Chapter 1

Understanding the Foundations of AWS Architecture

This chapter covers the following topics:

- **Essential Characteristics of AWS Cloud Computing**
- **AWS Cloud Computing and NIST**
- **Moving to AWS**
- **Operational Benefits of AWS**
- **Cloud Provider Responsibilities**
- **Security at AWS**
- **Migrating Applications**
- **The AWS Well-Architected Framework**
- **AWS Services Cheat Sheet**

The AWS Certified Solutions Architect – Associate (SAA-C03) exam that we are discussing in this book measures your technical competence in architecting workloads to run successfully in the Amazon Web Services (AWS) cloud. For any of their associate certification exams, AWS does not expect you to be an expert in every single cloud service, as that is an impossible task. However, AWS does expect you to be able to display a high level of competence about how to architect (design, deploy, monitor, and manage) workloads running on AWS cloud architecture based on the exam domains of knowledge. You can find the SAA-C03 exam guide here: **https://d1.awsstatic.com/training-and-certification/docs-sa-assoc/AWS-Certified-Solutions-Architect-Associate_Exam-Guide_C03.pdf**. The SAA-C03 exam guide lists the AWS services that could be tested on the exam, and what AWS services are not covered.

The goal of writing this book is to include enough technical details for all readers to absorb and pass the AWS Certified Solutions Architect – Associate (SAA-C03) exam. The following list should help you to gauge whether you should read this entire chapter or skim through the topics:

- If you are coming from a technical background but don't know anything about the AWS cloud, start with this first chapter and read it carefully.
- If you have a background working in the AWS cloud but this is your first certification attempt, you might not need to read the entire chapter, but you should review the first chapter's content, and study the final section, "AWS Services Cheat Sheet."
- If you already are certified as an AWS Certified Solutions Architect – Associate and it's time to re-certify, you might not need to read this chapter, but you should study the final section, "AWS Services Cheat Sheet," to ensure that you're up to speed on the latest AWS services covered on the exam.

And let's be clear, the goal of this book is to help you pass the AWS Certified Solutions Architect – Associate exam. If you ace the exam, great! However, passing the exam should be your overall goal. You need to get roughly 72% of the exam questions right to pass the exam; Amazon is not clear as to the exact percentage for passing the exam but it's in this range. The AWS SAA-C03 exam is 65 multiple choice questions. However, it's very important to understand that 15 of the 65 exam questions are beta questions that don't count! Therefore, there are 50 questions you must answer successfully. Answering approximately 37 questions correctly out of the 50 questions that count will achieve your goal of becoming an AWS Certified Solutions Architect – Associate.

The SAA-C03 exam is marked using what is defined as *scaled scoring*. The questions that you are presented on your exam most likely will not be the same as those presented to other exam candidates; the difficulty of each exam question is weighted to ensure the total knowledge level of each exam as a whole is maintained. Additional details on how to prepare to take the exam are fully covered in the last chapter of this book, **Chapter 16**, "**Final Preparation.**"

The following list of tasks will also help you greatly in the goal of becoming certified:

- **Read the FAQs:** Each AWS cloud service has a frequently asked questions (FAQs) summary that summarizes the service and its highlights. When learning about an AWS service, always start with the FAQ—you won't be disappointed. And be sure to take notes as you learn.

- **Read the AWS Well-Architected Framework PDFs:** The exam is based on the AWS Well-Architected Framework. Reading the PDF of each pillar is a great study aid for understanding the mindset of the exam questions, and will also prepare you to be a great AWS consultant/cloud architect. Make sure to review the Security Pillar, Reliability Pillar, Performance Efficiency Pillar, and the Cost Optimization Pillar. See https://aws.amazon.com/architecture/well-architected/.
- **Sign up for a free AWS cloud account:** This is the best method to practice hands-on tasks for the exam. Create multiple AWS accounts; you are not limited to one free AWS account, but a different e-mail address must be used as the root login for each AWS account that is created.
- **Complete AWS Well-Architected Labs:** Complete as many of the labs as possible that relate to the AWS Certified Solutions Architect – Associate exam topics. The labs are foundational (100), intermediate (200), and advanced (300), as partially shown in Figure 1-1 for the Security category. See https://wellarchitectedlabs.com/.

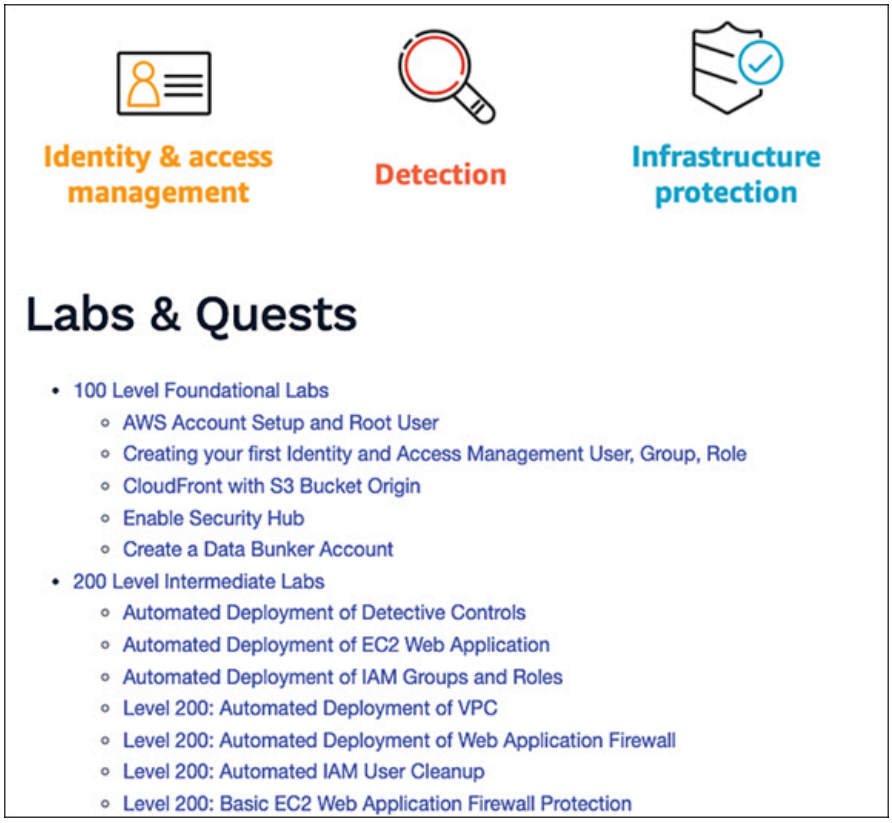

Figure 1-1 AWS Well-Architected Framework Hands-on Labs

- **Use the AWS Well-Architected Tool:** The AWS Well-Architected Tool is a self-paced utility that consists of Well-Architected Framework questions from each pillar to make you consider which best practices and procedures should be considered when hosting your workloads at

AWS. This is a great study aid for the exam, available at https://www.wellarchitectedlabs.com/well-architectedtool/.

- **Complete the AWS security workshops:** AWS offers a variety of security workshops that will help you understand AWS security best practices; see https://awssecworkshops.com/.
- **Answer as many sample exam questions as you can:** Included in this book is a test engine with hundreds of test questions. The hardest part of preparing to take the exam is getting used to answering multiple-choice test questions. The more practice you have, the better you will be prepared. AWS also has some sample questions for the SAA-C03 exam here:
  https://d1.awsstatic.com/training-and-certification/docs-sa-assoc/AWS-Certified-Solutions-Architect-Associate_Sample-Questions.pdf
  and here:
  https://explore.skillbuilder.aws/learn/course/external/view/elearning/13266/aw certified-solutions-architect-associate-official-practice-question-set-saa-c03-english?saa=sec&sec=prep
- **Browse the AWS Architecture Center:** The AWS Architecture Center (https://aws.amazon.com/architecture/) has many examples of how to deploy reference architecture for analytics, compute and HPC deployments, and databases, to name just a few. Walking through the step-by-step notes provides a great overview of the associated AWS services and can be helpful in visualizing how AWS architecture is designed and deployed.

Essential Characteristics of AWS Cloud Computing

In 2021, CEO Andy Jassy estimated that the cloud was currently less than 5% of global IT spending, which suggests that moving workloads to the cloud for many companies is really just beginning. The public cloud providers AWS and Microsoft Azure have been established for well over a decade and have strong infrastructure as a service (IaaS) and platform as a service (PaaS) offerings available around the world. Google Cloud Platform (GCP), Oracle Cloud, and IBM Cloud are also viable alternatives. Figure 1-2 shows the Gartner Magic Quadrant for Cloud Infrastructure and Platform Services (see https://www.gartner.com/en/research/methodologies/magic-quadrants-research), which indicates the current favorite cloud technology

providers companies can choose to align with. In the Leaders quadrant, Amazon Web Services leads, followed closely by Microsoft and then Google. Alibaba Cloud aligns with the Visionaries quadrant, and Oracle, Tencent Cloud, and IBM currently occupy the Niche Players quadrant.

When I started my career as a computer technician in the 1990s, most corporations that I supported used several computer-based services running on mainframes that were not located on premises. Accounting services were accessed through a fast (at the time) 1200-baud modem that was connected using one of those green-screened digital terminals. The serial cable, threaded through the drop ceiling to connect the terminal, was strong enough to pull a car.

Today we rely more and more on one or more public cloud providers for hosting many types of workloads on an ever-increasing collection of very specialized data centers and cloud services. There is no hardware ownership, the cloud provider owns the services, and customers rent cloud services as required.

Figure 1-2 Gartner's Magic Quadrant of Top Public Cloud Providers (https://www.gartner.com/en/research/methodologies/magic-quadrants-research)[1]

[1] Gartner does not endorse any vendor, product, or service depicted in its research publications and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

You might think that the public cloud only offers virtual resources, but the AWS cloud and others *can* provide bare-metal servers if requested. AWS will happily host your applications and databases on bare-metal servers hosted at AWS, or in your own data centers. Of course, more commonly, AWS offers you a wide variety of virtual servers in many different sizes and designs. AWS is also quite happy if you continue to operate your

on-premises data centers and coexist with cloud resources and services operating at AWS. AWS also offers AWS Outposts, which enables customers to run an ever-increasing number of AWS cloud services on premises. Microsoft Azure will offer to sell you a copy of its complete Azure cloud operating system, called Azure Stack, installed on servers in your data centers. It's getting harder to define the public cloud these days.

Applications that are hosted in the public cloud leverage virtual server, network, and storage resources combined with cloud services that provide monitoring, backup services, and more. Hardware devices, such as routers, switches, and storage arrays, have been replaced by AWS-managed cloud services built from the same virtual computers, storage, and networking components used by AWS themselves that are offered to each customer. This doesn't mean that companies aren't still using hardware devices on premises. However, it is possible to run hundreds or thousands of virtual machines in parallel, outperforming the functionality of a single hardware switch or router device. Most AWS cloud services are hosted on virtual machines called Amazon Elastic Cloud Compute (EC2) instances running in massive server farms powering the storage arrays, networking services, load-balancing, and auto-scaling services provided by AWS are part of Amazon Web Services (AWS). For example, AWS Config helps you manage compliance, and the AWS Backup service backs up AWS storage services.

AWS Cloud Computing and NIST

If you haven't heard of the National Institute of Standards and Technology (NIST), a branch of the U.S. government, you're not alone. Around 2010, NIST began documenting the emerging public cloud. After consulting the major cloud vendors, it released an initial report in June 2011, Special Publication 800-145, "The NIST Definition of Cloud Computing," defining the cloud services that were common across all public cloud vendors. The report's genius is that it defined in 2011 what the emerging public cloud actually became. NIST's cloud definitions have moved from mere definitions, to accepted standards that are followed by all of the public clouds we use today.

The five key NIST definitions of the public cloud have morphed into a definitive standard methodology of how cloud providers and thousands of customers operate in the public cloud. The report can be found here:

https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf. The five essential characteristics of the cloud model defined by NIST are

- On-demand Self-Service
- Broad Network Access
- Resource Pooling
- Rapid Elasticity
- Measured Service

The sections that follow describe these essential NIST characteristics.

On-Demand Self-Service

These days companies don't just *expect* cloud service to be delivered quickly; they *demand* it.

Every cloud provider, including AWS, offers a self-service management portal (see **Figure 1-3**). Request any cloud service, and in seconds, or minutes, it's available in your AWS account, ready to be configured or used. Gone are the days of requesting a virtual server via email and waiting several days until it's available. At AWS, a virtual server can be ordered and operational in under 5 minutes. Creating and using an Amazon Simple Storage Service (Amazon S3) bucket is possible within seconds. It is also possible to procure a software-defined network (called an Amazon Virtual Private Cloud) and have it operational in seconds. Using the AWS management console enables customers to order and configure many cloud services across many AWS regions. Any cloud service ordered is quickly delivered using automated procedures running in the background.

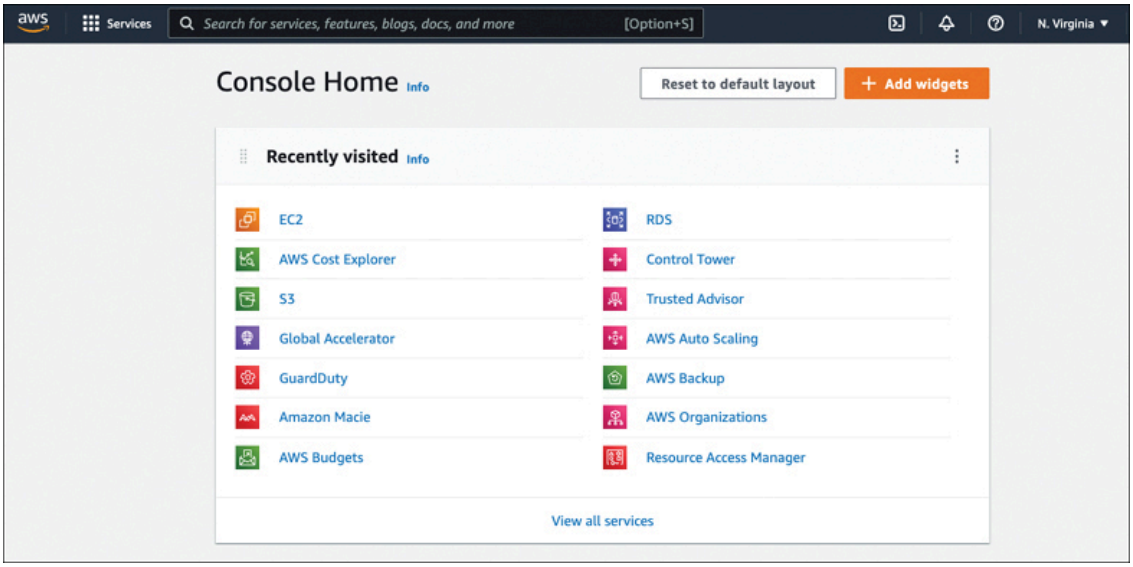

Figure 1-3 The AWS Management Console

Broad Network Access

Cloud services running at AWS can be accessed from anywhere there is an Internet connection, using just a web browser. AWS provides secure HTTPS endpoints to access every cloud service hosted at AWS. However, your company might not want or require what NIST defined as broad network access, which is public Internet network access to your workloads. Many companies that are moving to the AWS cloud have no interest in a publicly accessible software solution. They want their hosted cloud services to remain private, accessible only by their employees using private network connections. Each cloud customer ultimately defines their definition of broad network access: public Internet connections, private VPN or fiber connections, or both.

At AWS, applications and services can be made publicly available, or they can remain completely private. Virtual private network (VPN) connections from your place of work to AWS are commonplace. Customers can also order an AWS Direct Connect connection, a private fiber connection to AWS resources running at speeds up to 100 Gbps. Depending on the type of application you're hosting in the AWS cloud, high-speed network access may be essential.

It's also possible to administer AWS services from a smartphone by using an AWS app (see **Figure 1-4**). Certainly, accessing AWS from any device is possible.

Resource Pooling

Infrastructure resources for AWS cloud services are located across different geographical regions of the world in many data centers. A company running an on-premises private cloud will typically pool its virtual machines, memory, processing, and networking capabilities into one or two data centers offering a limited pool of compute and network resources.

AWS has clusters of data centers, stored in multiple availability zones (AZs) across each region, and each AZ has thousands of bare-metal servers and storage resources available and online, allowing customers to host their workloads with a high level of resiliency and availability. Without a massive pool of compute resources, AWS would not be able to allow customers to dynamically allocate compute resources to match their performance requirements and workload needs. Amazon S3 object storage is offered as unlimited; there is no defined maximum storage limit.

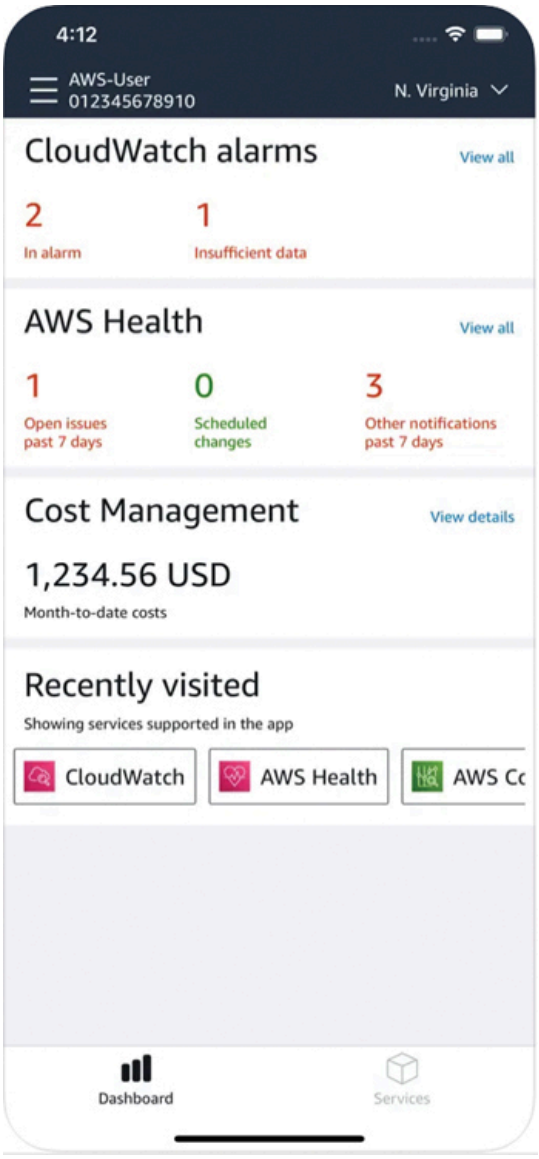

Figure 1-4 AWS Apps for Phone

Rapid Elasticity

Rapid elasticity in the public cloud is *the* key feature for hosted cloud applications. At AWS, compute and storage resources are defined as elastic. Workloads running in the AWS cloud for Amazon EC2 instances or Amazon Elastic Container Service (Amazon ECS) deployments have the capability to automatically scale using a scaling policy to dynamically resize an Auto Scaling group of web or application servers using several scaling policies, including target tracking (see **Figure 1-5**). In this example, EC2 Auto Scale will maintain CPU utilization of 65%; additional compute resources will be automatically added or removed to maintain the desired target value.

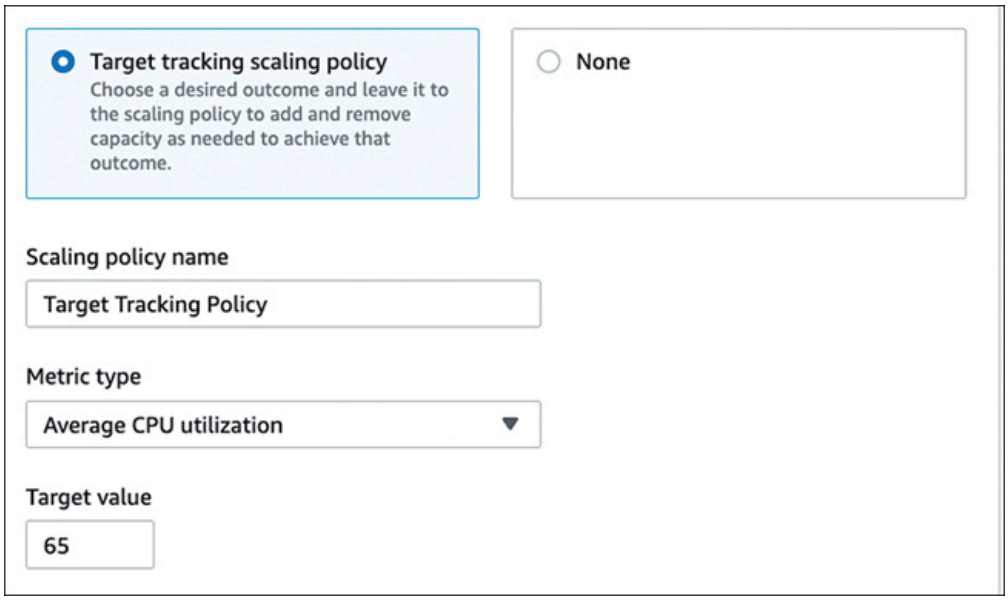

Figure 1-5 Workload Scaling Based on Demand

Elasticity—that is, dynamic scaling—is an automated solution scaling compute resources up or down in size based on workload needs. Administrators these days don't need to turn off virtual servers, add additional RAM, and turn the servers back on again; instead, they can deploy *horizontal scaling*—that is, automatically add or remove additional servers as required. AWS EC2 Auto Scaling is integrated with the Amazon CloudWatch monitoring service using metrics and event-driven alarms to dynamically increase or decrease compute resources as required.

Measured Service

In the AWS cloud, you are billed for only the services that you use or consume; this concept is referred to as a *measured service*. AWS charges can

be broken down into compute, storage, and data transfer charges. Packet flowing inbound (i.e., ingress to the AWS cloud) is usually free. By contrast, outbound packet flow (i.e., egress traffic across the Internet, a private network connection, or network replication traffic between a primary and alternate database server hosted on subnets in different availability zones) is charged an outbound data transfer fee. In the case of computer services such as AWS EC2 compute instances, charges are per hour for EC2 usage calculated by the second based on the size of the EC2 instance, operating system, and the AWS Region where the instance is launched. For storage services such as Amazon S3 storage or virtual hard drives (Amazon EBS), storage charges are per gigabyte used per month.

If a cloud service in your AWS account is on, charges will apply. Running hosted workloads in the AWS cloud requires a detailed understanding of how costs are charged; the management of costs at AWS is one of the most important tasks to understand and control. AWS has many useful tools to help you control your cloud costs, including the AWS Simple Monthly Calculator, AWS Cost Explorer, and AWS Budgets (see **Figure 1-6**).

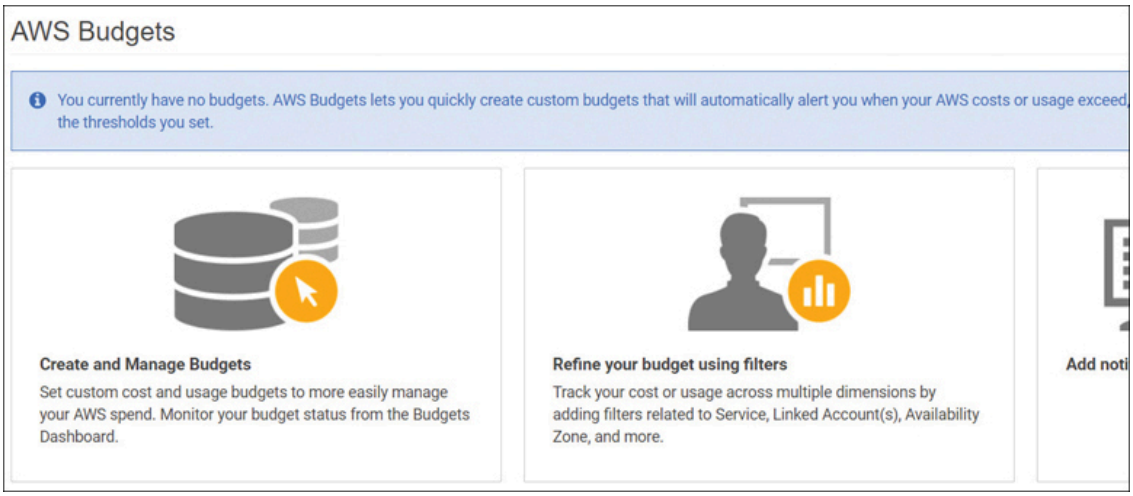

Figure 1-6 Using AWS Budgets to Track and Alert When Costs Are Over Budget

Being billed for consuming cloud services is a reality that we are all personally used to; for example, Netflix, Disney, and Dropbox are common services. However, billing at AWS is different from the flat per-month fees for personal software as a service (SaaS) services. Customers must understand and carefully monitor their compute, storage, and data transfer costs or else their monthly charges can become extremely expensive. For example, a load balancer can be ordered at AWS for approximately $18

per month. However, the data traffic transferred through the load balancer is also charged, so the overall monthly price could be substantial.

Moving to AWS

Once an organization has decided to move to the AWS cloud, countless moving parts begin to churn. People need to be trained, infrastructure changes must take place, developers need to develop applications with a different mindset, and administrators must get up to speed. Generally, people at companies beginning to utilize cloud services typically have several mindsets:

- **The corporate mentality:** You currently have data centers, infrastructure, and virtualized applications. Ever-increasing infrastructure and maintenance costs are driving you to look at what options are available in the AWS cloud. Your starting point could be to utilize the available IaaS offerings for servers, storage, monitoring, and networking services.
- **The born-in-the-cloud mentality:** You're a developer (or a nimble organization) with a great idea but not much startup funding. You also don't have a local data center, and want to get going as soon as possible. Your starting point could be to utilize the available IaaS offerings for servers, storage, monitoring, and networking, and the PaaS offerings, to speed up the development process.
- **The startup mentality:** You've just lost your job due to a merger or buyout and are determined to strike out on your own. Your brand-new company has no data center and lacks cash, but it has plenty of ideas. Your starting point will be the same as the born-in-the-cloud mentality example.

Each of these starting mindsets or outlooks will have differing points of view about how to migrate or design their cloud infrastructure and hosted applications. If you come from a corporate environment, you will probably expect the cloud provider to have a detailed service-level agreement (SLA) that you can change to match your needs. You will also probably have expectations about how much detail should be provided about the cloud provider's infrastructure and cloud services. AWS has service-level agreements for its cloud services and very detailed documentation for each hosted cloud service.

Note

AWS has options for developers who want to craft and deploy applications hosted at AWS. Visit https://aws.amazon.com/startups/ for further information about how you might be able to qualify for AWS Promotional Credit. There's a possibility of getting up to $15,000 in credits over 2 years, including AWS support and training.

Infrastructure as a Service (IaaS)

Many cloud services offered by AWS are defined as IaaS services, and are defined in this book as foundational services that are used by every customer (see Figure 1-7). Virtualized servers (Amazon EC2), container services (Amazon ECS), and database services (Amazon RDS) are hosted on a fast private software-defined network (SDN). Each customer's IaaS services are isolated from all other AWS customers by default. A robust security service named AWS Identity and Access Management (IAM) enables each customer to secure and control every ordered IaaS service as desired. A wide variety of supporting services, defined as Management and Governance services, also shown in Figure 1-7, provide monitoring (Amazon CloudWatch), audit services (AWS CloudTrail), scaling of compute resources (AWS Auto Scaling), governance (AWS Config), and event-driven automation (AWS Lambda).

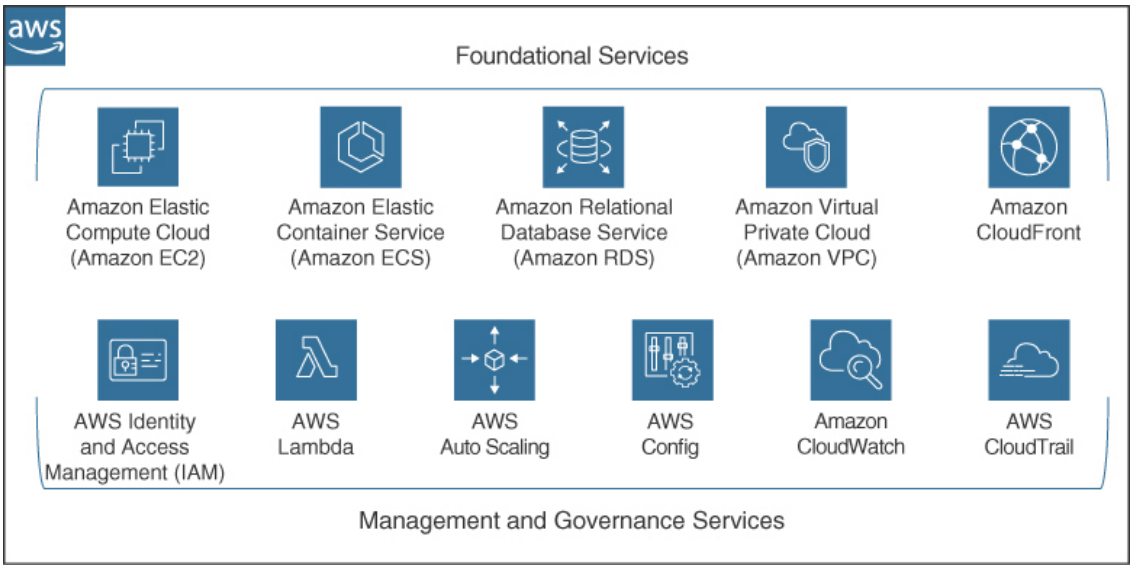

Figure 1-7 Infrastructure as a Service at AWS

Hosting compute workloads at AWS requires the creation of a network environment called Amazon Virtual Private Cloud (VPC) hosting web, ap-

plication, and database services on subnets. Customers have the flexibility to create whatever architectural stack is required at AWS, using the vast number of IaaS services and management services available. Many companies moving to AWS typically start with IaaS services, because the IaaS services at AWS closely mirror their current on-premises virtual environment.

Here are some examples of the essential cloud services at AWS:

- **Compute services:** The previously introduced Amazon EC2 is a cloud service that provides virtual servers (dedicated, multi-tenant, or bare-metal) in an ever-increasing variety of options. Amazon Elastic Container Service (Amazon ECS) supports Docker containers running at AWS, or on-premises using AWS Outpost deployments. Amazon Elastic Kubernetes Service (EKS) supports Kubernetes deployments at AWS or on-premises using AWS Outposts.
- **Storage services:** Amazon S3 is a cloud service that provides unlimited object storage in Amazon S3 buckets or archived storage in vaults. There are shared storage arrays: Amazon Elastic File System (Amazon EFS) for Linux, and Amazon FSx for Windows File Server for Microsoft Windows deployments, and virtual block storage volumes using the Amazon Elastic Block Store (Amazon EBS) service.
- **Database services:** AWS offers a fully managed database service called Amazon Relational Database Service (Amazon RDS). Choose from Amazon Aurora (with MySQL or PostgreSQL compatibility), MySQL, PostgreSQL, Oracle, and Microsoft SQL Server engines. Using Amazon RDS, AWS builds, hosts, maintains, backs up, and synchronizes HA pairs or clusters of primary/standby database servers, leaving customers the single task of managing their data records. Many other managed database services are also available at AWS, including Amazon DynamoDB, a NoSQL database; and Amazon ElastiCache, a managed in-memory caching service that supports Memcached and Redis deployments.
- **Automating AWS infrastructure:** AWS CloudFormation enables customers to automate the process of modeling and provisioning infrastructure stacks, complete with the required compute, storage, networks, load balancers, and third-party resources required for each workload. Template files are created using either JSON or YAML declarative code.

- **Auditing:** AWS CloudTrail is enabled in every AWS account, tracking and recording all application programming interface (API) calls and authentication calls. Customers can also configure AWS CloudTrail to store audit information in Amazon S3 Glacier archive forever.
- **Monitoring:** AWS CloudWatch is a powerful monitoring service with metrics for more than 70 AWS services that can be used to monitor resources and application operations using alarms to carry out automated actions when predetermined thresholds are breached.
- **VMware Cloud on AWS:** Many companies use VMware ESXi infrastructure for their on-premises application servers. Capital expenses and licensing costs are some of the biggest expenses incurred when running an ever-expanding on-premises private cloud. Virtualization was supposed to be the answer to controlling a company's infrastructure costs; however, the cost of hosting, running, and maintaining virtualization services became extremely high as deployments expand in size and complexity. Replacing on-premises VMware deployments with AWS-hosted virtualized servers running on AWS's hypervisor services removes a company's need for hypervisor administration expertise. Many applications used by corporations are also now widely available in the public cloud as hosted applications defined as a software as a service (SaaS) application. VMware ESXi is also available as VMware Cloud on AWS, using VMware's software-defined data center architecture running on AWS infrastructure.

Note

At AWS, infrastructure and platform services and resources are spread across the world in 31 different regions (2022), and additional regions are scheduled to be added. If you are in a large population center, the odds are that access to AWS cloud resources is close by. If AWS is not yet close by, you still might be able to connect using an edge location or a local point of presence connection. To review the current AWS infrastructure, visit **https://aws.amazon.com/about-aws/global-infrastructure/regions_az/**.

Platform as a Service (PaaS)

PaaS cloud providers enable your company's developers to create custom applications on a variety of popular development platforms, such as Java, PHP, and Python and Go. Your choice of language and development framework will determine the PaaS vendor you select. Using a PaaS provider means that developers don't have to manually build and manage the infrastructure components required for each workload; instead, the required infrastructure resources for each workload running in the development, testing, and production environments are created, hosted, and managed by the PaaS cloud provider. After an application has been developed and tested and is ready for production, end users can access the application using the application's public URL. In the background, the PaaS cloud provider hosts and scales the hosted SaaS workload based on demand. As the number of users using the workload changes, the infrastructure resources scale out or in as required. PaaS environments are installed on the IaaS resources of the PaaS cloud provider, as shown in Figure 1-8. In fact, IaaS is always behind all "as a service" monikers. Examples of PaaS providers include Google Cloud, Cloud Foundry, and Heroku.

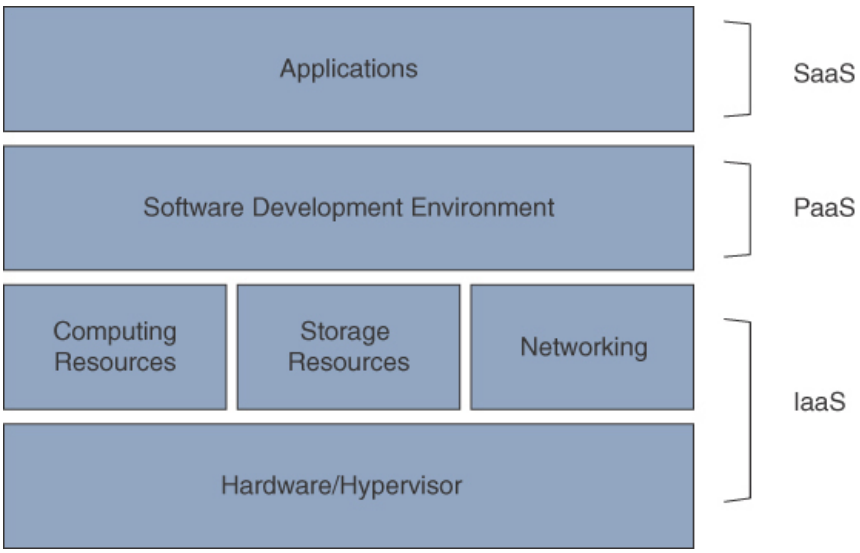

Figure 1-8 IaaS Hosting the PaaS Layer

The Cloud Foundry PaaS solution is offered for application development at IBM Cloud, running a customized version of the Cloud Foundry platform components. Developers can sign up and focus on writing applications. All application requests are handled by the PaaS layer interfacing with the IaaS layer, where the application's compute, storage, load-balancing, and scaling services operate.

Another popular solution for developing applications in the public cloud, Heroku, a container-based PaaS environment, enables developers to create and run applications using a variety of development platforms. Just as with IBM Cloud, once the workload is deployed into production, Heroku hosts, load-balances, and auto-scales each workload as required and sends each customer a bill for the infrastructure hosting costs used each month.

When developing applications at a PaaS provider, remember that programming languages change from time to time; therefore, the associated APIs offered by each cloud provider can change as well—and sometimes without much warning. Developers and companies must keep up to date with any ongoing changes or there can be issues when using a cloud-hosted PaaS development platform.

An additional reality is that one cloud provider's PaaS offering is not necessarily compatible with another cloud provider's PaaS offering. For example, Heroku and Microsoft Azure offer similar PaaS cloud services for developing applications, but internally, each cloud provider operates in a completely different fashion, with a completely different set of supporting APIs. There is no single standard for defining what PaaS must be. Compatibility issues can begin to reveal themselves at the lower levels of each vendor's proposed solution. RESTful interfaces, manifest file formats, framework configurations, external APIs, and component integration are not necessarily compatible across all cloud vendors.

AWS Elastic Beanstalk is Amazon's cloud service for deploying web applications. The service supports Java, .NET, PHP, Node.js, Python, Ruby, and Go. Code applications can be hosted on Apache, Nginx, Passenger, or IIS web servers, and containerized applications hosted on Docker.

Elastic Beanstalk acts as a managed service that frees customers from having to build out infrastructure configurations. It automatically handles scaling, load balancing, monitoring, capacity provisioning, and application updates. For additional details on AWS Elastic Beanstalk, see **Chapter 6**, "**Designing Resilient Architecture.**" AWS has also recently purchased Cloud9, an AWS-hosted integrated development environment (IDE) that supports more than 40 programming languages.

AWS has several cloud services to assist in developing applications, shown in Figure 1-9, including AWS CodeBuild, AWS CodeCommit, AWS Cloud9, and AWS CodeDeploy, that can be key components in your application deployment workflow at AWS.

Figure 1-9 Platform Options at AWS

Operational Benefits of AWS

Operating in the public AWS cloud has certain benefits provided by the previously discussed NIST five essential characteristics. Unlimited access to the many cloud services available at AWS may make it easier than expected to operate and manage workloads in the AWS cloud. Consider the following:

- **Servers:** Underutilized servers in your data center are expensive to run and maintain. Moving applications to the public cloud can reduce the size of your on-premises data center. When you no longer host as many physical servers, your total hosting costs (racking, powering, heating, and cooling) could be lower as well. You also don't have to pay for software licenses at the processer level because you're not responsible for running hypervisor services; that's now Amazon's job. You might think that moving to the AWS cloud means virtualized resources and only virtualization. However, with AWS, you can get an ever-increasing variety of EC2 instances, including dedicated virtual servers or bare-metal servers. Sizes range from a single-core CPU with 512 MB of RAM to hundreds of CPU cores and terabytes of RAM.
- **Storage:** Using cloud storage has huge benefits, including having unlimited amounts of storage. Amazon has shareable file solutions for both Linux and Windows Server workloads. Virtual hard disks are available using Amazon EBS to create the required volumes. Unlimited

storage and long-term archive storage are provided by Amazon S3 buckets and S3 Glacier archive storage.

- **Managed cloud services:** The AWS-managed cloud services, outlined in Table 1-1, may be able to replace or complement existing services and utilities currently used on premises after moving to the AWS cloud.

Table 1-1 Managed Services at AWS

IT Operation	On Premises	AWS Cloud
Monitoring	Nagios, SolarWinds	CloudWatch monitoring provides metrics for every AWS service with monitoring and logging data stored in unlimited Amazon S3 storage. Third-party monitoring solutions perform analysis of stored log data stored in S3 buckets.
Data backup	Backup tools such as Commvault and Veritas NetBackup	Many third-party vendors such as Veritas and Commvault (and many others) support the AWS cloud with compatible software appliances. AWS Storage Gateway can also be installed to move on-premises data records and virtual hard drive volumes to S3 storage while locally caching popular content. AWS

IT Operation	On Premises	AWS Cloud
		Backup enables you to centrally manage the backup of most data storage services at AWS to S3 storage.
Scale	Automation for increasing/decreasing the size of each virtual machine's RAM and CPU cores as required	Use EC2 Auto Scaling to automatically scale virtual machines (EC2 instances) or containers, dynamically increasing/decreasing the compute power required by applications.
Testing/development	Expensive provisioning of hardware for testing and development	Provisioning resources for short-term testing at AWS is incredibly inexpensive. Signing up for the AWS Free Tier enables customers to test a variety of AWS services for one year completely free of charge.
Identity management	Active Directory Domain Services for accessing corporate resources	It is possible to migrate or integrate on-premises Active Directory Domain Services to the AWS cloud using AWS Directory Services.

IT Operation	On Premises	AWS Cloud
		Deploy AWS single sign-on (SSO) services using IAM Identity Center to manage access to popular cloud business applications hosted by AWS or a third-party cloud provider.

Cloud Provider Responsibilities

AWS has published service-level agreements (SLAs) for most AWS cloud services. Each separate SLA lists the desired operational level that AWS will endeavor to meet or exceed. Current details on the SLAs offered by AWS can be viewed at **https://aws.amazon.com/legal/service-level-agreements**. AWS defines its commitments in each SLA about security, compliance, and overall operations. The challenge is to live up to these agreements when all services fail from time to time. Each cloud service SLA contains details about the acceptable outage times and the responsibility of the cloud provider when outages occur. Each SLA also contains statements about their level of responsibility for events outside the cloud provider's control. SLAs commonly use terms such as "best effort" and "commercially reasonable effort."

AWS is responsible for overall service operation and deployment, service orchestration and overall management of their cloud services, the security of the cloud components, and maintenance of each customer's privacy. A managed services SLA also spells out how a cloud consumer is to carry out business with the cloud provider. Each cloud consumer must fully understand what each cloud service offered provides—that is, exactly what the cloud service will, and will not, do.

Is it acceptable to expect AWS failures from time to time? It is a reality; everything does fail from time to time.

What happens when a key service or component of your workload hosted in the AWS cloud fails? Does a disaster occur, or is the failure manage-

able? When operating at AWS, customers must design each hosted workload to be able to continue operating as required when cloud services, or compute and storage failures occur. Designing high availability and failover for hosted workloads running at AWS is one of the key concepts of many of the questions on the AWS Certified Solutions Architect – Associate (SAA-C03) exam. Many questions will be based on the concepts of designing with a high availability, failover, and durability mindset. Customers must design workloads to meet the application requirements, considering that cloud services *do* fail from time to time.

All public cloud providers really have the same SLA summarized in nine short words when failures happen: "We are sorry; we will give you a credit." Here's another reality check: If your application is down, you might have to *prove* that it was actually down by providing network traces and appropriate documentation that leaves no doubt that it was down because of an AWS cloud issue.

Here's another further detail to be aware of: If you don't build redundancy into your workload design, don't bother asking for a credit. Application designs that have a single EC2 instance hosting a workload with no failover or high-availability design parameters have no cloud provider SLA protection. AWS expects customers to be serious about their application design. Each customer needs to carefully design, deploy, and maintain each hosted workload based on the business needs and requirements, ensuring that any high availability and failover requirements have been met.

Security at AWS

As you move to the AWS cloud, you need to consider a number of security factors, including the following:

- **Data security:** The reality is that your data is typically more secure and durable when stored in the public cloud than in on-premises physical servers due to the multiple physical copies of any data records stored in public cloud storage. All storage mediums at AWS can also be easily encrypted with the Advanced Encryption Standard (AES). Amazon EBS volumes—both boot and data volumes—can be encrypted at rest and in transit, using customer master keys provided by AWS or keys provided by the customer. Shared storage services such as

Amazon EFS and FSx for Windows File Server can also be encrypted at rest, as can all offered database engines. Amazon S3 buckets are encrypted with keys provided by the S3 service or the Key Management Service (KMS) shown in **Figure 1-10**. Data durability provides additional security as all data stored in the AWS cloud is stored in multiple physical locations. For example, each EBS volume has multiple copies replicated within the data center where they are created.
Amazon S3 objects are replicated across at least three separate availability zones within the selected AWS region, producing a very high level of durability.

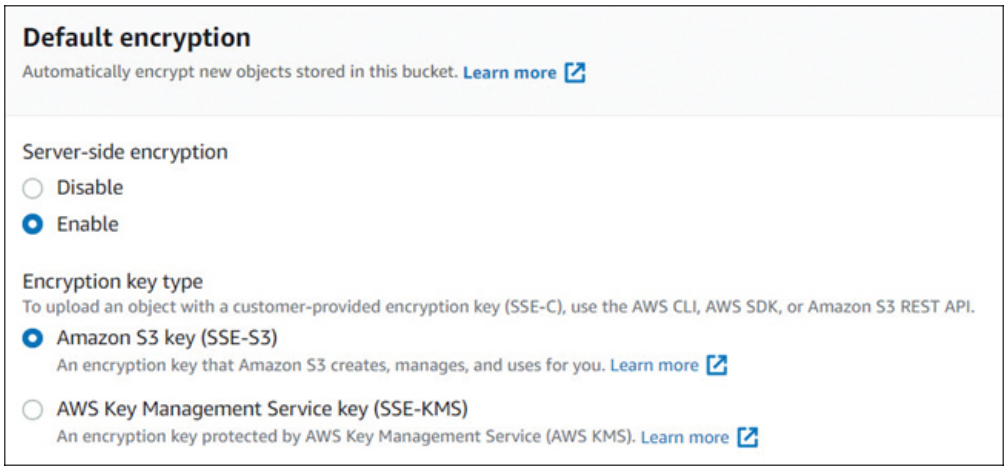

Figure 1-10 Encrypting S3 Buckets Using S3 Keys or AWS-KMS Managed Keys

- **Data privacy:** Amazon ensures that each AWS account's stored data records remain isolated from other AWS customers. In addition, data records are always created as a private resource. Each S3 bucket can be shared publicly; however, each customer assumes the responsibility when changing a private S3 bucket to be publicly accessible across the Internet.
- **Data control:** Customers are fully responsible for storing and retrieving their data records stored at AWS. It's the customer's responsibility to define the security and accessibility of all data records stored at AWS.
- **Security controls:** AWS Identity and Access management permission policies can be defined at a very granular level to control access to *all* resources at AWS. Customers can also enable multifactor authentication (MFA) as an additional security control for all IAM users authenticating to AWS, and on S3 buckets when deletion of data records is attempted. Resource policies defining the precise level of security and access can be directly attached to resources such as S3 buckets.

Network Security at AWS

At AWS, networking is managed at the subnet level, and subnets are first created as private subnets with no direct access to the outside world. Subnets that reside on your private networks at AWS are hosted in a virtual private cloud (VPC). Only by adding gateway services to a VPC and route table entries are subnets able to be accessed from either the Internet, a private VPN connection, or from an external network location. The following are examples of networking services and utilities at AWS that help control network traffic:

- Each subnet's ingress and egress traffic can be controlled by subnet firewalls called *network ACLs* that define separate stateless rules for inbound and outbound packet flow.
- Each EC2 instance hosted on a subnet is protected by a firewall called a *security group*, which defines what inbound traffic is allowed into the instance and where outbound traffic is allowed to flow to.
- VPCs can be further protected by deploying the AWS Network Firewall, providing control over all network traffic, such as blocking outbound Server Message Block (SMB) requests, bad URLs, and specific domain names.
- VPC flow logs can be enabled to capture network traffic for the entire VPC, for a single subnet, or for a network interface.

Application Security at AWS

Both web and application servers hosted at AWS are usually located on private subnets, which are not directly accessible from the Internet. Customers requesting access to the application will be directed by DNS services (Route 53) to the DNS name of the load balancer, which in turn directs incoming traffic from the public subnet to the targeted web servers hosted in private subnets.

For example, the end-to-end traffic pattern for a three-tier web application can be designed using many encryption/decryption points on its path from source to destination, as described in the list that follows and as shown in **Figure 1-11**:

- **AWS Web Application Firewall (WAF):** AWS WAF is a custom traffic filter that can be associated with an Application Load Balancer to protect against malicious traffic requests.

- **Application Load Balancer:** An application load balancer can accept encrypted HTTPS traffic on port 443 and provide Secure Sockets Layer/Transport Layer Security (SSL/TLS) decryption and, optionally, user authentication support.
- **EC2 instance hosting a web application:** EBS boot and data volumes can be encrypted using the AWS KMS service.
- **EC2 instance hosting an application server:** EBS boot and data volumes can be encrypted using the AWS KMS service.
- **RDS database server:** All boot and data volumes can be encrypted using the AWS KMS service.

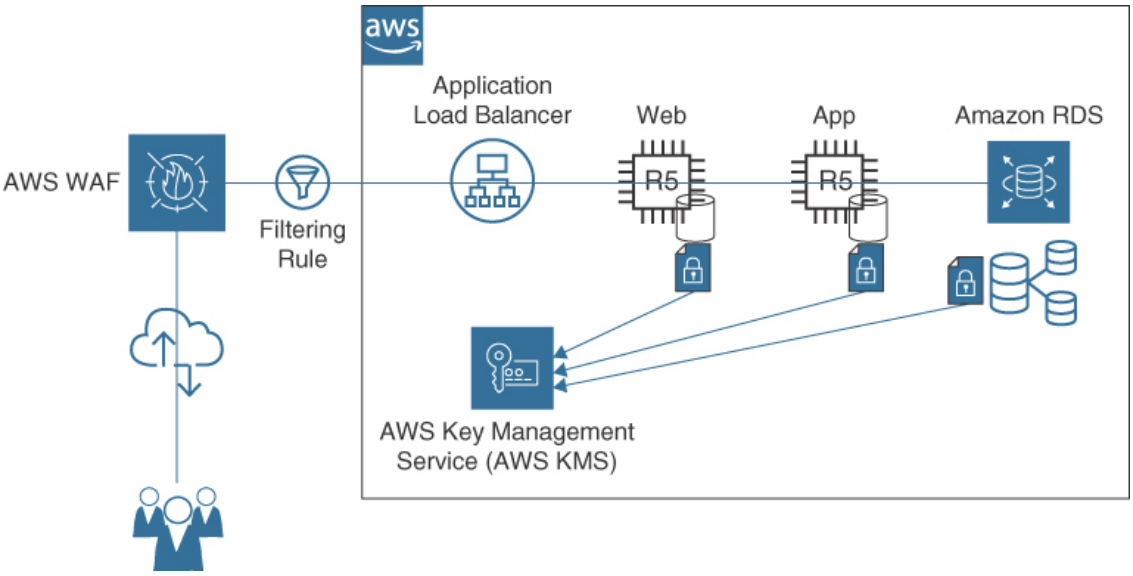

Figure 1-11 Encrypted Traffic Flow at AWS

Migrating Applications

For applications that have been chosen as starting candidates to move to the AWS cloud, several decisions need to be made about each application's journey or path. There are several options available for moving an application, depending on factors such as the age of the application and its operating system, and any local dependencies. The following sections walk through these options. Typical large organizations run many applications on thousands of virtual servers. When you move to AWS, you need to determine which applications can be moved to AWS and what applications should first be prioritized. Consider the following caveats before making these choices:

- **Define a value proposition:** Thousands of companies have successfully moved to AWS; you, too, can be successful. Start off with a defined value proposition that can be validated quickly—that is, in a matter of

months rather than years. For developing applications, you could consider developing with AWS Cloud9 (see **Figure 1-12**), a cloud-hosted integrated development environment (IDE) that supports more than 40 programming languages. Using Cloud9 and a browser, you can try your hand at developing a new application at AWS or at another PaaS provider such as Heroku. When you develop a completely new application at AWS, you are not constrained by factors such as the type of database that must be used, the type of programming language that must be used, or the type of compute that must be used. Starting new at AWS enables you to try out new methods to host applications, such as serverless computing, creating a mobile application using stateless components, or using DynamoDB as a NoSQL deployment instead of a SQL database. Developing and deploying a new workload at AWS without any legacy dependencies is where the real learning about what the AWS cloud can do for you begins.

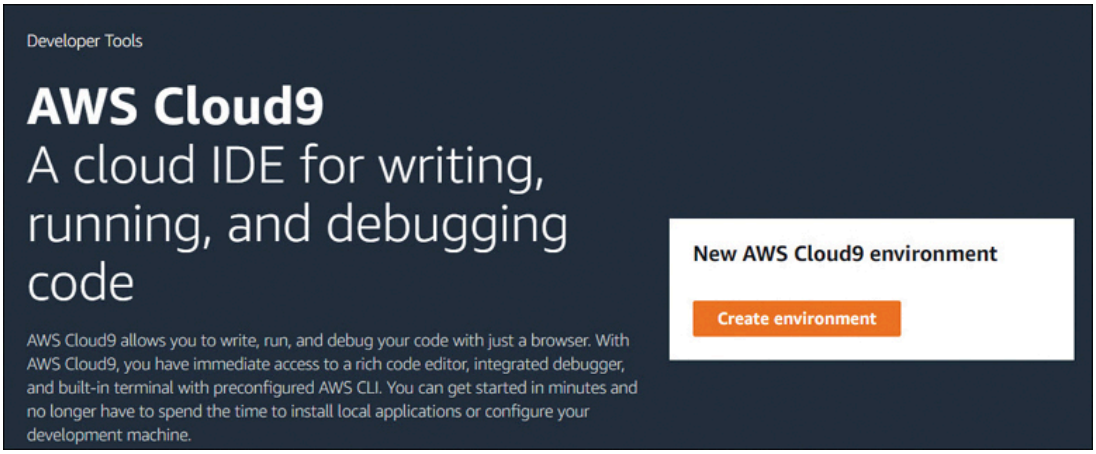

Figure 1-12 Cloud9 IDE at AWS for Application Development

- **Start with low value/low risk:** When choosing what application to move to the AWS cloud, many consultants begin by suggesting a starting point of selecting an already virtualized application stack with high value and low risk. However, it's probably going to take you many months or longer to successfully move a production application to the cloud. Think about choosing an application with low value first. This will enable you to do some additional planning and analysis without any added pressure. Many companies make the pronouncement that applications will be moving to the cloud quickly. It rarely happens as quickly as expected because there are so many things to learn and consider. Take your time and select a working application that has been virtualized and running successfully. Consider using the AWS Application Migration Service to migrate your first application to AWS.

After you are successful, document every step, including lessons learned and what to do differently for the next application chosen to be migrated. Moving additional applications to the cloud will generally be easier and faster thanks to the lessons learned and experience gained.

- **Solve a single problem:** Do you require additional storage? Perhaps that's a great starting point for moving resources to the AWS cloud. Archiving files in S3 Glacier could be as simple as ordering an external AWS Snowball device, connecting it up to your network, filling it with files that you would like to archive, and shipping it back to AWS. Archiving records in the AWS cloud would be an excellent first project in working with AWS.
- **Allowing access to on-premises data records:** The number-one problem for larger companies starting to work with cloud providers is working through the internal politics to allow access to on-premises data from the cloud. Be sure to consider data record access and the steps required for successful access before you begin moving to the cloud:
  - How can you access your on-premises data from the cloud?
  - What data records must stay on premises?
  - Are you bound by any compliance rules and regulations?
  - Is your current data in the right format for what you need?

Applications That Can Be Moved to AWS and Hosted on an EC2 Instance with No Changes

An application that fits into this category is referred to as *lift and shift* or *re-hosting*. Server migration tools and database migration tools can carry out these migrations quite effectively. AWS Application Discovery Service helps organizations plan migration projects by gathering information about their on-premises data centers and potentially thousands of workloads. Server utilization data and the mapping of any dependencies are useful first steps in the initial migration process. The collected data can be exported as a CSV file and used to estimate the total cost of ownership (TCO) of running workloads when planning migration to AWS.

AWS Application Migration Service (formally CloudEndure Migration) is the recommended migration service for performing lift-and-shift migrations to AWS because it automatically converts source servers from physical, virtual, or from existing third-party cloud providers to run at AWS. Supported physical servers include VMware vSphere and Microsoft

Hyper-V. EC2 instances can also be migrated between AWS regions or between AWS accounts.

However, applications that are lifted and shifted to the cloud are likely to have dependencies and issues that need to be considered before beginning the migration, including the following:

- If the application stores its data in a database, will the database remain on the premises or will it be moved to the cloud? The Database Migration Service can help in migrating many types of on-premises databases to the cloud.
- If the database for the application remains on premises, are there latency issues that need to be considered when communicating with the database? Each AWS site-to-site VPN connection supports a maximum throughput of up to 1.25 Gbps.
- Will a high-speed connection need to be established between the AWS cloud and the database remaining on premises? A high-speed private fiber AWS Direct Connect dedicated connection ranges from 1 to 100 Gbps.
- Are there compliance issues regarding the application data? Does the data have to be encrypted at rest? Does communication with the database need to be encrypted? AWS Artifact, available in the AWS Management console, provides compliance reports and agreements to review current compliance standards.
- Do users need to authenticate to the application across the corporate network? If so, are federation services required to be deployed at AWS for single sign-on (SSO)? IAM Identity Center provides SSO for multiple AWS accounts and SaaS cloud applications.
- Are there local dependencies installed on the application server that will interfere with the application server's operation in the AWS cloud? AWS Migration Hub Strategy Recommendations can be useful for alerting customers about potential migration conflicts for application migrations.
- Are there licensing considerations for both the operating system and the application when operating in the cloud? AWS License Manager can help track license usage across your environments.

Applications with Many Local Dependencies That Cause Problems

When Being Moved to the Cloud

For applications that fit in this category, consider the following:

- Application developers might have to refactor or restructure the source code of the application to take advantage of managed cloud services such as work queues (Amazon Simple Queue Service [SQS]), auto scaling (EC2 Auto Scaling), or hosted logging services (CloudWatch logs).
- Application developers might be able to take advantage of AWS cloud services by replacing the existing on-premises database with a database hosted in the cloud utilizing Amazon Relational Database Service (Amazon RDS).

Replacing an Existing Application with a SaaS Application Hosted by a Public Cloud Provider

With so many hosted cloud applications available in the public cloud, the odds are close to 100% that there will be an existing application that can replace a current on-premises application.

Applications That Should Remain On Premises and Eventually Be Deprecated

The following applications should not be moved to the cloud but should remain on premises or should be deprecated:

- The application is hosted on legacy hardware that is near end-of-life.
- The application cannot be virtualized.
- The application does not have technical support.
- The application is used by a small number of users.

The AWS Well-Architected Framework

Several years ago, AWS introduced the Well-Architected Framework to provide guidance to help cloud architects build secure, resilient, and well-performing infrastructure to host their applications. The framework describes recognized best practices developed over time, based on the experience of many AWS customers and AWS technical experts.

The documentation for the Well-Architected Framework (see https://docs.aws.amazon.com/wellarchitected/latest/framework) also presents many key questions customers should review. It is useful to dis-

cuss these questions with the other technical team members in your company to make key decisions about your infrastructure and workloads to be hosted at AWS. Each workload to be deployed at AWS should be viewed through the lens of the Well-Architected Framework following these six pillars:

- **Operational excellence:** Relates to how best to design, deploy, execute, and monitor applications running at AWS using automated deployment monitoring procedures, continuous improvement, and automated solutions for recovering from failures. Operational excellence questions to consider include:
  - How are disruptions to applications handled—manually or automatically?
  - How can you analyze the ongoing health of your applications and infrastructure components hosted at AWS?
- **Security:** Relates to how to best design systems that will operate reliably and securely while protecting customer information and data records. Security questions to consider include:
  - How are security credentials and authentication managed at AWS?
  - How are automated procedures secured?
- **Reliability:** Relates to how applications hosted at AWS recover from disruption with minimal downtime and how applications meet escalating demands. Reliability questions to consider include:
  - How do you monitor resources hosted at AWS?
  - How do applications hosted at AWS adapt to changes in demand by end users?
- **Performance efficiency:** Relates to how to use compute resources to meet and maintain your application requirements on an ongoing basis. Should your compute solution change from EC2 instances to containers or serverless? Performance efficiency questions to consider include:
  - Why did you select your database architecture?
  - Why did you select your current compute infrastructure?
- **Cost optimization:** Relates to how to design workloads that meet your needs at the lowest price point. Cost optimization questions to consider include:
  - How do you oversee usage and cost?
  - How do you meet cost targets?
  - Are you aware of current data transfer charges based on your AWS designs?

- **Sustainability:** Relates to designing workload deployments that minimize waste. Sustainability questions to consider include:
  - How do you select the most efficient storage and compute?
  - What managed service offerings could reduce current infrastructure deployments?

The Well-Architected Tool

In the AWS Management Console, you can search and find the AWS Well-Architected Framework tool. This tool, shown in **Figure 1-13**, provides a framework for documenting your workloads against AWS best practices, as defined in the Well-Architected Framework documentation. For each of the six pillars, there are many questions to consider before beginning to deploy an application. As questions for each pillar are considered and debated, milestones can be created marking important points about the workload architecture as teams discuss the questions and make changes to their workload design.

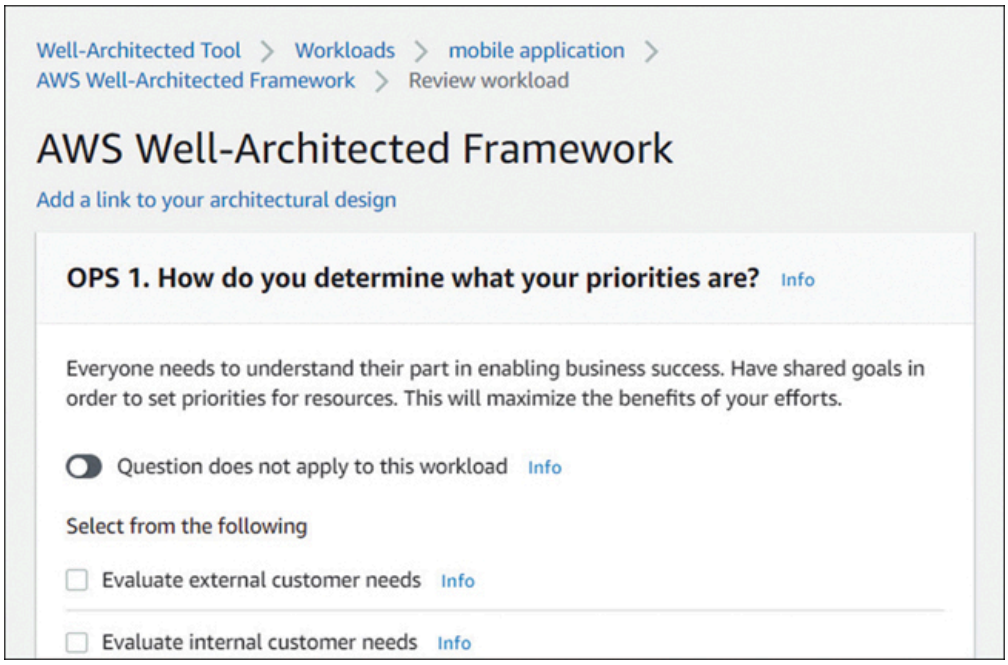

Figure 1-13 Evaluating Workloads Using the Well-Architected Framework Tool

The Well-Architected Framework tool provides tips and guidance on how to follow the best practices recommended by AWS while carrying out a full architectural review of an actual workload that you are planning to deploy at AWS. Your team will find that working with the Well-Architected Framework tool is well worth the time invested.

Before your architectural review begins, open the Well Architected Tool and select the AWS region where your application will be hosted, then de-

fine the workload and industry type, and whether the workload is in production or a pre-production environment. After all the pertinent questions have been answered, during the review process, the Well-Architected Framework tool helps you identify potential areas of medium and high risk, based on your answers to the questions. The six pillars of design success are also included in the plan for recommended improvements to your initial design decisions (see **Figure 1-14**).

Figure 1-14 Recommended Improvements Using the Well-Architected Framework Tool Review

AWS Services Cheat Sheet

Each section of the exam domains for the AWS Certified Solutions Architect – Associated (SAA-C03) exam is covered in a separate chapter in this book. You can quickly understand a variety of AWS services that are covered by the exam domains via a short explanation provided by the following list. You can review additional details on each of these services by reading the related FAQs for each service. There might be exam questions about some of these services, and then again, there might not be. For the purposes of preparing for the exam, the following details for these particular AWS services should be sufficient in answering the test questions that may mention them.

- **AWS AppSync:** A service designed for mobile applications that require user and data synchronization across multiple devices. AWS AppSync supports iOS, Android, and JavaScript (React and Angular). Select data records can be synchronized automatically across multiple devices using the GraphQL query language.
- **Amazon AppFlow:** A hosted integration service for securely exchanging data records, such as events from external SaaS applications such as Salesforce and ServiceNow.

- **Amazon Athena:** A serverless query service that analyzes Amazon S3 data. Queries can be performed in a variety of standards, including CSV, JSON, ORC, Avro, and Parquet. Queries can also be executed in parallel, resulting in extremely high performance.
- **AWS Audit Manager:** Audit Manager's prebuilt frameworks map your AWS resources to industry standards such as CIS AWS Foundations Benchmark, the General Data Protection Regulation (GDPR), and the Payment Card Industry Data Security Standard (PCI DSS).
- **Amazon Comprehend:** A natural language processing (NLP) service that uses machine learning to find meaning and insights in text.
- **Amazon Cognito:** Add mobile user sign-up, sign-in, and access controls to your web and mobile apps using a hosted identity store that supports both social media and enterprise identity federation.
- **Amazon Detective:** Analyze, investigate, and quickly identify the root cause of potential security issues or suspicious activities collecting log data from your AWS resources using machine learning, statistical analysis, and graph theory, ingesting data from AWS CloudTrail logs, Amazon VPC Flow Logs, and Amazon GuardDuty findings.
- **AWS Device Farm:** An application testing service that lets you improve the quality of your web and mobile apps during development by running tests concurrently on multiple desktop browsers and real physical mobile devices hosted at AWS. Device support includes Apple, Google, and Android devices.
- **AWS Data Exchange:** Supports the secure exchange of third-party data files and data tables into AWS. Customers can use the AWS Data Exchange API to copy selected third-party data from AWS Data Exchange into Amazon S3 storage. Data Exchange third-party products include weather, healthcare, data sciences, geospatial and mapping services.
- **AWS Data Pipeline:** Process and move data between different AWS compute and storage services, and from on-premises siloed data sources, and transfer the results into Amazon S3 buckets, Amazon RDS, Amazon DynamoDB, and Amazon EMR.
- **Amazon EMR:** EMR is a big data platform for data processing, interactive analysis, and machine learning using Apache Spark, Apache Hive, and Presto. Run petabyte-scale analysis much cheaper than traditional on-premises solutions.

- **Amazon Forecast:** Provides accurate time-sensitive forecasts for retail, manufacturing, travel demand, logistics, and web traffic markets.
- **Amazon Fraud Detector:** A managed fraud detector that helps identify potentially fraudulent online activities such as online payment fraud and fake account creation.
- **AWS Glue:** A fully managed extract, transform, and load (ETL) service that helps discover details and properties of data stored in Amazon S3 and Amazon Redshift for analytics, machine learning, and application development. AWS Glue has the following key components:
  - **AWS Glue Data Catalog:** Stores structural and operational metadata, including its table definition, physical location, and the data's historical and business relevance.
  - **Glue Crawlers:** Crawlers are used to scan various data stores populating the AWS Glue Data Catalog with relevant data statistics.
  - **AWS Glue Studio:** Create jobs that extract structured or semi-structured data from a data source.
  - **AWS Glue Schema Registry:** Validate and control streaming data using registered schemas for Apache Avro and JSON.
  - **AWS Glue DataBrew:** A visual data preparation tool that can be used by data analysts to clean and normalize data for analysis and machine learning.
- **Amazon Kendra:** Highly accurate machine learning enterprise search service for all unstructured data stored in Amazon S3 and Amazon RDS databases.
- **Amazon Kinesis:** Allows customers to connect, process, and analyze real-time streaming data to quickly gather insights to the incoming data flow of information. The use case for Amazon Kinesis is for ingesting, buffering, and processing streaming video, audio applications, logs, website clickstreams, and IoT telemetry data for machine learning, analysis, and storage at any scale.
  - **Amazon Kinesis Video Streams:** Developers can use the Kinesis Video Streams SDK to develop applications with connected camera devices, such as phones, drones, and dash cams, to securely stream video to custom real-time or batch-oriented applications running on AWS EC2 instances. The video streams can also be stored and encrypted for further monitoring and analytics.
  - **Amazon Kinesis Data Firehose:** Streaming data is collected and delivered in real time to Amazon S3, Amazon Redshift, Amazon Open

Search Service, custom HTTP/HTTPS endpoints, and to third-party service providers including Splunk, Datadog, and LogicMonitor. Kinesis Data Firehouse can also be configured to transform data records before the data is stored.

- **Amazon Kinesis Data Streams:** Collect and process gigabytes of streaming data that is generated continuously from thousands of locations such as log files, e-commerce purchases, game player activity, web clickstream data, and social media information. Multiple data streams ingested into Kinesis are sent into custom applications running on EC2 instances, or data stored in a DynamoDB table, Amazon S3 storage, Amazon EMR, or Amazon Redshift.

- **Amazon Lex:** Build conversational interfaces using voice and text powered by Alexa. Speech recognition and language understanding capabilities enable chatbots for applications published to Facebook Messenger, Slack, or Twilio SMS.
- **Amazon Managed Streaming for Apache Kafka (Amazon MSK):** Streaming data can be consumed using a full-managed Apache Kafka and Kafka Connect Clusters hosted at AWS, allowing Kafka applications and Kafka connectors to run at AWS without requiring expert knowledge in operating Apache Kafka.
- **Amazon Managed Service for Prometheus:** A monitoring and alerting service that collects and accesses performance and operational data from container workloads on AWS and on premises.
- **Amazon Managed Grafana:** Existing Grafana customers can analyze, monitor, and generate alarms on metrics, logs, and traces across AWS accounts, AWS regions, AWS CloudWatch, AWS X-Ray, Amazon Elasticsearch Service, Amazon Timestream, AWS IoT SiteWise, and Amazon Managed Service for Prometheus.
- **Amazon OpenSearch Service:** Perform log analysis and real-time application monitoring, providing visibility into your workload performance. Find relevant data within applications, websites, and data lakes using SQL query syntax. Data can be read using CSV tables or JSON documents.
- **Amazon Pinpoint:** An outbound and inbound marketing communications service allowing companies to connect with customers using email, SMS, push, voice messages, or in-app messaging to deliver promotional or transactional messages such as one-time passwords, reminders, or confirmation of orders.

- **Amazon Polly:** Turn text into lifelike speech for speech-enabled mobile apps and devices using lifelike voices in multiple languages; text sent to the Amazon Polly API returns an audio stream for use in your applications or devices.
- **AWS Personal Health Dashboard:** Receive notifications when AWS is experiencing issues on AWS services you are using, and alerts triggered by changes in the health of AWS services.
- **AWS Proton:** Allow platform teams to create rules for developers provisioning automated infrastructure as code. There are two supported methods:
  - AWS-managed provisioning uses CloudFormation templates to deploy infrastructure.
  - Self-managed provisioning uses Terraform templates to deploy infrastructure.
- **Amazon QuickSight:** A hosted business intelligence service powered by machine learning that provides data visualizations and insights from an organization's data records for reports or viewable dashboards. Accessed data records can be stored at AWS or stored in external locations including on-premises SQL Server, MySQL, and PostgreSQL databases, or in Amazon Redshift, RDS, Aurora, Athena, and S3 storage.
- **Amazon Rekognition:** Allows developers to add visual capabilities to applications using the following methods:
  - **Rekognition Image:** Searches, verifies, and organizes millions of images, detecting objects, scenes, and faces; identifies and extracts inappropriate content in images.
  - **Rekognition Video:** Extracts motion-based context from stored or live-stream videos for analysis, recognizing objects, celebrities, and inappropriate content in videos stored in Amazon S3 storage.
- **AWS Security Hub:** Provides a detailed view of your current security environment of a single or multiple AWS accounts by consuming and prioritizing the findings gathered from various AWS security services such as Amazon GuardDuty, AWS Config, Amazon Detective, AWS Firewall Manager, AWS IAM Access Analyzer, Amazon Inspector, Amazon Macie, and Amazon Trusted Advisor. Once enabled, AWS Security Hub executes continuous account-level configuration and security checks based on AWS best practices and industry standards.

- **Amazon SageMaker:** Build, train, and deploy machine learning (ML) models.
  - **Amazon SageMaker Autopilot:** Automatically inspect raw data and apply feature processors picking the best algorithm training and tuning multiple models and ranking each model based on performance.
  - **Amazon SageMaker Pipelines:** Create fully automated ML workflows.
- **Amazon Textract:** A document analysis service that detects and extracts printed text and handwriting from images and scans of uploaded documents.
- **Amazon Transcribe:** Converts speech to text.
- **Amazon Translate:** A neural machine translation service that delivers high-quality language translation.
- **AWS X-Ray:** Allows developers to analyze and debug applications in development and production to quickly identify and troubleshoot performance issues and errors, providing an end-to-end view of workload communication.
  - **Service map:** X-Ray creates a map of services and connections being used by your application and tracks all application requests.
  - **Identify:** Errors and bugs are highlighted by analyzing the response code for each request made to your application.
  - **Custom analysis:** X-Ray query APIs can be used to build your own analysis and visualization interfaces.

In Conclusion

In this initial chapter, we have looked at what the public cloud is and how AWS fits into the public cloud arena in terms of IaaS and PaaS services. This chapter also introduced the NIST definitions of the public cloud and how the AWS cloud fits into NIST's definition.

This chapter also introduced the AWS Well-Architected Framework, which is an essential guideline on accepted best practices for deploying and managing workloads in the AWS cloud using suggested best practices and procedures. If you are planning to take the AWS Certified Solutions Architect – Associate (SAA-C03) exam, you need to be familiar with the Well-Architected Framework. We finished with a summary of a variety of AWS services that you might encounter on the exam, with enough details to understand the purpose of each service for answering exam questions.