in this section, the International Standards Organization (ISO) provides an invaluable resource to security professionals in its ISO 27001:2013 information security standards. (See `www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en`.) Cloud security professionals should be intimately familiar with ISO 27001 controls, which can be mapped to address requirements for audit, ISMS, or other compliance measures. The 14 control sets of Annex A of ISO 27001 are as follows:

- **A.5 Information security policies**: How policies are written and reviewed
- **A.6 Organization of information security**: The assignment of responsibilities for specific tasks
- **A.7 Human resource security**: Ensuring that employees understand their responsibilities prior to employment and once they've left or changed roles
- **A.8 Asset management**: Identifying information assets and defining appropriate protection responsibilities
- **A.9 Access control**: Ensuring that employees can only view information that's relevant to their job role
- **A.10 Cryptography**: The encryption and key management of sensitive information
- **A.11 Physical and environmental security**: Securing the organization's premises and equipment

- **A.12 Operations security**: Ensuring that information processing facilities are secure
- **A.13 Communications security**: How to protect information in networks
- **A.14 System acquisition, development and maintenance**: Ensuring that information security is a central part of the organization's systems
- **A.15 Supplier relationships**: The agreements to include in contracts with third parties and how to measure whether those agreements are being kept
- **A.16 Information security incident management**: How to report disruptions and breaches and who is responsible for certain activities
- **A.17 Information security aspects of business continuity management**: How to address business disruptions
- **A.18 Compliance**: How to identify the laws and regulations that apply to your organization

In addition, cloud security professionals in the United States should be familiar with the National Institute of Standards (NIST) cybersecurity framework (v1.1). This framework identifies controls in these five broad areas:

- Identify
- Protect
- Detect
- Respond
- Recover

A common mnemonic used to remember the NIST CSF control areas is "In Public, Drink Reasonably Responsibly." See the following site for complete details: `www.nist.gov/cyberframework/online-learning/five-functions`.

Policies

Policies are a key part of any data security strategy. Without a policy in place, it is impossible to enforce requirements in any systematic way. Policies ensure that employees and management are aware of what is expected of them and where their responsibilities lie. Policies are an important piece of standardizing practices in an organization.

From a cloud computing perspective, policies can be an important tool in preventing isolation of activities (for treating data stored in one cloud provider in a completely different way than others). Policies also provide a way for leadership to specify appropriate uses of cloud services to ensure the organization realizes business benefits without also taking on additional business risk.

Organizational Policies

Companies use policies and procedures to outline rules as well as outline courses of action to deal with problems. Organizations use policies to make employees understand the organization's views and values on specific issues, and what actions will occur if they are not followed. At the organizational level, policies are used to reduce the chances of the following:

- Financial losses
- Loss of data
- Reputational damage
- Statutory and regulatory compliance issues
- Abuse or misuse of computing systems and resources

Functional Policies

A functional policy is a set of standardized definitions for employees that describe how they are to make use of systems or data. Functional policies typically guide specific activities crucial to the organization, such as appropriate handling of data, vulnerability management, and so on. These are generally codified in an operation ISMS and can consist of the following (not an exhaustive list):

- **Data classification policies**: Identifies types of data and how each should be handled
- **Network services policies**: How issues such as remote access and network security are handled
- **Vulnerability scanning policies**: Routines and limitations on internal scanning and penetration testing
- **Patch management policies**: How equipment is patched on what schedule
- **Acceptable use policies**: What is and is not acceptable to do on company hardware and networks
- **Email use policies**: What is and is not acceptable to do on company email accounts
- **Password policies**: Password complexity, expiration, reuse
- **Incident response policies**: How incidents are handled

Cloud Computing Policies

Employee provisioned cloud services may present significant data management risks and may be subject to changes in risk with or without notice. Almost all cloud services require

individual users to accept click-through agreements (which are rarely if ever read carefully). Agreements do not allow users to negotiate or clarify terms, often provide vague descriptions of services and safeguards, and often are subject to change without notice.

Cloud services should not be exempt from organizational policy application. These policies will define the requirements that users must adhere to in order to make use of the services. Because of the ease of provisioning cloud services, many organizations have specific policies in place that discourage or prohibit the use of cloud services by individuals outside of central IT oversight.

When pursuing cloud services, a security professional should pay close attention to the following:

- **Password policies**: If an organization has password policies around length, complexity, expiration, or multifactor authentication (MFA), it is important to ensure that these same requirements are met by a cloud service provider.
- **Remote access**: The same bar must be met for remote access to cloud providers as on-premises services. This may include items such as up-to-date patching, hard disk encryption, or MFA.
- **Encryption**: Policies about encryption strength and when encryption is required. Key escrow can be an important aspect of policy to focus on (for example who has the decryption keys?).
- **Data backup and failover**: Policies on data retention and backup must be enforced on cloud providers. If policies exist on data location for backups and redundancy, they must also be enforced on CSPs.
- **Third-party access**: What third parties might have access to data stored with the CSP? Can this access be logged and audited?
- **Separation of duties**: Can controls for the separation of key roles be enforced and maintained by the cloud provider?
- **Incident response**: What are the required steps in a response, including who is contacted for a variety of incidents?

In some instances, a cloud service provider cannot meet a company's requirements when it comes to adhering to a specific policy. If this happens, it is important to consider the risks of using the provider, and any deviations from policy should be carefully documented.

Identification and Involvement of Relevant Stakeholders

One key challenge in the provisioning of cloud computing resources is the identification of all company stakeholders that need to be involved in the decision process. Oftentimes,

this process is complicated by the need to fully understand the business processes that will be taking advantage of cloud computing services. This is not an easy task and requires visibility on what services are provided, how they are delivered, and what the current architecture looks like. This generally means that there will be stakeholders outside of IT, which necessitates coordination across multiple business functions. Once stakeholders are identified, it is important to document who should be involved in the decision process, because omitting key stakeholders can lead to a haphazard approach to the implementation of cloud services.

Stakeholder Identification Challenges

To identify stakeholders, some key challenges that a cloud security professional may face include the following:

- Defining the enterprise architecture currently used to delivery services.
- Evaluating potential cloud options and solutions objectively and without existing bias.
- Selecting the appropriate service(s) and cloud provider.
- Identifying what users may be affected by the service offerings and planning change management. This may particularly difficult if roles are being eliminated by a shift to cloud computing.
- Identifying both the direct costs (costs paid to CSP or third parties) and the indirect costs (training, loss of productivity due to learning a new system, etc.).
- Identifying appropriate parties to engage on risk management.

Governance Challenges

In addition to identifying stakeholders, cloud security professionals should be prepared to deal with governance challenges when moving to a cloud environment, including the following:

- Audit requirements and how to accomplish them in the cloud
- Regulatory and legal obligations
- Reporting and communication lines within the organization and with the cloud service provider
- Documentation of all operational procedure and process changes and where this documentation will be safely stored
- Review of all disaster recovery and business continuity planning documents to ensure updates for new cloud architecture

The benefits of the cloud can quickly and easily move beyond just the financial for many companies. The prevalence of distributed workforces, reliability and scalability, worldwide delivery of services, and backup and recovery possibilities are just a few benefits that have driven cloud adoption across countless industries. For those reasons, the emphasis and appetite to move to cloud computing may come from many different areas within a company. It is important to coordinate communication across the many departments that may be affected by cloud projects.

Specialized Compliance Requirements for Highly Regulated Industries

As we have detailed numerous times in this reference book, the responsibility for compliance to any relevant regulations ultimately rests with the company (not the cloud service provider). This is true for companies that fall under additional regulation (such as HIPAA for healthcare providers, PCI for credit card processors, GLBA for public financial companies, and North American Electric Reliability Corporation Critical Infrastructure Protection [NERC/CIP] for critical infrastructure providers such as power or water utilities). In some cases, these specialized requirements will make leveraging the cloud more difficult. Companies should always identify any geographic or jurisdiction-specific regulations (such as HIPAA, which might require PHI to be physically kept within the United States) and factor them into any cloud computing evaluation. Additionally, special scrutiny should be applied to cloud service providers when regulations demand levels of reliability above and beyond typical industry standards.

Impact of Distributed Information Technology Models

Distributed information technology (IT) models are become increasingly common as interconnected services are more easily provided from global suppliers. One example of a distributed model is the outsourcing of services such as network operation centers to third parties across the globe. In many ways, cloud computing makes distributed IT much simpler, as the cloud can provide services to IT across countries and continents without costly on-premises data center connections and associated hardware and personnel.

A cloud security professional should be aware of common issues caused by distributed IT models and how to mitigate them.

Communications

In a traditional IT model, teams are generally familiar with what roles and functions are filled by different personnel. Depending on the company, they may be on a first-name basis (and might play on the company softball team together). Distributed IT means that these roles and functions may be separated geographically and all correspondence takes place via technology.

In some cases, less formal processes (for example, "I call Jim in networks when I need to have a Client Access License [CAL] added to the system") move to more structured interactions using standardized processes (for example, "I submitted a ticket in service now for a new CAL to be provisioned").

From a security perspective, this is considered an improvement, despite the potential business impacts caused by increased effort to get IT functions accomplished quickly. Using formalized ticketing systems and processes means that changes are far more likely to be recorded, adding to the overall adherence to change management controls.

Coordination of Activities

Technology project management is a key component to the success of any IT department. Successful PMs oversee the implementation of hardware and software projects using project management methods and expertise that is a key skillset within any company, especially in an IT organization.

The specialized nature of distributed IT models (and cloud services offerings) makes it less likely that an organization will have personnel in house with the specialized knowledge required to successfully implement projects. For this reason, it is important to consider bringing in outside experts for implementation projects. This common practice gives the distinct advantage of having subject-matter experts involved in the rollout or deployment of a new service or offering. In addition, this relieves some of the burden on an organization's IT team in learning, coordinating, and provisioning new systems that they may be completely unfamiliar with. Consultant agreements generally allow a provider to deliver service to a customer's requirements with sufficient accountability and oversight from company representatives.

Governance of Activities

Effective governance can require additional steps in a distributed IT environment. When IT functions are dispersed across providers, countries, or continents, additional efforts may be needed to pull together information for program management, risk management, audit, compliance, and legal functions. Additionally, the distribution of these IT functions across jurisdictions where different statutory and regulatory issues may apply may increase the overall efforts needed to ensure effective governance.

Selecting IT service providers that are effective and responsive in communication is key to establishing repeatable governance models. Many vendors specialize in the automation of information collection for governance purposes, which can help streamline the collection of information from distributed providers.

Distributed IT models mean that effective governance will require the collection of information from multiple sources. Coordinating those efforts should include the up-front identification of how these processes will be managed to achieve the common goals of the organization.

UNDERSTAND IMPLICATIONS OF CLOUD TO ENTERPRISE RISK MANAGEMENT

If you compare how IT services were provisioned two decades ago to how they are done today, you would see a completely different landscape. In the dot-com era, provisioning systems took experts days or weeks to build out hardware, operating systems, and applications. Companies spent millions of dollars on physical hardware to support their computing efforts, wide area networks, and data centers. Today, anyone with a credit card (or even without one on free services) can provision a multitier web application in a few minutes. Services are spun up in seconds in some cases.

This shift in how IT services is provisioned is not insignificant to organizational risk. In the past, the thought of "accidentally" provisioning a server in another country or legal jurisdiction was unthinkable (and most likely impossible). Today, it takes a few errant clicks from a cloud administrator to expose an organization to risks unthinkable two decades ago.

It is vital that both the cloud customer and the cloud service provider understand risk and what strategies can be employed for risk mitigation. Risk must be accounted for in any migration to a cloud-based environment, and in some cases, the methods that a company has traditionally used to manage risk may need to evolve to reflect the realities of the cloud. In any case, it is vital for the cloud customer and the cloud provider to be aligned in policies and procedures as closely as possible in order to share the burden of risk management between them.

Assess Providers Risk Management Programs

Prior to establishing a relationship with a cloud provider, a cloud customer needs to analyze the risks associated with the adoption of a cloud-based solution for a particular system and plan a strategy to mitigate and control those risks. There are several steps that a cloud customer should perform to assess the risk management capabilities (controls, methodologies, and practices) of any cloud service provider (see Figure 6.2). Those steps are outlined in NIST special publication 800-37 Revision 2 (`csrc.nist.gov/publications/detail/sp/800-37/rev-2/final`) and include the following:

- Performing a risk assessment
- Identifying the best-fitting cloud architecture
- Selecting the most suitable cloud service
- Gaining necessary visibility into the cloud offering
- Defining and negotiating necessary risk and risk control mitigations before finalizing the SLA and proceeding with the cloud migration

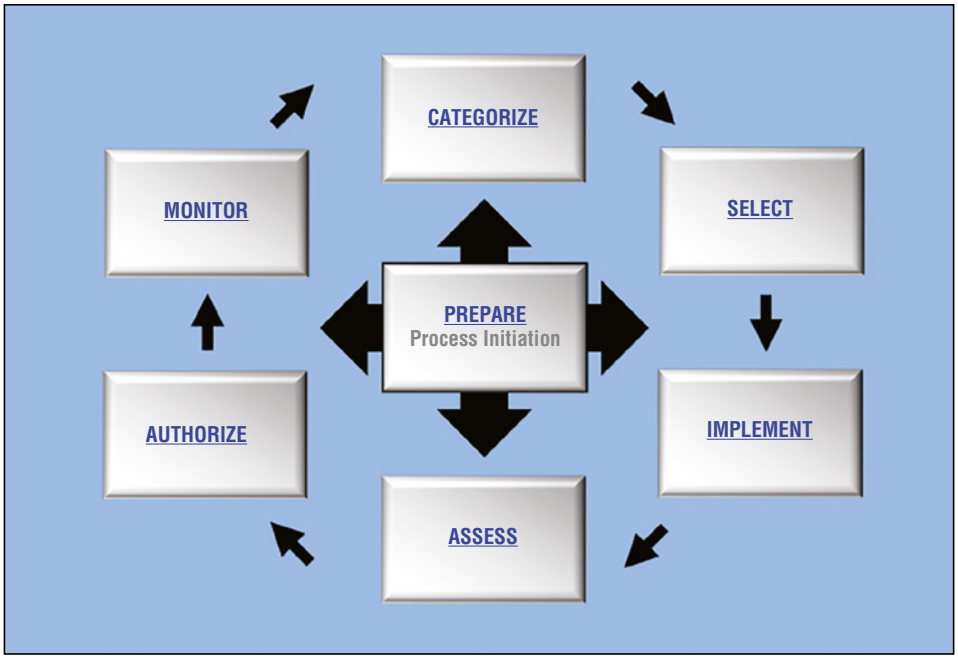

FIGURE 6.2 **NIST 800-37 rev 2 Risk Management Framework**

Cloud service providers by definition provide services to a wide variety of customers and industries. The architectures, service offerings, and security and privacy controls are therefore generally configured to meet the requirements of a large number of customers. Certification programs such as the Cloud Security Alliance's STAR program can give cloud customers a baseline on which to assess a given provider's controls, methodologies, and practices. Additionally, a cloud customer should apply a risk management framework as a standardized approach when vetting a cloud service. Several methods can be employed to assess a provider's risk management program. These can include reviewing the CSP's audit reports or conducting an independent risk assessment as an outside party.

A key concept for cloud security professionals to understand is the shared responsibility model of cloud computing. Depending on the delivery model of services, a cloud service provider takes responsibility for "security of the cloud"—protecting the infrastructure of the hardware, software, networking, and facilities that run the cloud services. The cloud customer takes responsibility for "security in the cloud," which varies depending on the cloud services and architectures implemented. The level of responsibility (and therefore risk) assumed by the customer can vary widely. When a company is evaluating a CSP for potential service, the methods used for evaluation are generally dictated by the company's risk profile and risk appetite.

Risk Profile

A company's risk profile generally comprises two factors. The first is a quantitative measure of the types of threats that an organization faces. This can vary widely by company and business sector, as a financial company can face very different risks than a landscaping provider. The second factor is the company's willingness to accept risk (risk tolerance). In mature companies, this can be represented as a numerical value to help quantify the calculations. Combining these with potential costs associated with the occurrence of each risk can provide a nonsubjective profile of risk.

A risk profile can help a company identify the level of risks that it is willing to accept. This profile is developed collaboratively with numerous stakeholders throughout the organization, such as business leaders, data owners, enterprise risk management, internal and external audit, legal, and compliance officers.

Risk Appetite

Risk appetite is just what it sounds like: the appetite a company has to accept risk. Every company is different in this arena, and it is directly affected by regulations and statutory requirements in a particular industry. The risk appetite of a company may be significantly larger if their success depends largely on time-to-market. They may be willing to accept larger risks in the cloud computing arena to achieve scalability or utilize cloud technologies to reach consumers in more geographic areas.

Differences Between Data Owner/Controller vs. Data Custodian/Processor

An important distinction in data is the difference between the data owner (data controller) and the data custodian (data processor). Let's start with some definitions:

- The data subject is the individual or entity that is the subject of the personal data.
- The data controller is the person (or company) that determines the purposes for which, and the way in which, personal data is processed.
- The data processor is anyone who processes personal data on behalf of the data controller.

For example, let's say a company that sells avocados takes personal data online to fulfill orders. They use cloud service provider AWS to host their website, and online payment vendor PayPal to process their transactions. In this case, any customer is the data subject, the avocado company is the data controller, and both AWS and PayPal are data processors (as they both process personal data on behalf of the avocado company).

The distinctions are important for regulatory and legal reasons. Data processors are responsible for the safe and private custody, transport, and storage of data according to

business agreements. Data owners are legally responsible (and liable) for the safety and privacy of the data under most international laws.

Regulatory Transparency Requirements

A cloud security professional should be aware of the transparency requirements imposed on data controllers by various regulations in the United States and internationally. The following is a short (and inexhaustive) description of some of the transparency requirements companies should consider as data owners.

Breach Notifications

As detailed earlier in this section, notification of data breaches is required by all 50 states, several federal statutes, and numerous international laws. In practice, nearly every human on the planet is familiar with these laws because they have been notified by numerous companies about the loss of their private data. In all cases (including cloud computing environments), the data controller is responsible under the law for notifying the appropriate authorities as well as the data subject of the loss of their personal data. Many states and international laws have timely notification requirements that companies must fulfill when personal data is lost.

Sarbanes–Oxley Act

If a company is publicly traded in the United States, they are subject to transparency requirements under the Sarbanes-Oxley Act (SOX) of 2002. Specifically, as data owners, these companies should consider the following:

- **Section 802**: It is a crime to destroy, change, or hide documents to prevent their use in official legal processes.
- **Section 804**: Companies must keep audit-related records for a minimum of five years.

SOX compliance is often an issue with both data breaches and ransomware incidents at publicly traded companies. The loss of data related to compliance due to external actors does not protect a company from legal obligations.

GDPR and Transparency

For companies doing business in the European Union or with citizens of the EU, transparency requirements under the GDPR are laid out in Article 12 (see `gdpr-info.eu/art-12-gdpr`). The exact language states that a data controller (data owner) "must be able to demonstrate that personal data are processed in a manner transparent to the data subject." The obligations for transparency begin at the data collection stage and apply "throughout the life cycle of processing."

The GDPR stipulates that communication to data subjects must be "concise, transparent, intelligible and easily accessible, and use clear and plain language." The methods of communication are also clearly defined. Information must be provided to the data subjects "in writing, or by electronic means" or orally upon request. All of this information must be provided free of charge and not be conditional upon the purchase of a service or the surrender of information.

In practical terms, this means that plain language must be used to explain why data is being collected and what it is being used for. Similar language is included in some current (CCPA) and forthcoming state regulations.

Risk Treatment

According to its definition, risk treatment is "the process of selecting and implementing of measures to modify risk." Risk treatment measures can include avoiding, modifying, sharing, or retaining risk. The risk treatment measures can be selected out of security controls that are used within a company's information security management system (ISMS).

Mitigation of risks will reduce either the potential exposure to a risk or the impact of a risk. If a risk is avoided, an organisation outside the EU may choose to stop, postpone, or cancel activities that present that risk. For example, a company can avoid the risk of a GDPR prosecution by not processing personal data of subjects who are in the EU. If a risk is modified, a company may work to reduce the effect that risk would have on a company (for example, defense-in-depth strategies to protect against data loss) or the likelihood of it occurring. Sharing a risk is the classic use case of insurance. By purchasing cyber-breach insurance, a company is sharing the risk with all of the customers facing the same risk. Retaining a risk is accepting the risk as part of the process of doing business.

Many of the strategies detailed in this chapter (such as controls, audits, compliance, etc.) are at their core about risk treatment. Cloud security, like all information security, is essentially about managing risk. In no scenario will risk treatment reduce a company's risks to zero. There is inherent risk in any business activity, and the risks retained by a company are referred to as *residual risk*.

Risk Frameworks

There are a number of risk frameworks that a company may use to help manage risk in their organization. Risk management is a business process that can affect all aspects of an enterprise. In the cloud computing arena, a cloud security professional should be familiar with the ISO 31000:2018 guidance standard, the European Network and Information Security Agency (ENISA)'s cloud computing risk assessment tool, and NIST standards such as 800-146 (cloud computing synopsis and recommendation) and 800-37 (risk management framework for information systems).

ISO 31000:2018

ISO 31000, "Risk management - Guidelines," was first published in 2009 and majorly revised in 2018. (See `www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en`.) This ISO standard provides generic recommendations for the design, implementation, and review of risk management processes within an organization. The 2018 update provides more strategic guidance as well as redefines risk from the concept of a "probability of loss" to a more holistic view of risk as the "effect of uncertainty of objectives," recasting risk as either a negative or positive effect. ISO 31000 recommends the following steps in planning for risk:

- Avoiding the risk by deciding not to start or continue with the activity that gives rise to the risk
- Accepting or increasing the risk to pursue an opportunity
- Removing the risk source
- Changing the likelihood
- Changing the consequences
- Sharing the risk with another party or parties (including contracts and risk financing)
- Retaining the risk by informed decision

ISO 31000 is a detailed framework, but is not designed to be used in certification (there is no such thing as "ISO 31000 certified"). Adopting this framework will require extensive management conformity to accountability standards as well as strategic policy implementation, communication, and review practices.

European Network and Information Security Agency

European Network and Information Security Agency (ENISA) produces a number of useful resources, and the "Cloud Computing Risk Assessment (2009)" is one of them (`www.enisa.europa.eu/publications/cloud-computing-risk-assessment`). This guide identifies 35 types of risk for companies to consider in the areas of policy and organizational, technical, legal, and risks not specific to the cloud. In addition, ENISA spells out 53 types of vulnerabilities that companies should be aware of and a top-eight security risk list based on the impact and likeliness of occurrence. This free resource is certainly worth review by any cloud security professional. The top eight security risks include the following:

- **Loss of governance**: Gaps in the security defenses caused by differences in the understanding of responsibility between the client and the CSP.
- **Lock-in**: The difficulty in leaving a CSP.

- **Isolation failure**: The potential failures caused by lack of separation in storage, memory, and other hardware between cloud clients.
- **Compliance risk**: The CSP provides a new challenge to achieving certification.
- **Management interface compromise**: Management interfaces for cloud environments provide an additional attack vector.
- **Data protection**: How CSPs handle data in a lawful way.
- **Insecure data deletion**: Secure deletion of the cloud is complicated by its distributed nature.
- **Malicious insiders**: Addition of a CSP adds high-risk access individuals who can comprise cloud architectures and data.

National Institute of Standards and Technology (NIST)

National Institute of Standards and Technology (NIST) makes the list of excellent standards once again with 800-146, "Cloud Computing Synopsis and Recommendations" (see `csrc.nist.gov/publications/detail/sp/800-146/final`). This risk framework was published in 2012, and is a comprehensive guide to the benefits and risks associated with SaaS, PaaS, and IaaS cloud environments. The guide also makes recommendations on management, data governance, security and reliability, and virtual machines, as well as software and applications. Another handy reference (albeit written specifically for U.S. federal government) is NIST 800-37, a generalized risk management framework for information systems that has well documented methods for controlling risk in IT environments.

Metrics for Risk Management

There are some key cybersecurity metrics that companies can track to present measurable data to company stakeholders. An incomplete list of some of these metrics is as follows:

- **Patching levels**: How many devices on your cloud networks are fully patched and up-to-date?
- **Intrusion attempts**: How many times have unknown actors tried to breach your cloud systems?
- **Mean time to detect (MTTD)**: How long does it take for security teams to become aware of a potential security incident?
- **Mean time to contain (MTTC)**: How long does it take to contain identified attack vectors?
- **Mean time to resolve (MTTR)**: How long until security threats are definitively dealt with?

- **Days to patch**: How long does it take a security team to fully patch a cloud system?
- **Access management**: How many users have administrative access?

A good method for communicating risks in a way that is easy to digest and understand is the risk register. As you have seen, a risk can be quantitatively expressed by multiplying a threat likelihood by the magnitude of impact.

RISK = Threat Likelihood × Magnitude of Impact

A risk register is a critical part of any incident response plan and provides a clear, visual way to see risks and their potential impacts. A risk register may contain the following (see Figure 6.3):

- A description of the risk
- The expected impact if the associated event occurs
- The probability of the event occurring
- Steps to mitigate the risk
- Steps to take should the event occur
- Rank of the risk

Risk Assessment

Risk Number	Risk Status	Data Identified	Risk Description	Impact (1–4)	Probability (1–4)	Risk Factor (I × P)	Mitigation Strategy/Status	Owner	Next Review or Expected Mitigation Date
1	Retired	22-Jul-15	Security compromise becomes public before July 31st—before mitigation is finalized and verified with Microsoft and Mandiant.	4	3	12	Whack-a-mole and monitor very closely. Bring in Mandiant and Microsoft ASAP. (We are not ready to scramble to remediate today!)		13-Aug-15
2	Retired	22-Jul-15	Security compromise becomes public after July 31st—after mitigation finalized but before planned communication date/time.	1	2	2	Go dark and have an intense review before beginning to execute communication plan. (Each day beyond August 3rd, situation should gradually improve and risk will diminish.)		14-Aug-15

FIGURE 6.3 **Risk register**

Risk is not an exact science, and calculating/categorizing risk is a difficult exercise that must be revisited often. A risk register can be an important tool for decision-makers to have a clear picture of what risks can be avoided, modified, shared, or retained.

Assessment of Risk Environment

As the cloud becomes more and more of a critical operating component for many companies, identifying and understanding the risks posed by a cloud service provider is vital

for any company that's using cloud services. Cloud providers are subject to change of business models, acquisition, and even bankruptcy and shutdown. It is important to consider a number of questions when considering a cloud service, vendor, or infrastructure provider.

- Is the provider subject to takeover or acquisition?
- How financially stable is the provider?
- In what legal jurisdiction(s) are the provider's offices located?
- Are there outstanding lawsuits against the provider?
- What pricing protections are in place for services contracted?
- How will a provider satisfy any regulatory or legal compliance requirements?
- What does failover, backup, and recovery look like for the provider?

And on and on. It can be a daunting challenge for any cloud customer to perform due diligence on their provider. But because the company maintains all legal and regulatory obligations as the data controller, it is an important part of any vendor selection. Fortunately, there are some industry certifications than can help companies make informed selections and mitigate risk.

ISO 15408-1:2009: The Common Criteria

The Common Criteria for Information Technology Security Evaluation is an international standard for information security certification. The evaluation process is designed to establish a level of confidence in a product or platform's security features through a quality assurance process. This is done through testing laboratories where the product of platform is evaluated.

Most cloud service providers do not have common criteria evaluations, but many cloud-based products do. One common example relevant to CCSPs are security tools designed to be deployed in virtual environments.

Does Common Criteria evaluation mean that a product is secure? Not necessarily. Though Windows operating systems have received certification under the program, they have later been found to have major security flaws. In addition, CC does not include any study of the administrative or business processes of a company that produced a product or platform; it is merely an evaluation of the technical merits of the platform itself. An up-to-date list of certified products can be found at `www.commoncriteriaportal.org/products`.

Cloud Security Alliance (CSA) STAR

It is worth revisiting the Security, Trust, and Assurance Registry (STAR) provided by the CSA (`cloudsecurityalliance.org/star/levels`) as a valuable tool in displaying transparency and assurance for cloud service providers. CSA STAR makes use of the cloud control matrix version 4.0. STAR consists of three levels of certification.

- Level 1, Self-Assessment, is a complimentary offering that documents the security controls provided by various cloud computing offerings, helping users assess the security of cloud providers they currently use or are considering using.
- Level 2, Attestation, is a collaboration between CSA and the AICPA to provide guidelines for CPAs to conduct SOC 2 engagements using criteria from the AICPA (Trust Service Principles, AT 101) and the CSA Cloud Controls Matrix. STAR Attestation provides for rigorous third-party independent assessments of cloud providers.
- Level 3, Continuous Auditing, consists of monitoring through automated processes that ensure security controls are monitored and validated at all times.

Since the registry of certified providers is publicly available, the STAR program makes it easy for a company to assess the relative risk of a provider and should certainly be consulted when assessing any new CSP.

ENISA Cloud Certification Schemes List and Metaframework

ENISA has provided another resource that can be used by cloud security professionals to assess the relative risk of a cloud service provider. (See `resilience.enisa.europa.eu/cloud-computing-certification`.) The Cloud Certification Schemes List (CCSL) provides an up-to-date summary of existing certification schemes that could be relevant for cloud computing. The list as of 2020 is as follows:

- Certified Cloud Service, TÜV Rheinland
- CSA Attestation, OCF Level 2
- CSA Certification, OCF Level 2
- CSA Self-Assessment, OCF Level 1
- EuroCloud Self-Assessment
- EuroCloud Star Audit Certification
- ISO/IEC 27001 Certification
- Payment Card Industry Data Security Standard v3
- Leet Security Rating Guide
- Service Organization Control (SOC) 1
- Service Organization Control (SOC) 2
- Service Organization Control (SOC) 3
- Cloud Industry Forum Code of Practice

ENISA describe the Cloud Certification Schemes Metaframework (CCSM) as an extension of CCSL. (See Figure 6.4.) It is a metaframework of cloud certification schemes. The goal of the metaframework is to provide a neutral high-level mapping from the customer's network and information security requirements to security objectives in existing cloud certification schemes, which facilitates the use of existing certification schemes during procurement. The agency also provides an online tool where it is possible to select from 27 different security objectives to see how the relative certifications compare.

CCSM security objectives	Certified Cloud Service - TÜV Rheinland	CSA Attestation - OCF Level 2	CSA Certification - OCF Level 2	CSA Self Assessment - OCF Level 1	EuroCloud Self Assessment	EuroCloud Star Audit Certification	ISO/IEC 27001 Certification	Leet Security Rating Guide	Service Organization Control (SOC) 2	Service Organization Control (SOC) 3	Cloud Industry Forum Code of Practice
Information security policy	●	●	●	●	●	●	●	●	●	●	●
Risk management	●	●	●	●	●	●		●	●	●	●
Security roles	●	●	●	●	●	●	●	●	●	●	●
Security in Supplier relationships	●	●	●	●	●	●	●	●	●	●	●
Background checks	●	●	●	●		●	●	●	●	●	
Security knowledge and training	●	●	●	●	●	●	●	●	●	●	●
Personnel changes	●	●	●	●		●	●	●	●	●	●
Physical and environmental security	●	●	●	●	●	●	●	●	●	●	●
Security of supporting utilities	●	●	●	●		●	●	●	●	●	●
Access control to network and information systems	●	●	●	●	●	●	●	●	●	●	●
Integrity of network and information systems	●	●	●	●	●	●	●	●	●	●	●
Operating procedures	●	●	●	●		●	●	●	●	●	●

FIGURE 6.4 **CSA CCSM online tool**

UNDERSTANDING OUTSOURCING AND CLOUD CONTRACT DESIGN

Outsourcing is defined as "the business practice of hiring a party outside a company to perform services and create goods that traditionally were performed in-house by the company's own employees and staff." (See `www.investopedia.com/terms/o/outsourcing.asp`.) Outsourcing was traditionally a practice undertaken by companies as a cost-cutting measure or to make use of outside expertise or specialization. As the practice gained favor, it affected a wide range of jobs, ranging from customer support to manufacturing to the back office.

So, what is cloud computing but outsourcing? Companies are taking advantage of specialized skills and lower costs to improve their business models. And they are doing it in a multibillion-dollar way that is growing year over year. Cloud security professionals are well served by understanding key contractual provisions that can make sure their companies can do business with cloud providers with their data and wallets intact.

Business Requirements

Before entering into a contract with a cloud service provider, it is important for any business to fully understand their own business needs and how they will be met. The evolution of the cloud means that more and more IT functions can use cloud technologies every year. Once a business has made the decision that a function is cloud ready and can codify the needs required of the system, a negotiation can take place with a cloud service provider. In legal terms, a cloud customer and a cloud service provider enter into a master service agreement (MSA), defined as any contract that two or more parties enter into as a service agreement.

But what exactly should be in a contract with a cloud service provider? The Cloud Standards Customer Council (CSCC) provides a good starting point for any contract in its "Practical Guide to Cloud Service Agreements." (See `www.omg.org/cloud/deliverables/CSCC-Practical-Guide-to-Cloud-Service-Agreements.pdf`.) This document provides a reference to help enterprise information technology and business decision-makers analyze cloud service agreements from different cloud service providers. They break cloud contracts down into three major areas:

- The "Customer Agreement" section of the agreement describes the overall relationship between the customer and cloud provider. Definitions of the roles, responsibilities, and execution of processes need to be explicitly identified and formally agreed upon. This section may outline communication processes and personnel.
- An acceptable use policy (AUP) is almost always included within a cloud contract. The AUP defines and prohibits activities that providers consider to be an improper or outright illegal use of their service. There is considerable consistency across cloud providers when it comes to AUP. Although specific details of acceptable use may vary somewhat among cloud providers, the scope and effect of these policies is typically similar, and these provisions typically generate the least concerns or resistance.
- The service level agreement (SLA) within the cloud contract describes levels of service using various attributes and metrics such as availability, serviceability, or performance. The SLA specifies thresholds, financial penalties, and who is responsible for monitoring violations of these thresholds. Well-designed SLAs can significantly contribute to avoiding future conflict and facilitating the resolution of issues before they escalate into a dispute.

Key SLA Requirements

Service level agreements can be a key factor in avoiding potential issues once a contract is in place. Service metrics, such as uptime or quality of service, can be included in this section of a contract. An uptime of 99 percent may seem adequate, but that level of allowed downtime would be equal to 87.6 hours a year! Service level agreements often include sections on the following:

- Uptime guarantees
- SLA violation penalties
- SLA violation penalty exclusions and limitations
- Suspension of service clauses
- Provider liability
- Data protection and management
- Disaster recovery and recovery point objectives
- Security and privacy notifications and timeframes

Vendor Management

The migration to the cloud relieves an IT department and IT executives of the duties of maintaining internal technology stacks in many ways. This has shifted the role that IT professionals play in vendor management, from primarily a point-in-time acquisition process (purchasing hardware or software) to an ongoing management relationship with a cloud vendor. This redefined relationship requires a great deal of trust and communication with vendors. Cloud professionals need strong project *and* people management skills. Managing cloud vendor relationships should focus on a number of areas.

- **Initial contract negotiations**: There are many issues to address before contracts are signed, including SLAs, personnel (point persons on the company and vendor side), division of labor, entry and exit strategies, and potentially proof of concept environments.
- **Service level agreement monitoring and project coordination**: Who is responsible for monitoring and alerting on SLA metrics? How often will these metrics be adjusted as project requirements evolve?
- **User community communications**: Company IT and the cloud vendor should not be communicating different things to the user community. Careful coordination of who sends information to the users is a key part of the relationship.

- **Security and business processes**: The coordination of both security requirements and business process integrations is a tightly shared duty of IT and the cloud service provider. Both parties should work together to ensure that requirements are considered from each perspective in an ongoing manner.
- **Upgrades**: The cloud is not a static environment. Cloud offerings change often and offer new capabilities. Managing these changes and ensuring that they are applied in an efficient and consistent manner requires working closely together on enhancements.

Contract Management

The management of cloud contracts is a core business activity that is central to any ongoing relationship with a cloud service provider. A company must employ adequate governance structures to monitor contract terms and performance and be aware of outages and any violations of stated agreements. The CSCC provides a prescriptive series of steps that should be taken by cloud customers to evaluate their cloud contracts to compare multiple cloud providers or to monitor and evaluate a contract with a selected provider. The following steps are important to discuss in detail (`www.omg.org/cloud/deliverables/CSCC-Practical-Guide-to-Cloud-Service-Agreements.pdf`):

- **Understanding of roles and responsibilities**: Clearly identifying who is responsible for which activities ensures that there are no gaps that could lead to problems when using cloud services.
- **Evaluation of business level policies**: Companies must consider policy requirements and how a cloud provider can meet them. The data policies of the cloud provider are perhaps the most critical business-level policies.
- **Understanding of service and deployment model differences**: To increase effectiveness, specific components of the contract should be stated in measurable terms and should include the services to be performed and outcome expectations, Key Performance Indicators (KPIs), how services are measured, and the reporting requirements.
- **Identification of critical performance objectives**: Cloud customers must decide which measures are most critical to their specific cloud environments and ensure these measures are included in their SLA.
- **Evaluation of security and privacy requirements**: This includes information on security controls, data classifications, auditing, and breach notification.
- **Identification of service management requirements**: This may consist of the internal controls, management, automation, and self-healing that comprise application performance management.

- **Preparing for service failure management**: Service failure management outlines what happens when the expected delivery of a cloud service does not occur. What happens when a service fails?
- **Understanding the disaster recovery plan**: "There is no cloud, there is just someone else's computer" is a popular slogan for IT professionals often found on T-shirts and stickers in an IT department. Just because businesses are outsourcing platforms to cloud service providers does not absolve them of the need for serious disaster planning.
- **Develop an effective governance process**: The control and oversight of the outside provider (the cloud service provider) and regular ongoing review of the contracted and actual use of each cloud service should be included in the governance process.
- **Understanding the exit process**: An exit clause should be part of every cloud agreement and describes the details of the exit process including the responsibilities of the cloud provider and consumer in case the relationship terminates prematurely or otherwise. Data ownership, transferability, and destruction should be clearly and prominently identified.

Cyber Risk Insurance

Cyber risk insurance is designed to help an organization modify risk by sharing the risk of a cyber incident with others via a policy that might offset costs involved with recovery after a cyber-related security incident such as a breach. This is a growing field, and many more companies are choosing to indemnify against the threats of harm from incidents. Cyber risk insurance usually covers costs associated with the following:

- **Investigation**: Costs associated with the forensic investigation to determine the extent of an incident. This often includes costs for third-party investigators.
- **Direct business losses**: Direct monetary losses associated with downtime or data recovery, overtime for employees, and oftentimes, reputational damages to the company.
- **Legal notifications**: Costs associated with required privacy and breech notifications required by relevant laws.
- **Lawsuits**: Policies can be written to cover losses and payouts due to class action or other lawsuits against a company after a cyber incident.
- **Extortion**: The insurance to pay out ransomware demands is growing in popularity. This may include direct payments to ensure data privacy or accessibility by the company.

Supply Chain Management

Imagine if you would that you are a highly successful CEO of a major American corporation overseeing your sixth year of growth. Your compensation package is in the tens of millions of dollars. Now imagine being forced out of your role because of lacking security controls at your company's HVAC vendor, which led to one of the largest breaches of private data in history (at the time). This is the story of Target CEO Gregg Steinhafel in 2014, after the holiday 2013 POS attacks at the department store.

This is just one high-profile example of how the weakest link in a security chain can lead to disastrous consequences. The cloud can either increase or reduce the risks posed to a company depending on how a company implements a cloud strategy. In any cloud strategy, it is important to understand the extent of the organization's reliance to a cloud service provider as well as the third parties that power and enable the CSP itself.

The supply chain should always be considered in any business continuity or disaster recovery planning. The same concepts of understanding dependencies, identifying single points of failure, and prioritizing services for restoration are important to apply to the entire supply chain. This includes the cloud services a company may employ to deliver their services *and* the associated third parties that enable that cloud provider. With a complete view of the potential risks posed by a CSP supply chain, a company can choose to avoid, modify, share, or retain the risks in a similar evaluation to other risks as discussed earlier.

ISO 27036: Information Security for Supplier Relationships

As we have seen with previous standards, the ISO provides a framework to allow for the evaluation and treatment of information risks involved in the acquisition of goods and services from suppliers. (See `www.iso.org/obp/ui/#iso:std:iso-iec:27036:-4:ed-1:v1:en`.) This is a four-part standard, part four (2016) of which "Guidelines for security of cloud services" provided excellent guidance for understanding and mitigating the risk of the cloud supply chain. According to ISO, the application of this standard should result in the following:

- Increased understanding and definition of information security in cloud services
- Increased understanding by the customers of the risks associated with cloud services to enhance the specification of information security requirements
- Increased ability of cloud service providers to provide assurance to customers that they have identified risks in their service(s) and associated supply chains and have taken measures to manage those risks

ISO 27036 is not yet available for free but provides a strong reference for understanding supply chain risk in the cloud context. Additional resources worth review include the draft NISTIR 8276 (Key Practices in Cyber Supply Chain Risk Management:

Observations from Industry), NIST SP800-161 (Supply Chain Risk Management Practices for Federal Information Systems and Organizations), and the 2015 ENISA publication on supply chain integrity.

SUMMARY

A cloud security professional must be constantly aware of the legal and compliance issues in migrating and maintaining systems in the cloud. Understanding the legal requirements, privacy issues, audit challenges, and how these relate to risk and contracts with cloud providers is a must for any company taking advantage of cloud services. A cloud security professional must also be well versed in the frameworks provided by professional organizations such as ENISA, NIST, ISO, and the CSA. All information security is essentially a business risk, and understanding and mitigating the risks faced by a company is critical to any business strategy. Any cloud security professional should understand the role that IT plays in this larger picture and have the communication and people skills to involve the appropriate business, legal, and risk decision-makers within a company.

Index

A

B

C

D

E

F

G

H

I

J–K

L

M

N

O

P

Q

R

S

T

U

V

W

X–Y–Z

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.