Chapter 1

# Cloud Principles and Design

## THE FOLLOWING COMPTIA CLOUD ESSENTIALS+ EXAM CLO-002 OBJECTIVES ARE COVERED IN THIS CHAPTER:

✓ **1.1 Explain cloud principles.**

- Service models
  - SaaS
  - IaaS
  - PaaS
- Deployment models
  - Public
  - Private
  - Hybrid
- Characteristics
  - Elastic
  - Self-service
  - Scalability
  - Broad network access
  - Pay-as-you-go
  - Availability
- Shared responsibility model

✓ **1.4 Summarize important aspects of cloud design.**

- Redundancy
- High availability
- Disaster recovery
- Recovery objectives
  - RPO
  - RTO

The computer industry is an industry of big, new trends. Every few years, a new technology comes along and becomes popular, until the next wave of newer, faster, and shinier objects comes along to distract everyone from the previous wave. Thinking back over the past few decades, there have been several big waves, including the rise of the Internet, wireless networking, and mobile computing.

The biggest recent wave in the computing world is cloud computing. Its name comes from the fact that the technology is Internet based; in most computer literature, the Internet is represented by a graphic that looks like a cloud. It seems like everyone is jumping on the cloud (pun intended, but doesn't that sound like fun?), and perhaps you or your company have used cloud technologies already. But to many people, the cloud is still nebulous and maybe even a little scary. There's a lot to know about the nuts and bolts of cloud computing.

The CompTIA Cloud Essentials+ certification exam assesses cloud knowledge from the perspective of the business analyst. Business analysts come from a variety of backgrounds, including technophiles who have business-facing roles, those who have business acumen but little technical experience, and anywhere in between. This certification—and this study guide—attempts to balance technical data with practical business information. You'll find some techspeak here, but we won't get into the hard-core inner workings of cloud management.


If you're interested in the more technical side of cloud management, consider the CompTIA Cloud+ certification. To help prepare for it, look for the *CompTIA Cloud+ Study Guide, Second Edition*, by Todd Montgomery and Stephen Olson (Sybex, 2018).

This chapter starts off by diving into the fundamentals of cloud principles and cloud design. A business analyst (BA) working with cloud providers should understand common service and deployment models, cloud characteristics, and important design aspects such as redundancy, high availability, and disaster recovery.

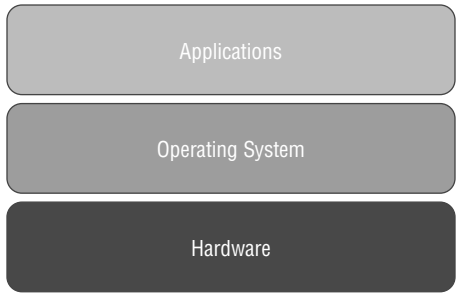
# Understanding Cloud Principles

You hear the term a lot today—*the cloud*. You're probably already using cloud services whether you realize it or not. For example, if you store music in iCloud, use Gmail, or use Microsoft Office 365, you are using cloud-based services. But what exactly is the cloud? The way it's named—and it's probably due to the word *the* at the beginning—makes it sound as if it's one giant, fluffy, magical entity that does everything you could ever want a computer to do. Only it's not quite that big, fluffy, or magical, and it's not even one thing. To help illustrate what the cloud is, let's first consider a pre-cloud example—we'll call it *traditional computing*.

Imagine that you are at a small business in the 1980s—you probably have big hair and make questionable choices in the music you listen to—and need to decide if the company should purchase a computer. This decision might sound silly today, but before the mid-1990s, it was one that required many companies to seriously weigh the pros and cons. Desktop computers were almost a luxury item, and kids' lemonade stands certainly didn't have card readers to accept mobile payments! Perhaps the accounting department could automate payroll, the sales team could make promotional materials, or the boss just wanted to play solitaire. Regardless of the reason, someone went out and priced computers and made the business case to purchase one.

The computer would be traditional in all senses of the word. It would have a collection of hardware such as a processor, memory, and hard drive, an operating system (OS) that interfaced with the hardware, and one or more applications that allowed employees to complete tasks. Figure 1.1 illustrates this traditional computing model. You can see that the hardware is considered the base, and the OS and apps build upon it.

**FIGURE 1.1** Traditional computing model




Over the years, the company expands, and more employees need computers. Eventually the computers need to talk to each other, and things like centralized storage and a database are required. So along with computers, you have to buy expensive server hardware, storage devices, and networking equipment such as switches, routers, and a firewall. The costs are adding up, and every year it seems like the IT budget gets squeezed. To top it off, every few years much of the hardware becomes obsolete and really should be replaced. The already tight budgets become even more challenging to manage. To add even another complication, software companies keep producing new versions with features that employees say are critical, so the software needs to be upgraded as well. (The same holds true for OSs.)

The pace of innovation can be staggering, and the cost of keeping up can be overwhelming for many businesses. But in the traditional model of computing, it was just the cost of doing business. Then in the late 2000s, cloud computing started changing everything.

*Cloud computing* is a method by which you access remote servers that provide software, storage, database, networking, or compute services for you. Instead of your company needing to buy the hardware and software, another company does, and yours essentially rents it from them and accesses it over the Internet. There isn't one cloud but hundreds of

commercial clouds in existence today. Many of them are owned by big companies, such as Amazon, Microsoft, Google, HP, and Apple. The three most popular ones for businesses are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud.

One great feature of the cloud is that using it is pretty simple in most cases. If a user can open a web browser and navigate to a website, they can use the cloud. We'll get into more details of different types of cloud access in Chapter 2, "Cloud Networking and Storage."

There are many advantages to cloud computing, and the most important ones revolve around money. Cloud providers get economies of scale by having a big pool of resources available to share among many clients. It may be entirely possible for them to add more clients without needing to add new hardware, which results in greater profit. From a client company's standpoint, the company can pay for only the resources it needs without investing large amounts of capital into hardware and software that will be outdated in a few years. Using the cloud is often cheaper than the alternative.

For clients, there are several other advantages of cloud computing as well, including the following:

- Access to resources from anywhere in the world from multiple types of client computers and devices
- Flexibility and scalability, because additional resources can be quickly added or removed
- Access to newer technologies without someone needing to take a long time to set them up
- Data protection through the use of enterprise-level backups
- Reduced IT staff and administration costs

Plus, if there is a hardware failure within the cloud, the provider handles it. If the cloud is set up right, the client won't even know that a failure occurred.

The biggest disadvantage of the cloud has been security. The client company's data is stored on someone else's server, and company employees are sending it back and forth via the Internet. Cloud providers have dramatically increased their security mechanisms over the last several years, but there can still be issues, especially if the data is highly sensitive material or personally identifiable information (PII). We'll talk more about security in the "Shared Responsibility Model" section later in this chapter. Other disadvantages include potential downtime, either from the service provider being down or from the lack of an Internet connection, and limited control. Some companies don't like the fact that they don't own the assets or have the ability to perform some administrative tasks.

If it seems like there's a lot to learn about the cloud, there is. We'll break it down into manageable sections to help it make more sense. Within this chapter, we will cover the following topics:

- Service models
- Deployment models
- Cloud characteristics
- Shared responsibility model
- Cloud design

Before we get into those topics, though, we're going to take a slight detour into the technology that makes cloud computing possible—virtualization.

## Virtualization

Cloud computing is possible thanks to a concept called *virtualization*, which means that there isn't necessarily a one-to-one relationship between a physical server and a logical (or virtual) server or services. In other words, there might be one physical server that virtually hosts cloud services for a dozen companies, or there might be several physical servers working together as one logical server. From the end user's side, the concept of a physical machine versus a virtual machine (VM) doesn't even come into play, because it's all handled behind the scenes.


Virtualization has been around in the computer industry since 1967, but it has only recently exploded in popularity thanks to the flexibility that the Internet offers.

Perhaps the easiest way to understand virtualization is to compare and contrast it to what we called the traditional computing model earlier in this chapter. In the traditional computing model, a computer is identified as being a physical collection of hardware that is running some combination of software, such as an OS and various applications. There's a one-to-one relationship between the hardware and the OS.

For the sake of illustration, imagine that a machine is a file server and now it needs to perform the functions of a web server as well. To make this happen, the administrator would need to ensure that the computer has enough resources to support the service (processor, memory, network bandwidth), install web server software, configure the appropriate files and permissions, and then bring it back online as a file and web server. These would be relatively straightforward administrative tasks.

But now imagine that the machine in question is being asked to run Windows Server and Linux at the same time. Now there's a problem. In the traditional computing model, only one OS can run at one time, because each OS completely controls the hardware resources in the computer. Sure, an administrator can install a second OS and configure the server to dual-boot, meaning which OS to run is chosen during the boot process, but only one OS can run at a time. So if the requirement is to have a Windows-based file server and a Linux-based web server, there's a problem. Two physical computers are needed.

Similarly, imagine that there is a Windows-based workstation being used by an applications programmer. The programmer has been asked to develop an app that works in Linux, macOS, or anything other than Windows. When the programmer needs to test the app to see how well it works, what does she do? She can configure her system to dual-boot, but once again, in the traditional computing model, she's limited to one OS at a time per physical computer. Her company could purchase a second system, but that quickly starts to get expensive when there are multiple users with similar needs.

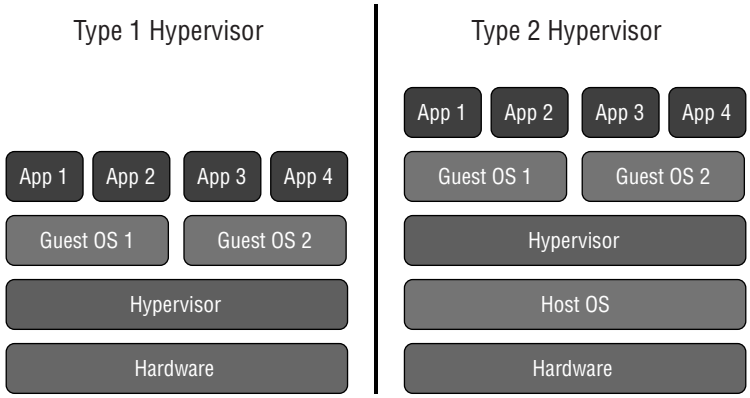
This is where virtualization comes in. The term *virtualization* is defined as creating virtual (rather than actual) versions of something. In computer jargon, it means creating

virtual environments where "computers" can operate. We use quotation marks around the word *computers* because they don't need to be physical computers in the traditional sense. Virtualization is often used to let multiple OSs (or multiple instances of the same OS) run on one physical machine at the same time. Yes, they are often still bound by the physical characteristics of the machine on which they reside, but virtualization breaks down the traditional one-to-one relationship between a physical set of hardware and an OS.

## The Hypervisor

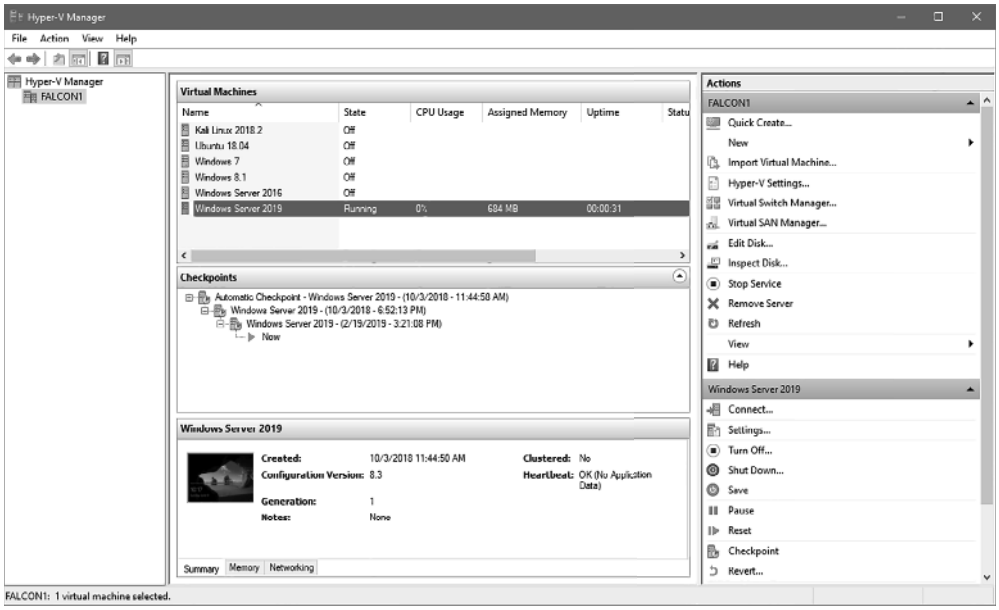
The key enabler for virtualization is a piece of software called the *hypervisor*, also known as a *virtual machine manager (VMM)*. The hypervisor software allows multiple OSs to share the same host, and it also manages the physical resource allocation to those virtual OSs. As illustrated in Figure 1.2, there are two types of hypervisors: Type 1 and Type 2.

**FIGURE 1.2** Type 1 and Type 2 hypervisors




A Type 1 hypervisor sits directly on the hardware, and because of this, it's sometimes referred to as a *bare-metal* hypervisor. In this instance, the hypervisor is basically the OS for the physical machine. This setup is most commonly used for server-side virtualization, because the hypervisor itself typically has very low hardware requirements to support its own functions. Type 1 is generally considered to have better performance than Type 2, simply because there is no host OS involved and the system is dedicated to supporting virtualization. Virtual OSs are run within the hypervisor, and the virtual (guest) OSs are completely independent of each other. Examples of Type 1 hypervisors include Microsoft Hyper-V, VMware ESX, and Citrix XenServer. Figure 1.3 shows an example of the Microsoft Hyper-V interface, running on a Windows 10 workstation. In the "Virtual Machines" pod, you can see that this system is running six VMs, including ones for Linux, Windows clients, and Windows Server. These VMs can be turned on or off at any time and can provide flexibility in production or testing environments.

**FIGURE 1.3** Microsoft Hyper-V




A Type 2 hypervisor sits on top of an existing OS (called the host OS), such as Microsoft Windows. This is most commonly used in client-side virtualization, where multiple OSs are managed on the client machine as opposed to on a server. An example of this would be a Windows user who wants to run Linux at the same time as Windows. The user could install a hypervisor and then install Linux in the hypervisor and run both OSs concurrently and independently. The downsides of Type 2 hypervisors are that the host OS consumes resources, such as processor time and memory, and a host OS failure means that the guest OSs fail as well. Examples of Type 2 hypervisors include Microsoft's Windows Virtual PC and Azure Virtual Server, Oracle VM VirtualBox, VMware Workstation, and Linux KVM.


Linux version 2.6.20 (released in 2007) and newer come with built-in open source virtualization software called the *Kernel-based Virtual Machine (KVM)*. (The CompTIA acronyms list refers to it as *Kernel Virtual Machine*.) The KVM software allows any Linux installation to be turned into a hypervisor and run multiple VMs.

## Virtualization Resource Requirements

As you might expect, running multiple OSs on one physical computer can require more resources than running a single OS. There's no rule that says a computer being used for virtualization is required to have more robust hardware than another machine, but for performance reasons, the system should be fairly well equipped. This is especially true for

systems running a Type 2 hypervisor, which sits on top of a host OS. The host OS will need resources too, and it will compete with the VMs for those resources.

If you are going to purchase a computer for the purpose of running virtualization, be sure to get a system with plenty of capability; for example, consider a fast multicore processor, lots of memory, a large and fast hard drive (preferably SSD), and a fast network card.

### Real World Scenario

#### Why Understanding Virtualization Is Important

Virtualization isn't a Cloud Essentials+ exam objective, and it's highly unlikely that you will see exam questions on it. So then why do we talk about virtualization and hypervisors? There are two real-world situations where this knowledge could be helpful to you.

First, if you are working with a cloud service provider (CSP) and asking for services, it's good to know a little about what's happening behind the curtain. It can help you ask the right questions and receive the appropriate services without overpaying for them. In addition, cloud concepts such as elasticity, scalability, and broad network access (which we cover later in this chapter) will make a lot more sense!

Second, you might have users in your company who could benefit from a virtualization workstation. Software developers are a great example. Armed with this knowledge, you can help determine whether a virtualization workstation meets the business needs and understand why you might need to pay for a slightly more powerful machine and additional software to support the user's needs.

Exercise 1.1 walks you through some steps you can take to assess the virtualization needs for a user in your company. Some of the items are more technical. If you're not comfortable with them, enlist someone from the IT department as a partner.

**EXERCISE 1.1**

### Determining a User's Virtualization Needs

1. Understand why the client needs virtualization.

   For example, do they have a Windows or Mac computer and need to run Linux? Or perhaps they have a Windows computer and want to run macOS at the same time? Determine their needs and then determine what is needed to secure the additional OSs, such as licenses.

2. Evaluate their computer to ensure that it can support the VM. If not, the user may require a new computer.

   - Does the processor support virtualization?
   - How much RAM does the computer have? Is it enough to meet the minimum requirements of all installed OSs?

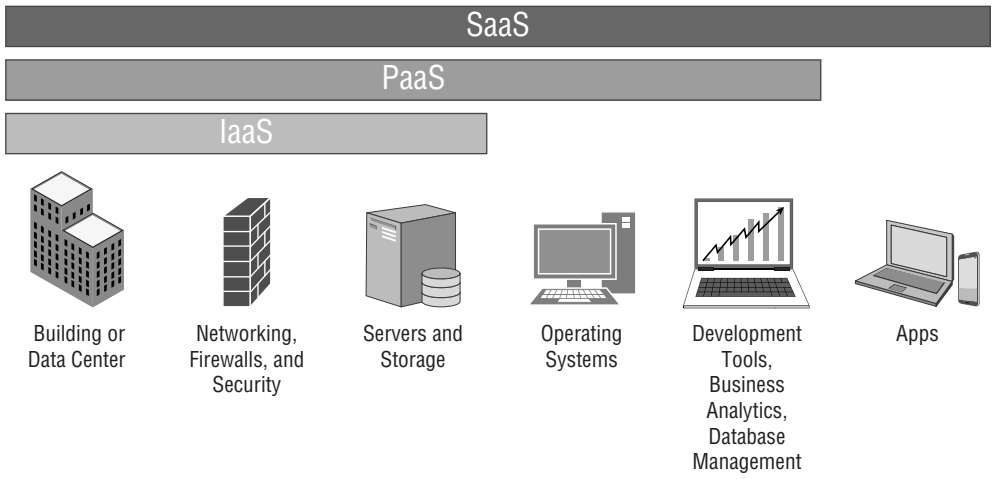
- How much free hard drive space is there? It needs to be enough to install the hypervisor and the guest OS as well as to store any files that need to be stored from within the guest OS.
- Does the system have a fast enough network connection if the guest OS needs access to the network?

3. Consider which hypervisor to use. Is there a version of the hypervisor that's compatible with the host OS?
4. Consider security requirements. If the guest OS will be on the Internet, be sure to obtain the proper security software.
5. After all of these have been considered, make the business case for why or why not to proceed. Is the cost required worth the benefits provided?

## Service Models

Cloud providers sell everything "as a service." The type of service is named for the highest level of technology provided. For example, if compute and storage are the highest level, the client will purchase infrastructure as a service. If applications are involved, it will be software as a service. Nearly everything that can be digitized can be provided as a service. Figure 1.4 introduces the key differences between the three primary types of services: IaaS, PaaS, and SaaS. We'll define those acronyms and describe each of them in further detail in the following sections, starting from the bottom.

**FIGURE 1.4** Cloud service models



## Infrastructure as a Service

Let's say that a company needs extra network capacity, including processing power, storage, and networking services (such as a firewall, which we will discuss in Chapter 2) but doesn't have the money to buy more network hardware. Instead, it can purchase *infrastructure as a service (IaaS)*, which is virtual hardware that replaces physical, on-premises IT infrastructure. The client decides what they need, and the CSP provisions it.

You will hear the term *provisioning* a lot. It just means that the CSP allocates those resources to a client.

IaaS can quickly scale up or down based on client needs. Paying for IaaS is often a lot like paying for utilities—the client pays for what it uses. Common IaaS use cases include the following:

**Hosting of Websites and Web Apps** This includes the rapidly growing industry of the Internet of Things (IoT), which we will discuss further in Chapter 3, "Assessing Cloud Needs."

**High-Performance Computing** Examples include big data analytics, machine learning, and simulations and modeling. Big temporary projects (sometimes called *batch jobs*) can be run. Once they are finished, the resources are turned off and no longer used or paid for.

**Data Storage, Backup, and Recovery** IaaS can replace traditional on-premises data centers in organizations.

**Testing and development** IaaS allows a client to quickly scale up an environment to test a new product or service and then shut that environment back down.

Most providers offer features such as self-serve web-based interfaces and management tools. When looking for an IaaS solution, there are three aspects to pay particular attention to.

**High Security Standards** If not properly implemented, cloud security can be a major business risk. Check to ensure that the company has third-party security assessments or Statement on Standards for Attestation Engagements (SSAE) 18 audits.

SSAE 18 audits are a set of rigorous auditing standards for CSPs, which ensure that the CSP is providing adequate security. The details of an SSAE 18 audit are beyond the scope of the CompTIA Cloud Essentials+ exam. If you are interested in more information, check out `https://ssae-18.org` or `https://www.aicpa.org/research/standards/auditattest/ssae.html`.

**High Availability** IaaS resources aren't useful if you can't access them. Seek to understand what the provider has for guaranteed uptime offerings, and work that into the

service level agreement (SLA). We'll discuss this more in the "High Availability" section later in this chapter.

**Flexible Pricing Arrangements** One of the nice features of IaaS is the ability to scale resources up or down as needed. Be sure that the CSP offers pricing at the hourly level or in even shorter increments if that's what you need.

Of the three service models, IaaS requires the most network management expertise from the client. In an IaaS setup, the client provides and manages the software. Examples of IaaS products include AWS, Microsoft Azure, Google Compute Engine (GCE), Rackspace, and Joyent Triton.

## Platform as a Service

Companies and individuals developing software typically want to focus on their area of expertise, which is coding. They don't want the extra hassle of managing hardware platforms to slow them down. That's where PaaS comes in.

*Platform as a service (PaaS)* adds features to IaaS that include OSs and software development tools such as runtime environments. Sometimes you will hear this collection of tools called *cloudware*. The vendor manages the various hardware platforms, freeing up the software developer to focus on building their application and scaling it. Overall, this should speed up the software development process and decrease its cost. Some key features of PaaS include the following:

- The ability to develop and deploy apps in a wide variety of languages, such as Java, JavaScript Object Notation (JSON), .NET, Python, PHP, and Node.js
- Cross-platform support, such as Windows, macOS, iOS, Android, and Linux
- Support for multiple, geographically separated development teams
- Integrated app lifecycle management

Other services that some PaaS providers may offer include help with provisioning, app deployment, load balancing, autoscaling, machine learning, integration with third-party application programming interfaces (APIs), and development operations (DevOps). Finally, some CSPs even have online drag-and-drop tools and templates that make app development possible for nonexperts. Examples of PaaS solutions include Google App Engine, Microsoft Azure, Red Hat OpenShift PaaS, AWS Elastic Beanstalk, Oracle Cloud Platform (OCP), Salesforce aPaaS, and Mendix aPaaS.

## Software as a Service

The highest level of the three common cloud services is *software as a service (SaaS)*, which is a software licensing and delivery model where apps are subscribed to and accessed over the Internet. It has the largest market size of the three standard services and is likely the one you are most familiar with.

Clients using SaaS pay a monthly or yearly subscription fee to access software. Providers manage the software, software updates, and patches, as well as all underlying infrastructure needed to run it. Common examples include Google Docs and Microsoft Office 365 as

well as email clients such as Gmail. Custom business applications can be built and deployed in a SaaS model as well, such as those for accounting, customer relationship management (CRM), or design.

As with other cloud services, the primary advantages are cost, flexibility, and ease of access. Clients will save money in the long run by not needing to purchase new software every few years, and any user with a web browser can access the software, regardless of the type of device they are using.

The disadvantages are few but can be significant. The biggest worry is generally data security. When working with a cloud, all data is being transmitted over the Internet, and data is stored "somewhere else" other than your on-premises data center. A smaller concern is that there are a relatively limited number of web apps available. This will change as the market matures, but finding the right type of software may be a challenge. Third, an Internet connection is required to use the apps and save data. In the event of a dropped connection, most apps have the ability to temporarily store changed data in the local device's memory, but that's not guaranteed. A lost Internet connection could result in data loss or the inability to access data that's needed.

Even though you may already be familiar with using software online, it never hurts to get hands-on experience to further your learning. Exercise 1.2 gives you some practice using SaaS, in the form of Google Apps.

## EXERCISE 1.2

### Using Google's Cloud Services

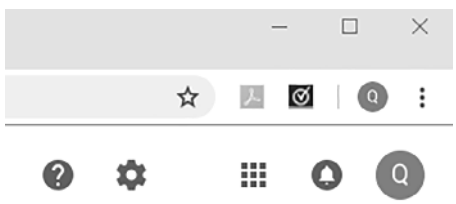
1. Open Google at `www.google.com`.

   If you do not already have a Google account, you will need to create one. With it, you use Google's online apps and storage as well as a Gmail account.

2. If you are doing this exercise on your own, create a second account to share files and folders with.

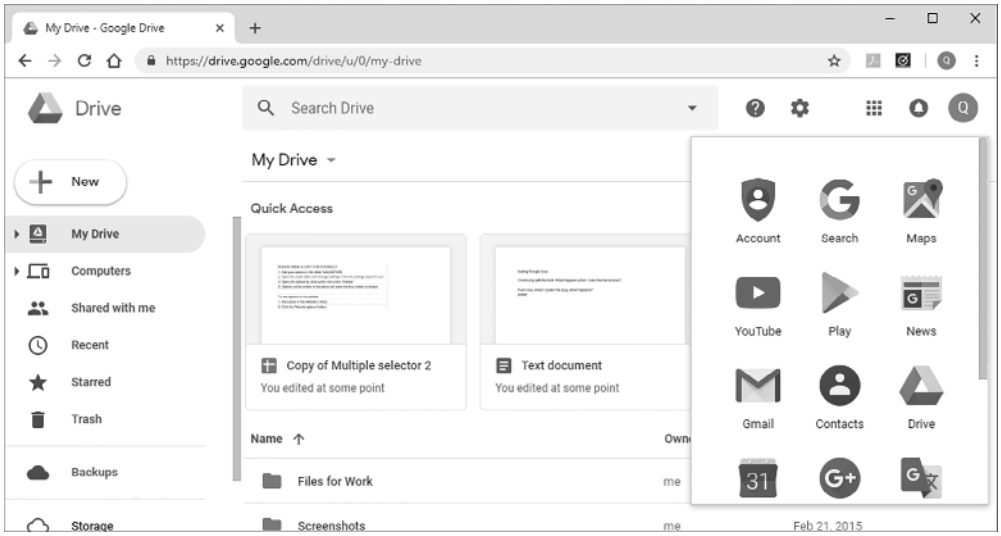
3. Once you're logged in, click the Apps icon in the upper-right corner. It's the one that has nine small squares (see Figure 1.5).

**FIGURE 1.5** Google icons



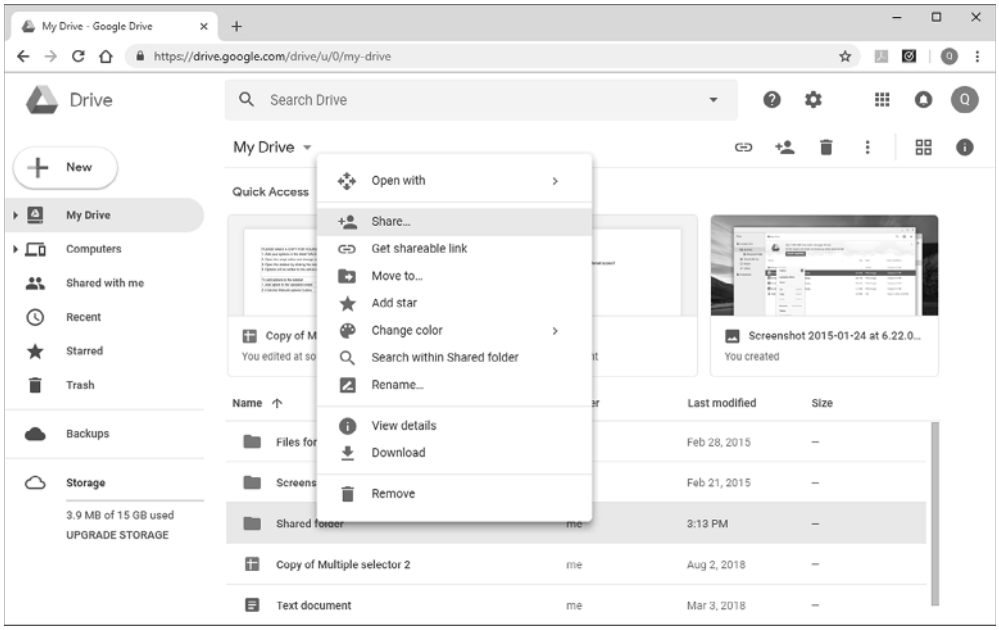
This will open Apps, as shown in Figure 1.6.

**FIGURE 1.6** Google Apps




4. In Apps, click Drive. This will open Google Drive, as shown in Figure 1.7.

**FIGURE 1.7** Google Drive



**EXERCISE 1.2 *(continued)***

5. Create a folder by clicking New ➢ Folder and share it with another account.
6. Also create a document or spreadsheet using Google's online software.

   How easy or difficult was it?
7. If necessary, log out and log in to the other account that you created to access the resources that were shared with you.

   How easy or difficult was it?

## Other Service Models

The three services we've already discussed are the most common, but you will see many other types in the market as well. They are all offshoots and combinations of IaaS, PaaS, or SaaS. In some cases, CSPs will come up with their own names in an effort to make the service sound unique or special. Don't be fooled by marketing gimmicks! If you hear a term you're not familiar with, seek to understand exactly what it is that you're getting.

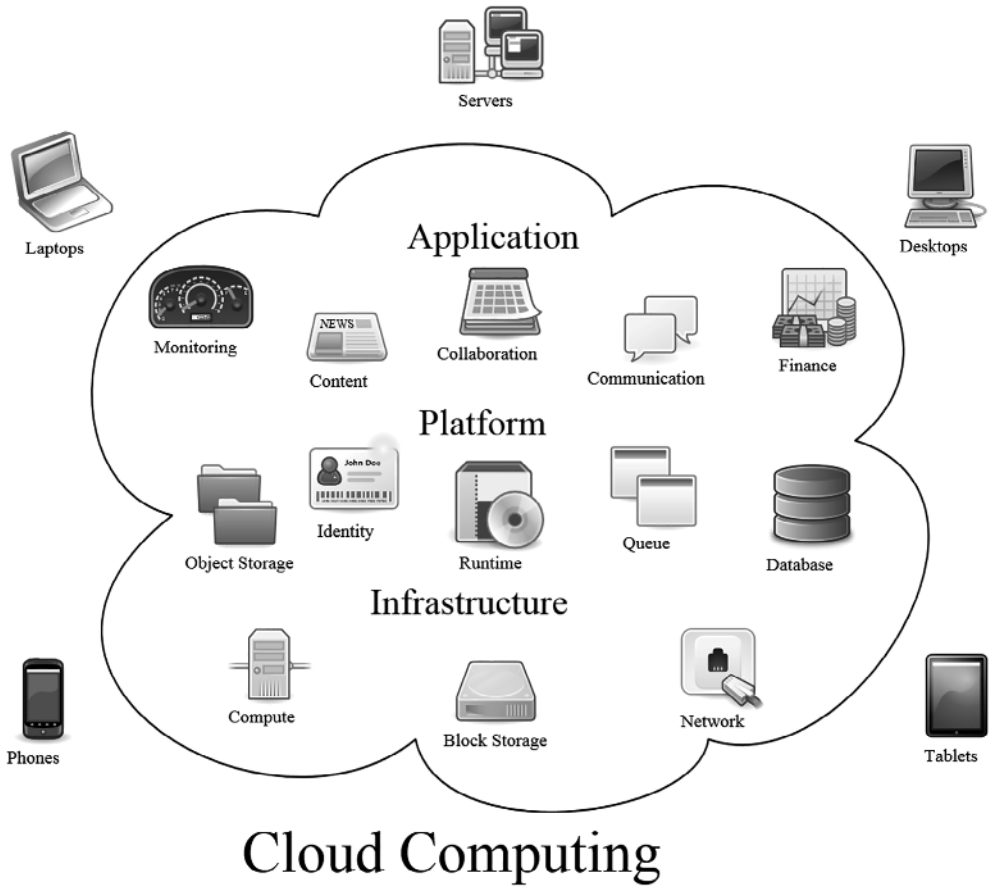
Here's a list of other service levels that you may see in the real world:

- *Business processes as a service (BPaaS)* provides business processes such as payroll, IT help desk, or other services.
- *Communications as a service (CaaS)* provides things like Voice over IP (VoIP), instant messaging, and video collaboration.
- *Desktop as a service (DaaS)* provides virtual desktops so that users with multiple devices or platforms can have a similar desktop experience across all systems.
- *Data as a service (also DaaS)* provides for a mash-up of multiple sources of data.
- *Database as a service (DBaaS)* is the fourth most common service. It's literally a database hosted by a CSP.
- *Disaster recovery as a service (DRaaS)* is when cloud providers provide replication and full environmental hosting to enable recovery after a disaster.
- *Functions as a service (FaaS)* is a relatively new way of developing and architecting cloud apps. Development teams provide application code as a series of functions, and the cloud provider runs those functions.
- *Hardware as a service (HaaS)* is similar to IaaS but is more likely related specifically to data storage.
- *Information technology as a service (ITaaS)* aims to provide a client with all needed IT services, such as hardware, software, administration, and technical support in one package.
- *Monitoring as a service (MaaS)* tracks the performance of applications, devices, or services within a cloud network.

- *Network as a service (NaaS)* provides network infrastructure similar to IaaS.
- *Anything/everything as a service (XaaS)* is a combination of the services already discussed.

The level of service between the provider and the client is specified in the SLA, which is the contract between the two parties. It should be very clear which party has responsibility for specific elements, such as applications or hardware, should anything go awry. Figure 1.8 wraps up the service models discussion with an overview of how they all fit together. SaaS is the same as the Application layer shown in the figure.

**FIGURE 1.8** Reviewing the cloud service models




"Cloud computing" by Sam Johnston. Licensed under CC BY-SA 3.0 via Wikimedia Commons.


As a reminder, the three service models in the exam objectives are SaaS, IaaS, and PaaS.

## Deployment Models

The traditional type of cloud that usually comes to mind is a *public cloud*, like the ones operated by the third-party companies we mentioned earlier. These clouds offer the best in scalability, reliability, flexibility, geographical independence, and cost effectiveness. Whatever the client wants, the client gets. For example, if the client needs more resources, it simply scales up and uses more. Of course, the client will also pay more, but that's part of the deal.

Using the cloud is not restricted to big companies offering services over the Internet. Companies can purchase virtualization software to set up individual clouds within their own network. That type of setup is referred to as a *private cloud*. Running a private cloud pretty much eliminates many of the features that companies want from the cloud, such as rapid scalability and eliminating the need to purchase and manage computer assets. The big advantage, though, is that it allows the company to control all of its own security within the cloud environment.

Some clients choose to combine public and private clouds into a *hybrid cloud*. This gives the client the great features of a public cloud while simultaneously allowing for the storage of more sensitive information in the private cloud. It's the best of both worlds. This could be a good model in a situation where the company needs to keep tight control over some data—a database hosted in a private cloud, for example—but also wants to take advantage of productivity software such as Google Docs.

A fourth type of cloud is a *community cloud*. These are created when multiple organizations with common interests, such as schools or merging companies, combine to create a cloud. In a sense, it's like a public cloud but with better security. The clients know who the other clients are and, in theory, can trust them more than they could trust random people on the Internet. The economies of scale and flexibility won't be as great as with a public cloud, but that's the trade-off for better security.

With the exception of private clouds, all cloud types use the concept of *shared resources*. A pool of resources is purchased, and each participant in the cloud pays for a fraction of those resources. Those resources will most likely be external to the company using them, as opposed to internal resources that they would have if they managed a private cloud.

### Using a Multicloud Environment

Literature on cloud usage will sometimes refer to a *multicloud environment*. As the name implies, it's the implementation of multiple clouds for one organization. It's not the same thing as a hybrid cloud, which refers to multiple deployment models within the same cloud. Instead, it means using different clouds for different types of services.

For example, a company might use a separate cloud provider for its infrastructure (IaaS) and its software (SaaS). Or, perhaps a company is large enough that it prefers to use multiple providers for IaaS. The IaaS providers might each handle separate infrastructure workloads, or they could be used to load balance the entire workload between them.

An extreme version of multicloud is for a company to have two entirely separate but identical clouds. One cloud can serve as a backup and can spring into action in the event of a primary cloud failure. Of course, a solution such as this is very expensive, but it can be worth it depending on business needs.

Using multiple clouds provides flexibility to customize your network environment to meet the business needs. Downsides can include cost, added complexity for your internal cloud administrators, and occasional interoperability issues between cloud providers.


As a reminder, the three deployment models in the exam objectives are public, private, and hybrid.

If you are interested in learning more about cloud service or deployment models, Amazon has a great reference source at `https://aws.amazon.com/types-of-cloud-computing/`.

## Cloud Characteristics

Within the U.S. Department of Commerce, there is a nonregulatory agency named the National Institute of Standards and Technology (NIST). Its stated mission is to promote innovation and industrial competitiveness. Because it's nonregulatory, though, it doesn't have the power to enforce standards compliance. In 2011, NIST published de facto principles for cloud computing. These principles include three cloud service models and four deployment models, which you have already learned, and an official definition of cloud computing, as follows:

> "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

The definition calls out five key characteristics of cloud computing: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The CompTIA Cloud Essentials+ exam objectives list six characteristics, using slightly different terms than what NIST put forward. Here, we will examine the six CompTIA cloud characteristics while calling out how they relate to their NIST counterparts.

**Elastic** Cloud services must be *elastic*, meaning that they can grow or shrink as the client's needs change. Clouds can do this because, thanks to virtualization, they employ *resource pooling*. The provider's resources are seen as one large pool that can be divided up among clients as needed. Clients should be able to access additional resources as needed, even though the client may not be aware of where the resources are physically located. In most cases, clients can get more resources instantly (or at least very quickly), and that is

called *rapid elasticity*. For the client, this is a great feature because they can scale up without needing to purchase, install, and configure new hardware. Elasticity can also work backward; if fewer resources are required, the client may be able to scale down and pay less without needing to sell hardware. Typical pooled resources include network bandwidth, storage, processing power (CPU), and memory (RAM).

**Self-Service** From an end user's standpoint, *self-service* (often labeled as *on-demand self-service* by providers) is one of the cloud's best features. This characteristic allows users to access additional resources (such as storage, network, or compute) automatically, 24 hours a day, 7 days a week, 365 days a year, without requiring intervention from the service provider.

**Scalability** We've talked about the need for clouds to be elastic and allow clients to use more or fewer resources as needed. The ability to use more or fewer resources is called *scalability*. Most public clouds have a very high degree of scalability—it's highly unlikely that your company will cause AWS to run out of storage space, for example!

Scalability can also refer to location. Public clouds can in theory be accessed from anywhere in the world. They can scale geographically to suit any business situation.

**Broad Network Access** Resources need to be accessible over the network by different types of clients, such as workstations, laptops, and mobile phones, using common access software such as web browsers. This also applies to clients using different OSs, such as Windows, macOS, Linux, iOS, or Android, and is called *broad network access*. The ability for users to get the data they want, when they want, and how they want is also sometimes referred to as *ubiquitous access*.

**Pay-as-You-Go** Cloud providers track clients' usage and then charge them for services used, much as utility companies charge for electricity, gas, or water used at a residence. In some billing models, clients pay only for the services they use, which is called *pay-as-you-go*. Sometimes you will hear this referred to as *metered service* or *measured service*. This is as opposed to a model where a client signs a contract for a fixed resource and pays regardless of usage. For example, a client might reserve 20 terabytes of storage space, use only 10, but still pay for 20. Regardless of the type of contract, resource usage should be monitored by the provider and reported to the client in a transparent fashion. We'll talk about charging and reporting models more in Chapter 5, "Management and Technical Operations."

**Availability** Simply stated, *availability* means that cloud resources are accessible and responsive whenever a client needs them. We'll get into more detail on this concept later in this chapter, in the "High Availability" section.

**So Many Organizations and Standards!**

In this section we introduced you to NIST, which published de facto standards for cloud computing. In the acronym list for Cloud Essentials+, CompTIA lists other bodies and standards as well, such as ISO and ITIL. You might see even others such as ANSI and

PMI. We don't want to go down too deep a rabbit hole, considering that none of these is an exam objective, but here's a quick explanation of what each acronym stands for and what it means:

**American National Standards Institute (ANSI)** ANSI is a private nonprofit institute that helps develop and facilitate the use of voluntary consensus standards in the United States. These are usually referred to as "open" standards because of ANSI's collaborative and consensus-based development and approval process. NIST and ANSI have worked together on creating standards, but ANSI is not specifically involved with cloud technology.

**International Organization for Standardization (ISO)** The ISO is an international body that develops international standards, frequently in the technology, workplace safety and health, and environmental arenas. For example, ISO 9001 is an internationally recognized workplace quality management standard. ISO does not have any cloud-specific standards. (And yes, their name and acronym don't really match!)

**Information Technology Information Library (ITIL)** ITIL is a framework of best practices for aligning IT services with business needs and delivering those services. It was first developed in the 1980s by the British government's Central Computer and Telecommunications Agency (CCTA) and is periodically revised and updated. ITIL has recommendations on how to best use cloud services but did not participate in standards development.

**Project Management Institute (PMI)** PMI is a global nonprofit professional organization for project management. PMI is not focused on cloud standards development.


The six cloud characteristics found in the exam objectives are elastic, self-service, scalability, broad network access, pay-as-you-go, and availability.

## Shared Responsibility Model

An old network technician adage is, "A server is perfectly secure until you install a network card." While this is true, a server without a network connection won't be a very useful server. And since we're talking about cloud services, everything is Internet-based. You're probably well aware of the myriad security challenges the Internet presents. Yet it's a critical piece of infrastructure for most modern businesses.

When business usage of cloud services started to explode in the late 2000s, security was the biggest concern. As with any new technology, there was a lot of ambiguity over who owned what. In this case, most businesses assumed cloud providers would handle all security measures. After all, if a business had servers, infrastructure, and data of its own stored in an on-site data warehouse, it expected to manage all aspects of security. In a new model

where a third party managed the services, infrastructure, apps, and data storage, shouldn't that third party also be responsible for all security measures?

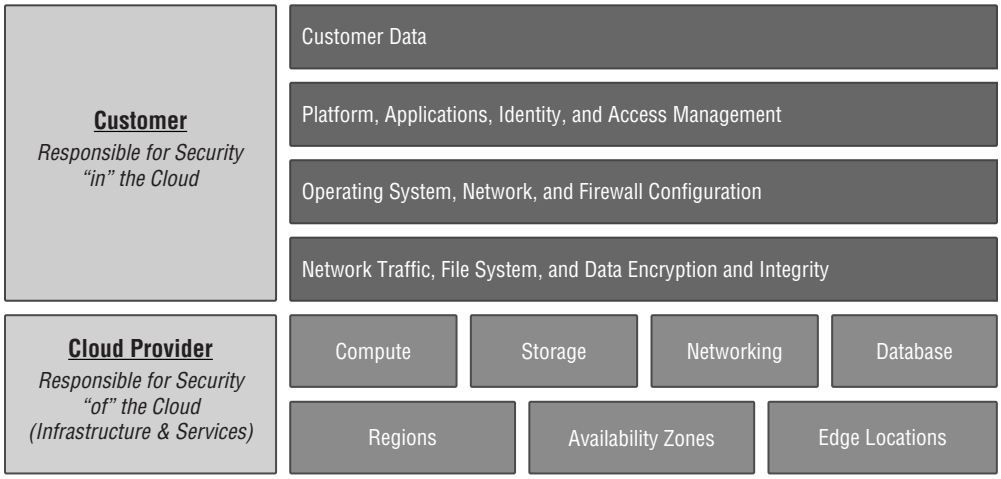
Cloud providers had a different perspective. They didn't feel that they should be held responsible for things outside of their control. They agreed that they needed to secure the cloud hardware and infrastructure, but could they really be held responsible for all of their clients' data? What if clients uploaded data with viruses or attempted to hack into data owned by another client?

The ambiguity and assumptions made by both parties unfortunately led to some security holes, and hackers were more than happy to exploit them. Out of this environment rose the concept of the shared responsibility model.

In the *shared responsibility model*, the CSP and the client share security duties, which improves the overall security of the system. The CSP is responsible for security "of" the cloud, while the client is responsible for security "in" the cloud. The CSP takes responsibility for what it can realistically control, and the rest is up to the client. Figure 1.9 shows how the shared responsibility model works.


Don't worry if you're not familiar with some of the cloud terms in Figure 1.9, such as *regions*, *availability zones*, and *edge locations*. You will learn about them in Chapter 5.

**FIGURE 1.9** The shared responsibility model




As you can see in Figure 1.9, the security responsibilities are pretty clearly defined. Even as a business client using a cloud, you need to have IT resources that are capable of managing the tasks needed to keep everything in the customer portion of the figure under control.

**Breaking the Shared Responsibility Model**

New technologies follow a predictable pattern. They are launched, and there's a breaking-in period during which users figure out the best way to apply them. The technology then gains popularity, creating more buzz and interest. Then, enterprising companies and individuals figure out how to expand the technology to make it more powerful or give it additional functionality. The cloud is no exception, and recent enhancements have stretched the shared responsibility model nearly to its breaking point.

Most of the security vectors shown in Figure 1.9 are clear-cut. It's the CSP's responsibility to secure the hardware and infrastructure. Patching an OS, installing a firewall, and securing data on the server falls to the client—easy enough. A few newer cloud implementations, such as containers and serverless (also known as *FaaS*), aren't as easily defined.

The in-the-weeds details of how containers and serverless work isn't too important to your studies at the moment—they are covered in Chapter 5. What is important to know is that each of these models shifts some of the traditional security responsibility from one partner to the other. Most CSPs are aware of situations where they need to take on more responsibility, which they do for FaaS, and have implemented appropriate security procedures. Containers can shift more responsibility to the client, and not all clients fully understand that.

The key takeaway from this story is to think of security holistically from end to end. Have all security needs mapped out, from hardware and infrastructure to software and apps to processes. Then partner closely with your CSP to ensure that all aspects of security are being covered so there are no gaps for hackers to exploit.


As a reminder, the shared responsibility model is in the exam objectives. Be sure to remember that the CSP is responsible for security of the cloud, and the client is responsible for security in the cloud!

# Exploring Cloud Design

A well-designed cloud solution can save a company money in several ways. First, it can eliminate the sunk cost of buying new hardware and software every few years, just to have it go obsolete. Second, it can enable users to have faster and better access to data, freeing their time to build the business. Third, a good cloud solution can help a company quickly recover from what would otherwise be a catastrophic data loss. In this section, you will learn about the last two, in the forms of redundancy and high availability and disaster recovery.

## Redundancy and High Availability

If someone at work calls you redundant, that's not typically meant as a compliment. It could mean that your job is not needed or isn't useful, or it could mean that the absence of the role wouldn't result in any loss. If two people have redundant jobs, it doesn't bode well for their employment prospects. Clearly, as employees we do our best to not be redundant!

In computer terms, though, *redundancy* is often a good thing. It means that there is some sort of device, system, or process that can take over in the event of a failure and keep the system running. Data redundancy, for example, can mean that critical data is accessible at any time from any location. Network redundancy can mean that a key server or the Internet is always available. Redundancy plans enable another key feature of clouds—high availability—which means that users have uninterrupted service and good responsiveness from cloud resources.

### Redundancy Plans

Building redundancy into your network architecture nearly always means spending more money versus not having redundancy. That negative will always present itself during budget discussions, but if your company will be hurt by a loss of access to resources or data, the positives will outweigh the negative.


Not all data or infrastructure components need redundancy. If you determine that you can live without a certain resource for a few hours or a few days, then redundancy might be too expensive.

When putting together your cloud needs, build in a redundancy plan that can help eliminate issues caused by a single point of failure. Then, work with your CSP to ensure that proper redundancy is built into the SLA. In most cases, you will need to balance the risk tolerance for a specific failure with the cost of implementing a redundant system. Redundancy plans should cover the following areas:

**Hardware** Most network administrators will understand the concept of hardware redundancy, usually referred to as *fault tolerance*. A cloud, of course, uses hardware—all computers do. As a species, humans haven't yet figured out how to create computing devices without a physical component.

Hardware redundancy solutions from a cloud provider should come with an uptime guarantee. That is, what percentage of the time will they guarantee data and services to be available? We'll look at a few examples of this later in the "High Availability" section.

**Network** For most companies, the computer network needs to be available at all times. One way to accomplish this is through network redundancy. Network redundancy is closely linked to hardware, but it's just more specific. Whereas hardware redundancy can refer to things such as storage, processors, memory, or network devices, network redundancy is pretty specific. Cloud providers typically differentiate the two by saying that *hardware redundancy* often refers to the servers themselves, whereas *network redundancy* typically refers specifically to network infrastructure such as switches and routers. Network redundancy can also involve guaranteeing a certain amount of network bandwidth.

**Geographic** Looking at hardware on an even bigger scale, geographic redundancy means that two or more physical locations have replicas of the same data. Those physical locations should be far enough apart that if one were to suffer a catastrophic natural disaster or failure, the other or others would be spared. Another term for this is *geo-redundancy*, which we'll talk about again in Chapter 5.


Geographically redundant locations can also help with load balancing for your cloud services. For example, having geo-redundant locations in the United States, Europe, and Asia can help spread the workload, since users from those regions can be directed to their regional site as their primary access point. Users in each region will have significantly better performance than if they accessed one centrally located site.

**Process** Business processes are frequently overlooked when it comes to redundancy, but they shouldn't be ignored. If a process fails, it could mean catastrophic results for a business. For example, assume that a company sells products online. The ordering system needs to interface with the billing system, which also interfaces with the fulfillment system. And perhaps the fulfillment system includes several subsystems—one that sends the order to the warehouse, one that contacts the delivery company, and another that emails the client updates on the order status. Most likely, there's an automated process that manages this workflow. If one link in the chain breaks, the entire workflow could be disrupted, and orders won't get filled. That means a loss of sales, which is never a good thing.

Map out critical business processes and understand which require high availability. Then figure out what the single points of failure are—they could be hardware, software, or network infrastructure—and ensure there's a plan in place to avoid issues.

**Software** All computers rely upon software, from OSs to applications to drivers that enable communication between the hardware and software. Software redundancy can take one of two forms—redundancy of the program or redundancy within the program.

The concept of redundancy at the program level means that you have a second program that can be run to perform the task in the event that the first one fails. At a very simplistic level, this could mean something like using Google Docs if Microsoft Word fails. Usually word processing isn't critical enough to need a backup, but you get the idea.

Redundancy within the program is a little trickier and is typically an issue only if your company has software developers writing code. It often means placing some sort of failsafe mechanism within the software, so if there's a critical error, the software can recover. It can also mean creating self-adaptive and self-checking programs. Self-adaptive programs are a basic implementation of artificial intelligence.

**Data** Last but certainly not least, there is data redundancy. Most often, this takes the form of data backups. Backups come in all shapes and sizes, from ones that are mirrored copies of data and instantly accessible to tapes or other media that needs to be restored before its contents are available.

Data loss can come in a variety of forms. The obvious example is hard drive failure, but that accounts for only a small percentage of overall data loss. Malware such as viruses and ransomware, inadvertent deletion, and malicious acts by hackers, employees, or contractors are much bigger problems.

Implementing some sort of data backup as part of the redundancy plan (or disaster recovery plan) may be legally required for a company, depending on its industry. Even if not, it's the most important thing a company can do. While that might sound like hyperbole, it's absolutely not. Boston Computing has found that 60 percent of companies that experience a data loss without a backup shut down within six months. With cloud computing, data backups and redundancy are incredibly easy to implement and generally not very expensive. It would be foolish to ignore them.


If you are implementing data backups, be sure to secure the data in those backups via encryption!

## High Availability

A great aspect of cloud services is that they provide users with *high availability*, meaning that users have uninterrupted service and responsive access to services. When we say uninterrupted, though, we should probably caveat that by saying it's *mostly* uninterrupted. The level of uptime guaranteed by the CSP will be specified in the SLA.

Service availability is measured in terms of "nines," or how many nines of uptime the provider guarantees. For example, "three nines" means that the service will be available 99.9 percent of the time, whereas "four nines" means it will be up 99.99 percent of the time. More nines means more money, and different aspects of your service contract might require different levels of uptime. For example, a critical medical records database might need more guaranteed uptime than would a word processing application. As we've said before, the level of service you should get depends on how much risk your company is willing to take on and the trade-off with cost. Table 1.1 shows how much downtime is acceptable based on the number of nines of guaranteed uptime.

**TABLE 1.1** Availability Downtime

| Availability | Downtime per year | Downtime per day |
|---|---|---|
| Three nines (99.9%) | 8.77 hours | 1.44 minutes |
| Four nines (99.99%) | 52.6 minutes | 8.64 seconds |
| Five nines (99.999%) | 5.26 minutes | 864 milliseconds |
| Six nines (99.9999%) | 31.56 seconds | 86.4 milliseconds |

Guaranteeing that services will be available with the possible exception of less than one second per day seems pretty impressive, as is the case with five nines. You might see other combinations too, such as "four nines five," which translates into 99.995 percent availability, or no more than 4.32 seconds of downtime per day. The majority of CSPs will provide at least three nines or three nines five.

## Disaster Recovery

Having great plans for redundancy and high availability can help avoid service outages and keep your business running smoothly. And with those plans in place, you might think that you've covered your bases in the case of a system failure. That's not true though—unforeseen things can still happen to take out a critical system, application, or data set. There's one more aspect of cloud design to consider, and that's disaster recovery.

*Disaster recovery* is defined as the ability to become operational after a disaster. The disaster can be catastrophic hardware or software failures, hackers or other cyber criminals, human error, or forces of nature that cripple one or more computer systems. Every organization should have a disaster recovery plan outlining the steps to take to regain operational status in the event of a disaster.

Disaster recovery plans need to account for all aspects of corporate network infrastructure, including hardware, software, network infrastructure, power, and of course data. The plans should also specify what employees should do in the event of a disaster. For example, if a tornado or flood cripples the corporate headquarters overnight, what should employees do the next workday?

The good news is, if you are buying cloud services from a CSP, they can help with disaster planning. And depending on how you set up your SLA, the disaster recovery and response could fall on them, not on you. Policies and procedures for recovery should be clearly agreed to between you and the CSP and spelled out in the SLA—assuming they are going to take care of everything could be a tragic mistake. Two recovery objectives that need to be defined in a disaster recovery plan are the recovery point objective and recovery time objective.

### Recovery Point Objective

Imagine a scenario in which your company gets hit by a catastrophic natural disaster. The company is essentially down, and access to critical data is gone. The good news is nobody was hurt, so senior management has shifted to voicing concerns about how they will recover the data. Fortunately, you have a disaster plan with your CSP, and they've been backing up your data per the SLA. Now the question is, how old is the data that they will restore?

This is where the *recovery point objective (RPO)* becomes critical. The RPO defines the maximum age of files that must be recovered from backups in order to restore normal operations. In other words, how old can restored data be and still be useful to use to run the business? The client defines the RPO (and RTO, which we discuss next), and they get added to the SLA.

Different applications and data sets can and should have different RPOs, depending on their criticality and how often they change. OS core files don't change often, and most likely you have the ability to restore an OS and necessary patches from installation media or online. Therefore, an RPO for these types of files isn't needed. If it's a product list that changes once per month, then the RPO should be one month. If it's something like medical data or a financial transactions database, or other high-importance, potentially fast-changing data, the RPO needs to be close to zero.

Catalog the different types of data your company has, and then decide what the maximum age for "good enough" data is. That will help you determine your RPOs.

## Recovery Time Objective

Your company has just been hacked and the public website, network infrastructure, and data systems that are hosted in the cloud have effectively been taken down by the attack. Everyone is panicking because every minute of downtime costs the company money. How long will it take to get everything operational again? Never mind the question of how do you prevent that attack from happening again—that's important too!

The maximum amount of time a system can be offline in the event of a disaster is called the *recovery time objective (RTO)*. It defines how long the CSP has to get everything operational, including network access and data restoration.

Much like RPOs, RTOs can and should vary by system. If your company has a website that does not handle any e-commerce transactions, then a failure might be annoying but not essential to business operations. The RTO for that can be a little longer than for a database that hosts the online ordering system, for example.

### Assumptions Are Dangerous

Earlier we discussed the shared responsibility model, where both the client and the CSP have responsibility to secure cloud-based systems. Recall that the client is responsible for security *in* the cloud, and the CSP is responsible for security *of* the cloud. Even though there's no official term for it, think of redundancy and disaster recovery in the same way. Don't make assumptions about what the CSP is going to provide in those areas.

CSPs do have redundancy and data recovery services available, but they must be agreed to in the SLA. If they are not specified, then they are not covered. This includes absolutely essential services such as creating and testing data backups. Again, don't assume—that could be a dangerous mistake if you suffer a data loss and have no recovery options!


There are four aspects of cloud design listed in the exam objectives: redundancy, high availability, disaster recovery, and recovery objectives. The two recovery objectives are RPO and RTO.

# Summary

We began this first chapter by introducing you to core cloud concepts. Before getting into cloud concepts, though, you learned about virtualization, which, along with the Internet, enables cloud providers to provide scaled services to multiple clients at the same time. Virtualization is made possible through the use of software called hypervisors. Cloud services benefit client companies by allowing them to forego investing in expensive hardware, and they benefit cloud providers by generating economies of scale.

There are three primary service models in cloud computing: software as a service (SaaS), infrastructure as a service (IaaS), and platform as a service (PaaS). You learned about all three, with IaaS being the lowest level of service, PaaS providing more, and SaaS being the highest level of service. There are also several more types of services that CSPs will offer, such as database as a service (DBaaS) and functions as a service (FaaS), but all are offshoots of the three we previously mentioned.

Next, we discussed three deployment models: public, private, and hybrid. Public models are quite common, with dozens of providers in the marketplace. The biggest two for businesses are AWS and Microsoft Azure. Google Cloud is also widely used. Each has pros and cons and different pricing and service levels. Private clouds enable more security and control but require the company to buy all of the necessary hardware to support the cloud. Hybrid clouds are an implementation of public and private at the same time.

Clouds have characteristics that make them attractive to client companies. Those are elasticity, self-service, scalability, broad network access, pay-as-you-go models, and high availability. Cloud security is enabled through the shared responsibility model. That means cloud suppliers are responsible for security of the cloud, whereas clients are responsible for security within the cloud.

Finally, the chapter finished with a discussion of important aspects of cloud design, including redundancy, high availability, disaster recovery, and recovery objectives. Redundancy is like fault tolerance, where systems have backups in case something fails. Redundancy is one way to ensure high availability, which means that clients have uninterrupted access to resources. A disaster recovery plan is important for any network, whether it's internal or hosted in the cloud. Two important disaster recovery concepts to understand are the recovery point objective and recovery time objective.

# Exam Essentials

**Know what SaaS is.** Software as a service means that the CSP provides software for a client as well as the underlying hardware needed to run the software. Popular examples include Microsoft Office 365 and Google Docs.

**Know what IaaS is.** In an infrastructure as a service setup, the CSP provides hardware infrastructure. This often means compute, storage, and network capabilities.

**Know what PaaS is.** In platform as a service, the CSP provides both hardware and a development platform. That includes software development tools such as runtime environments. This lets the client focus on software development while the CSP manages the hardware and development environment.

**Know what the three cloud deployment models are.** The three deployment models are public, private, and hybrid. You might also see literature that refers to multicloud environments. Multicloud refers to having multiple clouds doing the same or specialized tasks, and not to the deployment model.

**Be able to list pros and cons for each of the three deployment models.** A public cloud means that the client does not need to invest in hardware that will go obsolete in a few years, but downsides can be lack of control and security. Private clouds allow the company to retain security and resource control but usually cost more because of the hardware replacement life cycle. Hybrid clouds are the best and worst of both worlds, because they combine features of public and private clouds.

**Know the six characteristics of clouds.** Clouds are elastic, provide self-service, offer scalability, have broad network access, may offer pay-as-you-go, and have high availability.

**Be able to explain what it means for a cloud to be elastic.** Elasticity within a cloud means that the cloud's resources can grow or shrink as the client's needs change. Clouds can do this because they utilize resource pooling. Sometimes you will hear this called rapid elasticity.

**Be able to describe what self-service means.** In a cloud environment, much is automated. This includes allowing clients to request or access additional features without requiring intervention from the service provider.

**Be able to explain what scalability in the cloud is.** Scalability is closely linked with elasticity. It just means clients have the ability to use more or fewer resources as needed.

**Be able to describe what broad network access means.** Cloud clients can access resources regardless of the hardware or OS they run. This can mean access for desktops, laptops, and mobile devices running Windows, macOS, or mobile OSs like iOS or Android.

**Be able to explain what pay-as-you-go means.** Cloud providers track client usage and charge them for services used, much as public utilities do. In some payment models, clients only pay for the services they use, which is called pay-as-you-go.

**Be able to describe what availability means in the cloud.** It means that resources are available whenever the client needs them. Most commonly, you will hear this referred to as high availability. The goal for the CSP is to provide uninterrupted service.

**Understand how the shared responsibility model works.** The shared responsibility model pertains to cloud security. The CSP is responsible for security of the cloud, and the client is responsible for security in the cloud.

**Know how redundancy works in a cloud environment.** Cloud providers can provide redundant systems, which is like fault tolerance. That means if one component or system fails, there is no loss of functionality because the redundant system takes over. Redundancy can apply to hardware, software, network infrastructure, geography, processes, or data.

**Understand what disaster recovery is.** Disaster recovery is the ability to become operational after a disaster happens. Disasters can include natural disasters, cyberattacks, hardware or software failures, or human error.

**Know what an RPO is.** A recovery point objective specifies how old data can be to still be considered current or useful to restore normal operations.

**Know what an RTO is.** A recovery time objective defines how long the system or services can be down before being restored, in the event of a disaster or outage.

# Written Lab

Fill in the blanks for the questions provided in the written lab. You can find the answers to the written labs in Appendix A.

1. The ____ specifies the oldest that data can be to be acceptable to restore for business use.
2. Microsoft Office 365 is an example of the _____ cloud service model.
3. Cloud security responsibilities are defined by the _____.
4. Cloud users need to access resources from PCs and mobile devices. This is called _____.
5. Cloud clients can get more compute power without CSP intervention based on the _____ characteristic.
6. The _____ cloud deployment model places all security responsibilities on the company that owns it.
7. Thanks to the concept of pooled resources, cloud resources are said to be _____.
8. An SLA that specifies five nines is referring to _____.
9. One month, a cloud client uses twice as many resources as they normally do. The next month, they use their normal amount. They pay for only what they used. This is called _____.
10. The two recovery objectives that should be in a disaster recovery plan are ____ and ____.

# Review Questions

The answers to the chapter review questions can be found in Appendix B.

1. The CIO wants to reduce the cost of network ownership for the company, specifically for app licenses. Which service model should you recommend to her?
   - **A.** XaaS
   - **B.** SaaS
   - **C.** PaaS
   - **D.** IaaS

2. Which of the following is a potential outcome of implementing a multicloud environment for your company?
   - **A.** Combined features of public and private clouds
   - **B.** Decreased complexity for internal cloud administrators
   - **C.** Increased security for cloud-based resources
   - **D.** Increased flexibility of using separate SaaS and IaaS vendors

3. You are buying new cloud services. The internal network administration team is worried about cloud access from different OSs, such as Windows, macOS, and Android. What should you tell them?
   - **A.** Resources will be available to only Windows and macOS clients via broad network access.
   - **B.** Resources will be available to only Windows and macOS clients via self-service.
   - **C.** Resources will be available to all client OSs via broad network access.
   - **D.** Cloud services are usable only by Android and iOS mobile devices.
   - **E.** Resources will be available to all client OSs via self-service.

4. The company CIO asks you to ensure that the new cloud solution provides fault tolerance. Which aspect of cloud design does this refer to?
   - **A.** High availability
   - **B.** Shared responsibility
   - **C.** Disaster recovery
   - **D.** Redundancy

5. Which of the following are examples of IaaS that a cloud provider might offer? (Choose two.)
   - **A.** Compute
   - **B.** Applications
   - **C.** Database
   - **D.** Storage

6. Which of the following should you ask for in an SLA if you want to ensure the highest availability of resources for cloud users?
   - **A.** Four nines five
   - **B.** Four nines
   - **C.** Three nines five
   - **D.** Three nines

7. When shopping for public cloud services, the CSP tells you that if your company needs more or fewer resources, the CSP can instantly accommodate that. What cloud characteristic does this refer to?
   - **A.** Elasticity
   - **B.** Self-service
   - **C.** Broad network access
   - **D.** Availability

8. Your CSP makes daily backups of important files and hourly backups of an essential database, which will be used to restore the data if needed. Which aspect of cloud design does this represent?
   - **A.** Redundancy
   - **B.** High availability
   - **C.** Disaster recovery
   - **D.** RTO

9. Which of the following best describes the purpose of the shared responsibility model?
   - **A.** The CSP and client share responsibility for cloud security.
   - **B.** The CSP and client share responsibility for cloud costs.
   - **C.** Clients in a cloud share security responsibility with each other.
   - **D.** Clients in a cloud share costs with each other.

10. Which of the following cloud service models best supports a software development team with members in the United States and Japan?
    - **A.** IaaS
    - **B.** SaaS
    - **C.** DBaaS
    - **D.** PaaS

11. You are negotiating an SLA with a CSP. Which two things need to be included as part of the recovery objectives?
    - **A.** Recovery process objective
    - **B.** Recovery point objective
    - **C.** Recovery time objective
    - **D.** Recovery cost objective

12. In the cloud security model, who is responsible for securing access management and firewall configurations?
    - **A.** CSP
    - **B.** Client
    - **C.** CSP and client
    - **D.** All clients share this responsibility.

13. Which of the following cloud deployment models offers the best scalability and cost effectiveness?
    - **A.** Public
    - **B.** Private
    - **C.** Community
    - **D.** Hybrid

14. Which of the following best describes an RTO?
    - **A.** The oldest that data can be to be useful when it's restored
    - **B.** The amount of time services can be down after a disaster before being restored
    - **C.** The cost associated with restoring services after a disaster
    - **D.** The chain of command for notifications when a disaster occurs within the cloud

15. Which of the following is NOT a common characteristic of public clouds?
    - **A.** Self-service
    - **B.** Pay-as-you-go
    - **C.** Higher cost
    - **D.** Availability

16. Your company uses a financial transactions database that updates frequently. If a natural disaster occurred, any data backup older than one hour would not be useful to the company. To ensure that backups are always more current, what needs to be specified in the disaster recovery plan?
    - **A.** RPO
    - **B.** RTO
    - **C.** TSO
    - **D.** SLA

17. Your network hardware is outdated and needs to be replaced. The CIO suggests using the most cost-effective cloud solution. However, he insists that the company database remain 100 percent controlled by your company. Which solution is the best choice?
    - **A.** Public
    - **B.** Private
    - **C.** Community
    - **D.** Hybrid

18. Your data science team runs complex simulations that require more compute resources. They do about one simulation per month, and it takes one to two days. You want to ensure that the team has available compute resources but that you only pay for the resources when they are in use. Which cloud characteristic do you need?
    - **A.** Scalability
    - **B.** Availability
    - **C.** Pay-as-you-go
    - **D.** Elasticity

19. Which of the following is NOT a type of redundancy typically available in public cloud solutions?
    - **A.** Process
    - **B.** OS
    - **C.** Hardware
    - **D.** Data

20. Several universities want to band together to get cloud service. They want to share scholarly research but also take advantage of common cloud-based productivity apps. Which solution is the best option for their needs?
    - **A.** Public
    - **B.** Private
    - **C.** Hybrid
    - **D.** Community