

# Chapter 1

## Introduction to AWS

**THE AWS CERTIFIED SOLUTIONS ARCHITECT ASSOCIATE EXAM OBJECTIVES COVERED IN THIS CHAPTER MAY INCLUDE, BUT ARE NOT LIMITED TO, THE FOLLOWING:**

**Domain 1.0: Designing highly available, cost-efficient, fault-tolerant, scalable systems**

**✓1.1 Identify and recognize cloud architecture considerations, such as fundamental components and effective designs.**

**Content may include the following:**

- How to design cloud services
- Planning and design
- Familiarity with:
  - Best practices for AWS architecture
  - Hybrid IT architectures (e.g., AWS Direct Connect, AWS Storage Gateway, Amazon Virtual Private Cloud [Amazon VPC], AWS Directory Service)
  - Elasticity and scalability (e.g., Auto Scaling, Amazon Simple Queue Service [Amazon SQS], Elastic Load Balancing, Amazon CloudFront)

**Domain 2.0: Implementation/Deployment**

**✓ 2.1 Identify the appropriate techniques and methods using Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3), AWS Elastic Beanstalk, AWS CloudFormation, AWS OpsWorks, Amazon VPC, and AWS Identity and Access Management (IAM) to code and implement a cloud solution.**

**Content may include the following:**

- Operate and extend service management in a hybrid IT architecture.
- Configure services to support compliance requirements in the cloud.
- Launch instances across the AWS global infrastructure.



In 2006, Amazon Web Services, Inc. (AWS) began offering IT infrastructure services to businesses in the form of web services, now commonly known as *cloud*

*computing*. One of the key benefits of cloud computing is the opportunity to replace up-front capital infrastructure expenses with low variable costs that scale with your business. With the cloud, businesses no longer need to plan for and procure servers and other IT infrastructure weeks or months in advance. Instead, they can instantly spin up hundreds or thousands of servers in minutes and deliver results faster.

Today, AWS provides a highly reliable, scalable, and low-cost infrastructure platform in the cloud that powers hundreds of thousands of businesses in more than 190 countries around the world.

This chapter provides an introduction to the AWS Cloud computing platform. It discusses the advantages of cloud computing and the fundamentals of AWS. It provides an overview of the AWS Cloud services that are fundamentally important for the exam.

# What Is Cloud Computing?

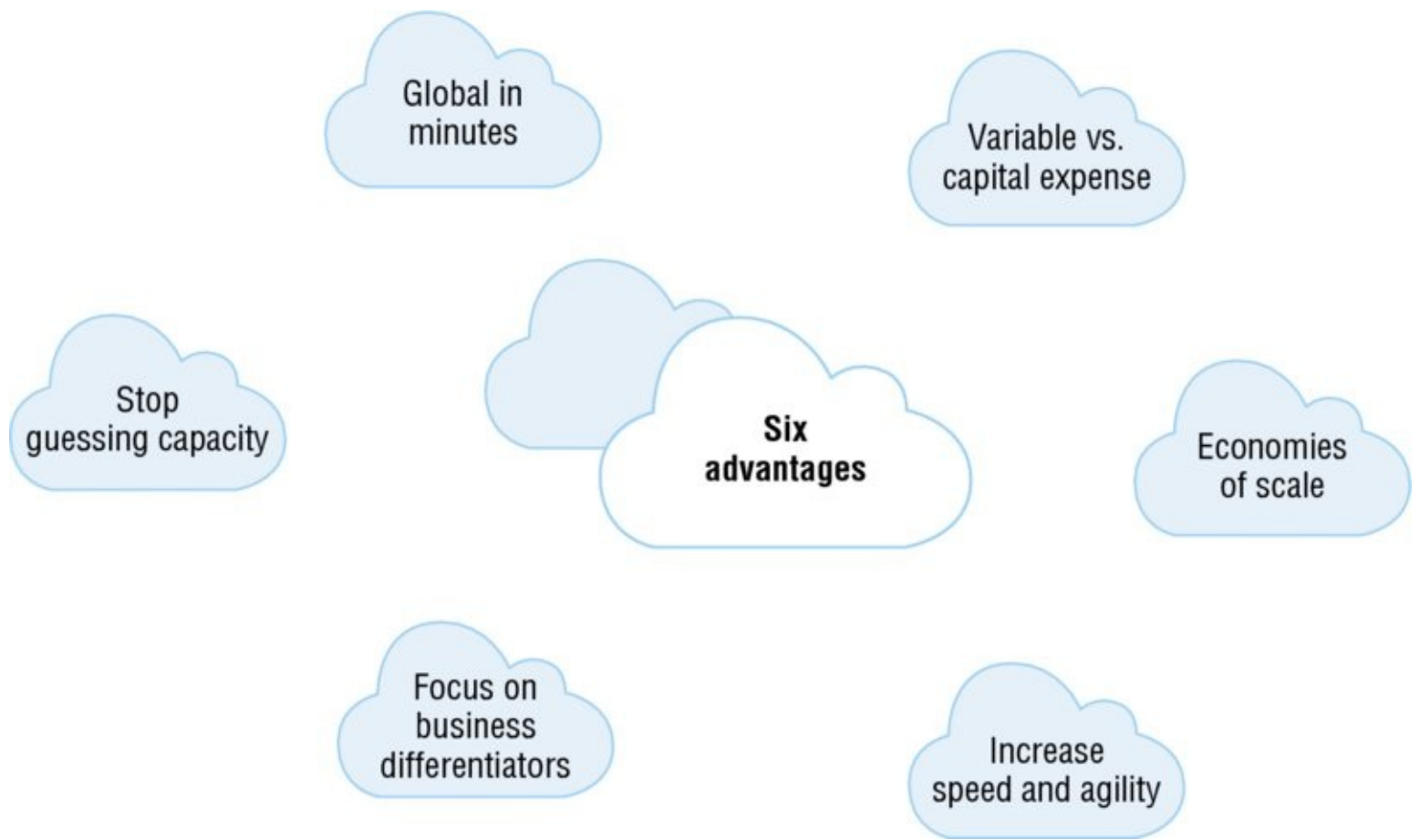
*Cloud computing* is the on-demand delivery of IT resources and applications via the Internet with pay-as-you-go pricing. Whether you run applications that share photos to millions of mobile users or deliver services that support the critical operations of your business, the cloud provides rapid access to flexible and low-cost IT resources. With cloud computing, you don't need to make large up-front investments in hardware and spend a lot of time managing that hardware. Instead, you can provision exactly the right type and size of computing resources you need to power your newest bright idea or operate your IT department. With cloud computing, you can access as many resources as you need, almost instantly, and only pay for what you use.

In its simplest form, cloud computing provides an easy way to access servers, storage, databases, and a broad set of application services over the Internet. Cloud computing providers such as AWS own and maintain the network-connected hardware required for these application services, while you provision and use what you need for your workloads.

## Advantages of Cloud Computing

Cloud computing introduces a revolutionary shift in how technology is obtained, used, and managed, and in how organizations budget and pay for technology services. With the ability to reconfigure the computing environment quickly to adapt to changing business requirements, organizations can optimize spending. Capacity can be automatically scaled up or down to meet fluctuating usage patterns. Services can be temporarily taken offline or shut down permanently as business demands dictate. In addition, with pay-per-use billing, AWS Cloud services become an operational expense instead of a capital expense.

While each organization experiences a unique journey to the cloud with numerous benefits, six advantages become apparent time and time again, as illustrated in [Figure 1.1](#).



**FIGURE 1.1** Six advantages of cloud computing

### Variable vs. Capital Expense

Let's begin with the ability to *trade capital expense for variable operational expense*. Instead of having to invest heavily in data centers and servers before knowing how you're going to use them, you can pay only when you consume computing resources and pay only for how much you consume.

### Economies of Scale

Another advantage of cloud computing is that *organizations benefit from massive economies of scale*. By using cloud computing, you can achieve a lower variable cost than you would get on your own. Because usage from hundreds of thousands of customers is aggregated in the cloud, providers such as AWS can achieve higher economies of scale, which translates into lower prices.

### Stop Guessing Capacity

When you make a capacity decision prior to deploying an application, you often end up either sitting on expensive idle resources or dealing with limited capacity. With cloud computing, organizations can *stop guessing about capacity requirements* for the infrastructure necessary to meet their business needs. They can access as much or as little as they need and scale up or down as required with only a few minutes' notice.

### Increase Speed and Agility

In a cloud computing environment, new IT resources are one click away, which allows

organizations to reduce the time it takes to make those resources available to developers from weeks to just minutes. This results in a dramatic *increase in speed and agility* for the organization, because the cost and time it takes to experiment and develop is significantly lower.

## Focus on Business Differentiators

Cloud computing allows organizations to focus on their business priorities, instead of on the heavy lifting of racking, stacking, and powering servers. By embracing this paradigm shift, organizations can *stop spending money on running and maintaining data centers*. This allows organizations to focus on projects that differentiate their businesses, such as analyzing petabytes of data, delivering video content, building great mobile applications, or even exploring Mars.

## Go Global in Minutes

Another advantage of cloud computing is the ability to *go global in minutes*. Organizations can easily deploy their applications to multiple locations around the world with just a few clicks. This allows organizations to provide redundancy across the globe and to deliver lower latency and better experiences to their customers at minimal cost. Going global used to be something only the largest enterprises could afford to do, but cloud computing democratizes this ability, making it possible for any organization.

While specific questions on these advantages of cloud computing are unlikely to be on the exam, having exposure to these benefits can help rationalize the appropriate answers.

## Cloud Computing Deployment Models

The two primary cloud computing deployment models that the exam focuses on are “all-in” cloud-based deployments and hybrid deployments. It is important to understand how each strategy applies to architectural options and decisions.

An *all-in cloud-based application* is fully deployed in the cloud, with all components of the application running in the cloud. Applications in the cloud have either been created in the cloud or have been migrated from an existing infrastructure to take advantage of the benefits of cloud computing. Cloud-based applications can be built on low-level infrastructure pieces or can use higher-level services that provide abstraction from the management, architecting, and scaling requirements of core infrastructure.

A *hybrid deployment* is a common approach taken by many enterprises that connects infrastructure and applications between cloud-based resources and existing resources, typically in an existing data center. The most common method of hybrid deployment is between the cloud and existing on-premises infrastructure to extend and grow an organization’s infrastructure while connecting cloud resources to internal systems. Choosing between an existing investment in infrastructure and moving to the cloud does not need to be a binary decision. Leveraging dedicated connectivity, identity federation, and integrated tools allows organizations to run hybrid applications across on-premises and cloud services.

# AWS Fundamentals

At its core, AWS provides on-demand delivery of IT resources via the Internet on a secure cloud services platform, offering compute power, storage, databases, content delivery, and other functionality to help businesses scale and grow. Using AWS resources instead of your own is like purchasing electricity from a power company instead of running your own generator, and it provides the key advantages of cloud computing: Capacity exactly matches your need, you pay only for what you use, economies of scale result in lower costs, and the service is provided by a vendor experienced in running large-scale networks.

AWS global infrastructure and AWS approach to security and compliance are key foundational concepts to understand as you prepare for the exam.

## Global Infrastructure

AWS serves over one million active customers in more than 190 countries, and it continues to expand its global infrastructure steadily to help organizations achieve lower latency and higher throughput for their business needs.

AWS provides a highly available technology infrastructure platform with multiple locations worldwide. These locations are composed of regions and Availability Zones. Each *region* is a separate geographic area. Each region has multiple, isolated locations known as *Availability Zones*. AWS enables the placement of resources and data in multiple locations. Resources aren't replicated across regions unless organizations choose to do so.

Each region is completely independent and is designed to be completely isolated from the other regions. This achieves the greatest possible fault tolerance and stability. Each Availability Zone is also isolated, but the Availability Zones in a region are connected through low-latency links. Availability Zones are physically separated within a typical metropolitan region and are located in lower-risk flood plains (specific flood zone categorization varies by region). In addition to using a discrete uninterruptable power supply (UPS) and on-site backup generators, they are each fed via different grids from independent utilities (when available) to reduce single points of failure further. Availability Zones are all redundantly connected to multiple tier-1 transit providers. By placing resources in separate Availability Zones, you can protect your website or application from a service disruption impacting a single location.



You can achieve high availability by deploying your application across multiple Availability Zones. Redundant instances for each tier (for example, web, application, and database) of an application should be placed in distinct Availability Zones, thereby creating a multisite solution. At a minimum, the goal is to have an independent copy of each application stack in two or more Availability Zones.

## Security and Compliance

Whether on-premises or on AWS, information security is of paramount importance to

organizations running critical workloads. Security is a core functional requirement that protects mission-critical information from accidental or deliberate theft, leakage, integrity compromise, and deletion. Helping to protect the confidentiality, integrity, and availability of systems and data is of the utmost importance to AWS, as is maintaining your trust and confidence.

This section is intended to provide a very brief introduction to AWS approach to security and compliance. Chapter 12, “Security on AWS,” and Chapter 13, “AWS Risk and Compliance,” will address these topics in greater detail, including the importance of each on the exam.

## Security

Cloud security at AWS is the number one priority. All AWS customers benefit from data center and network architectures built to satisfy the requirements of the most security-sensitive organizations. AWS and its partners offer hundreds of tools and features to help organizations meet their security objectives for visibility, auditability, controllability, and agility. This means that organizations can have the security they need, but without the capital outlay and with much lower operational overhead than in an on-premises environment.

Organizations leveraging AWS inherit all the best practices of AWS policies, architecture, and operational processes built to satisfy the requirements of the most security-sensitive customers. The AWS infrastructure has been designed to provide the highest availability while putting strong safeguards in place regarding customer privacy and segregation. When deploying systems on the AWS Cloud computing platform, AWS helps by sharing the security responsibilities with the organization. AWS manages the underlying infrastructure, and the organization can secure anything it deploys on AWS. This affords each organization the flexibility and agility they need in security controls.

This infrastructure is built and managed not only according to security best practices and standards, but also with the unique needs of the cloud in mind. AWS uses redundant and layered controls, continuous validation and testing, and a substantial amount of automation to ensure that the underlying infrastructure is monitored and protected 24/7. AWS ensures that these controls are consistently applied in every new data center or service.

## Compliance

When customers move their production workloads to the AWS Cloud, both parties become responsible for managing the IT environment. Customers are responsible for setting up their environment in a secure and controlled manner. Customers also need to maintain adequate governance over their entire IT control environment. By tying together governance-focused, audit-friendly service features with applicable compliance or audit standards, AWS enables customers to build on traditional compliance programs. This helps organizations establish and operate in an AWS security control environment.



Organizations retain complete control and ownership over the region in which their data is physically located, allowing them to meet regional compliance and data residency requirements.



The IT infrastructure that AWS provides to organizations is designed and managed in alignment with security best practices and a variety of IT security standards. The following is a partial list of the many certifications and standards with which AWS complies:










- Service Organization Controls (SOC) 1/International Standard on Assurance Engagements (ISAE) 3402, SOC 2, and SOC 3
- Federal Information Security Management Act (FISMA), Department of Defense Information Assurance Certification and Accreditation Process (DIACAP), and Federal Risk and Authorization Management Program (FedRAMP)
- Payment Card Industry Data Security Standard (PCI DSS) Level 1
- International Organization for Standardization (ISO) 9001, ISO 27001, and ISO 27018

AWS provides a wide range of information regarding its IT control environment to help organizations achieve regulatory commitments in the form of reports, certifications, accreditations, and other third-party attestations.



# AWS Cloud Computing Platform

AWS provides many cloud services that you can combine to meet business or organizational needs (see [Figure 1.2](#)). While being knowledgeable about all the platform services will allow you to be a well-rounded solutions architect, understanding the services and fundamental concepts outlined in this book will help prepare you for the AWS Certified Solutions Architect – Associate exam.

Enterprise Applications	 Virtual Desktops			 Sharing and Collaboration	
Platform Services	Databases	Analytics	App Services	Deployment and Management	Mobile Services
	Relational	Hadoop	Queuing	Containers	Identity
		Real-Time	Orchestration	DevOps Tools	Syns
	NoSQL	Data Warehouses	App Streaming	Resources Templates	Mobile Analytics
	Caching	Data Workflows	Email Search	Usage Tracking	Notifications
Foundation Services	 Compute (VMs, Auto Scaling and load Balancing)		 Storage (Object, Block and Archive)		 Security and Access Control
					 Networking
Infrastructure	 Regions	 Availability Zones	 Content Delivery Networks and Points of Presence		

**FIGURE 1.2** AWS Cloud computing platform

This section introduces the major AWS Cloud services by category. Subsequent chapters provide a deeper view of the services pertinent to the exam.

## Accessing the Platform

To access AWS Cloud services, you can use the AWS Management Console, the AWS Command Line Interface (CLI), or the AWS Software Development Kits (SDKs).

The *AWS Management Console* is a web application for managing AWS Cloud services. The console provides an intuitive user interface for performing many tasks. Each service has its own console, which can be accessed from the AWS Management Console. The console also provides information about the account and billing.

The *AWS Command Line Interface (CLI)* is a unified tool used to manage AWS Cloud services. With just one tool to download and configure, you can control multiple services from the command line and automate them through scripts.

The *AWS Software Development Kits (SDKs)* provide an application programming interface (API) that interacts with the web services that fundamentally make up the AWS platform. The SDKs provide support for many different programming languages and platforms to allow you to work with your preferred language. While you can certainly make HTTP calls directly

to the web service endpoints, using the SDKs can take the complexity out of coding by providing programmatic access for many of the services.

## Compute and Networking Services

AWS provides a variety of compute and networking services to deliver core functionality for businesses to develop and run their workloads. These compute and networking services can be leveraged with the storage, database, and application services to provide a complete solution for computing, query processing, and storage across a wide range of applications. This section offers a high-level description of the core computing and networking services.

### Amazon Elastic Compute Cloud (Amazon EC2)

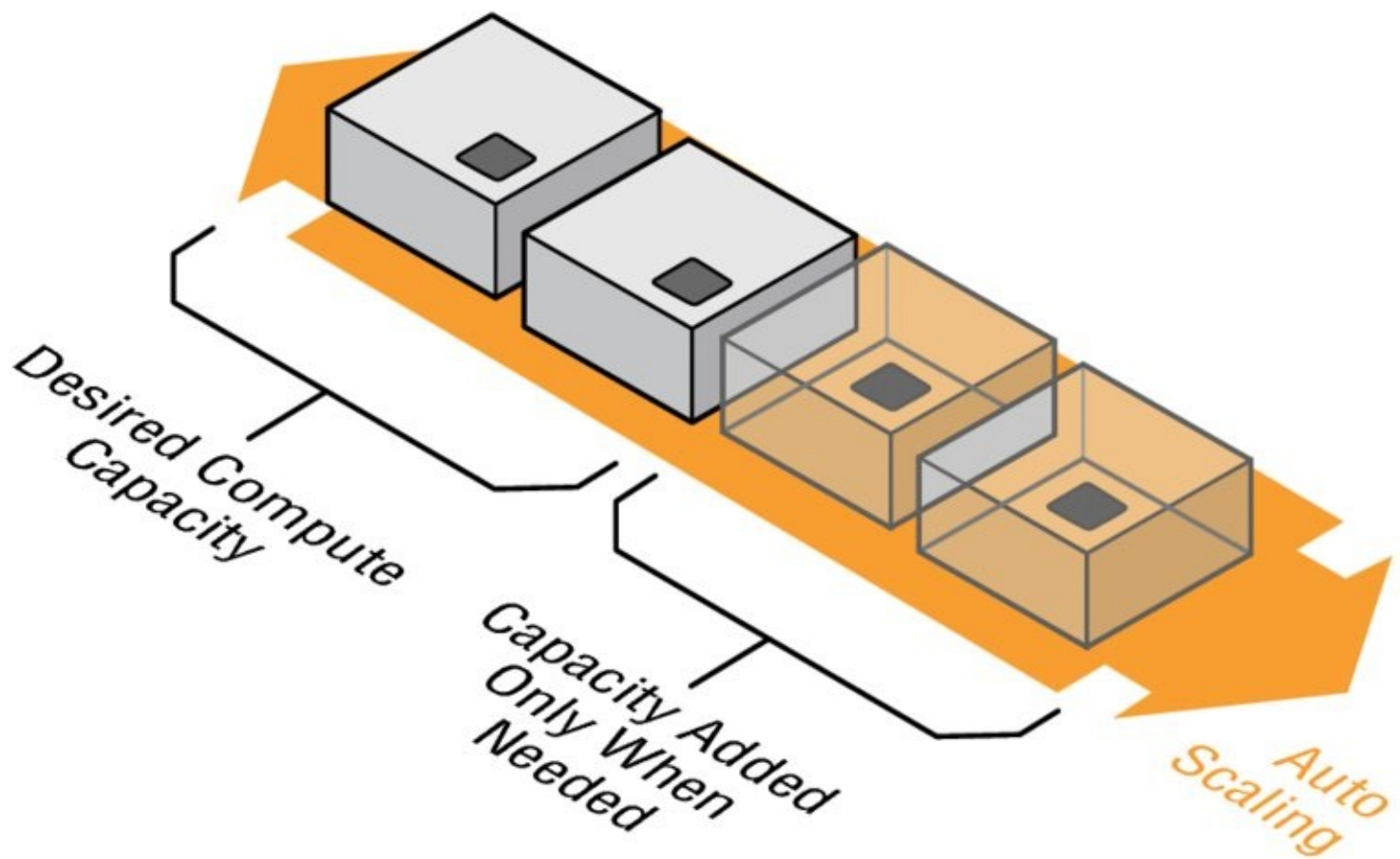
*Amazon Elastic Compute Cloud (Amazon EC2)* is a web service that provides resizable compute capacity in the cloud. It allows organizations to obtain and configure virtual servers in Amazon's data centers and to harness those resources to build and host software systems. Organizations can select from a variety of operating systems and resource configurations (memory, CPU, storage, and so on) that are optimal for the application profile of each workload. Amazon EC2 presents a true virtual computing environment, allowing organizations to launch compute resources with a variety of operating systems, load them with custom applications, and manage network access permissions while maintaining complete control.

### AWS Lambda

*AWS Lambda* is a zero-administration compute platform for back-end web developers that runs your code for you on the AWS Cloud and provides you with a fine-grained pricing structure. AWS Lambda runs your back-end code on its own AWS compute fleet of Amazon EC2 instances across multiple Availability Zones in a region, which provides the high availability, security, performance, and scalability of the AWS infrastructure.

### Auto Scaling

*Auto Scaling* allows organizations to scale Amazon EC2 capacity up or down automatically according to conditions defined for the particular workload (see [Figure 1.3](#)). Not only can it be used to help maintain application availability and ensure that the desired number of Amazon EC2 instances are running, but it also allows resources to scale in and out to match the demands of dynamic workloads. Instead of provisioning for peak load, organizations can optimize costs and use only the capacity that is actually needed.



**FIGURE 1.3** Auto scaling capacity

Auto Scaling is well suited both to applications that have stable demand patterns and to applications that experience hourly, daily, or weekly variability in usage.

### **Elastic Load Balancing**

*Elastic Load Balancing* automatically distributes incoming application traffic across multiple Amazon EC2 instances in the cloud. It enables organizations to achieve greater levels of fault tolerance in their applications, seamlessly providing the required amount of load balancing capacity needed to distribute application traffic.

### **AWS Elastic Beanstalk**

*AWS Elastic Beanstalk* is the fastest and simplest way to get a web application up and running on AWS. Developers can simply upload their application code, and the service automatically handles all the details, such as resource provisioning, load balancing, Auto Scaling, and monitoring. It provides support for a variety of platforms, including PHP, Java, Python, Ruby, Node.js, .NET, and Go. With AWS Elastic Beanstalk, organizations retain full control over the AWS resources powering the application and can access the underlying resources at any time.

### **Amazon Virtual Private Cloud (Amazon VPC)**

*Amazon Virtual Private Cloud (Amazon VPC)* lets organizations provision a logically isolated section of the AWS Cloud where they can launch AWS resources in a virtual network that they define. Organizations have complete control over the virtual environment, including selection of the IP address range, creation of subnets, and configuration of route tables and

network gateways. In addition, organizations can extend their corporate data center networks to AWS by using hardware or software *virtual private network (VPN)* connections or dedicated circuits by using AWS Direct Connect.

## **AWS Direct Connect**

*AWS Direct Connect* allows organizations to establish a dedicated network connection from their data center to AWS. Using AWS Direct Connect, organizations can establish private connectivity between AWS and their data center, office, or colocation environment, which in many cases can reduce network costs, increase bandwidth throughput, and provide a more consistent network experience than Internet-based VPN connections.

## **Amazon Route 53**

*Amazon Route 53* is a highly available and scalable Domain Name System (DNS) web service. It is designed to give developers and businesses an extremely reliable and cost-effective way to route end users to Internet applications by translating human readable names, such as [www.example.com](http://www.example.com), into the numeric IP addresses, such as 192.0.2.1, that computers use to connect to each other. Amazon Route 53 also serves as domain registrar, allowing you to purchase and manage domains directly from AWS.

## **Storage and Content Delivery**

AWS provides a variety of services to meet your storage needs, such as Amazon Simple Storage Service, Amazon CloudFront, and Amazon Elastic Block Store. This section provides an overview of the storage and content delivery services.

### **Amazon Simple Storage Service (Amazon S3)**

*Amazon Simple Storage Service (Amazon S3)* provides developers and IT teams with highly durable and scalable object storage that handles virtually unlimited amounts of data and large numbers of concurrent users. Organizations can store any number of objects of any type, such as HTML pages, source code files, image files, and encrypted data, and access them using HTTP-based protocols. Amazon S3 provides cost-effective object storage for a wide variety of use cases, including backup and recovery, nearline archive, big data analytics, disaster recovery, cloud applications, and content distribution.

### **Amazon Glacier**

*Amazon Glacier* is a secure, durable, and extremely low-cost storage service for data archiving and long-term backup. Organizations can reliably store large or small amounts of data for a very low cost per gigabyte per month. To keep costs low for customers, Amazon Glacier is optimized for infrequently accessed data where a retrieval time of several hours is suitable. Amazon S3 integrates closely with Amazon Glacier to allow organizations to choose the right storage tier for their workloads.

### **Amazon Elastic Block Store (Amazon EBS)**

*Amazon Elastic Block Store (Amazon EBS)* provides persistent block-level storage volumes for use with Amazon EC2 instances. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect organizations from component failure, offering high



availability and durability. By delivering consistent and low-latency performance, Amazon EBS provides the disk storage needed to run a wide variety of workloads.

## **AWS Storage Gateway**

*AWS Storage Gateway* is a service connecting an on-premises software appliance with cloud-based storage to provide seamless and secure integration between an organization's on-premises IT environment and the AWS storage infrastructure. The service supports industry-standard storage protocols that work with existing applications. It provides low-latency performance by maintaining a cache of frequently accessed data on-premises while securely storing all of your data encrypted in Amazon S3 or Amazon Glacier.

## **Amazon CloudFront**

*Amazon CloudFront* is a content delivery web service. It integrates with other AWS Cloud services to give developers and businesses an easy way to distribute content to users across the world with low latency, high data transfer speeds, and no minimum usage commitments. Amazon CloudFront can be used to deliver your entire website, including dynamic, static, streaming, and interactive content, using a global network of edge locations. Requests for content are automatically routed to the nearest edge location, so content is delivered with the best possible performance to end users around the globe.

## **Database Services**

AWS provides fully managed relational and NoSQL database services, and in-memory caching as a service and a petabyte-scale data warehouse solution. This section provides an overview of the products that the database services comprise.

### **Amazon Relational Database Service (Amazon RDS)**

*Amazon Relational Database Service (Amazon RDS)* provides a fully managed relational database with support for many popular open source and commercial database engines. It's a cost-efficient service that allows organizations to launch secure, highly available, fault-tolerant, production-ready databases in minutes. Because Amazon RDS manages time-consuming administration tasks, including backups, software patching, monitoring, scaling, and replication, organizational resources can focus on revenue-generating applications and business instead of mundane operational tasks.

### **Amazon DynamoDB**

*Amazon DynamoDB* is a fast and flexible NoSQL database service for all applications that need consistent, single-digit millisecond latency at any scale. It is a fully managed database and supports both document and key/value data models. Its flexible data model and reliable performance make it a great fit for mobile, web, gaming, ad-tech, Internet of Things, and many other applications.

### **Amazon Redshift**

*Amazon Redshift* is a fast, fully managed, petabyte-scale data warehouse service that makes it simple and cost effective to analyze structured data. Amazon Redshift provides a standard SQL interface that lets organizations use existing business intelligence tools. By leveraging

columnar storage technology that improves I/O efficiency and parallelizing queries across multiple nodes, Amazon Redshift is able to deliver fast query performance. The Amazon Redshift architecture allows organizations to automate most of the common administrative tasks associated with provisioning, configuring, and monitoring a cloud data warehouse.

## Amazon ElastiCache

*Amazon ElastiCache* is a web service that simplifies deployment, operation, and scaling of an in-memory cache in the cloud. The service improves the performance of web applications by allowing organizations to retrieve information from fast, managed, in-memory caches, instead of relying entirely on slower, disk-based databases. As of this writing, Amazon ElastiCache supports Memcached and Redis cache engines.

## Management Tools

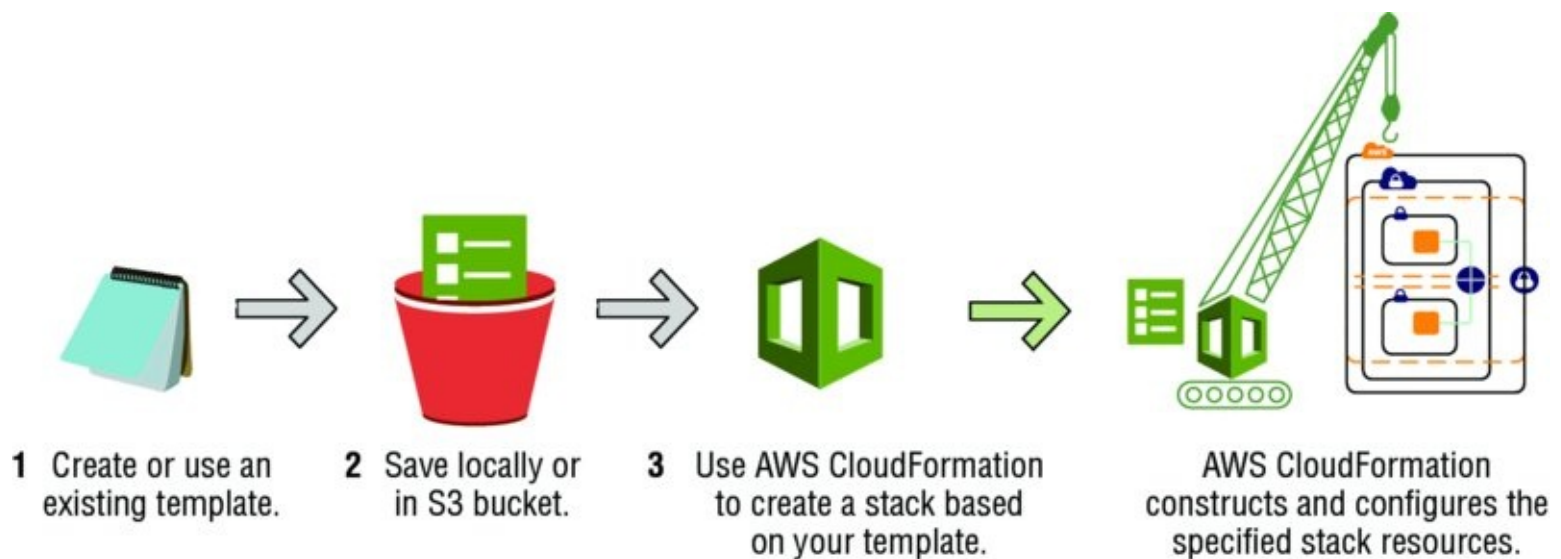
AWS provides a variety of tools that help organizations manage your AWS resources. This section provides an overview of the management tools that AWS provides to organizations.

## Amazon CloudWatch

*Amazon CloudWatch* is a monitoring service for AWS Cloud resources and the applications running on AWS. It allows organizations to collect and track metrics, collect and monitor log files, and set alarms. By leveraging Amazon CloudWatch, organizations can gain system-wide visibility into resource utilization, application performance, and operational health. By using these insights, organizations can react, as necessary, to keep applications running smoothly.

## AWS CloudFormation

*AWS CloudFormation* gives developers and systems administrators an effective way to create and manage a collection of related AWS resources, provisioning and updating them in an orderly and predictable fashion. AWS CloudFormation defines a JSON-based templating language that can be used to describe all the AWS resources that are necessary for a workload. Templates can be submitted to AWS CloudFormation and the service will take care of provisioning and configuring those resources in appropriate order (see [Figure 1.4](#)).



**FIGURE 1.4** AWS CloudFormation workflow summary

## **AWS CloudTrail**

*AWS CloudTrail* is a web service that records AWS API calls for an account and delivers log files for audit and review. The recorded information includes the identity of the API caller, the time of the API call, the source IP address of the API caller, the request parameters, and the response elements returned by the service.

## **AWS Config**

*AWS Config* is a fully managed service that provides organizations with an AWS resource inventory, configuration history, and configuration change notifications to enable security and governance. With AWS Config, organizations can discover existing AWS resources, export an inventory of their AWS resources with all configuration details, and determine how a resource was configured at any point in time. These capabilities enable compliance auditing, security analysis, resource change tracking, and troubleshooting.

## **Security and Identity**

AWS provides security and identity services that help organizations secure their data and systems on the cloud. The following section explores these services at a high level.

### **AWS Identity and Access Management (IAM)**

*AWS Identity and Access Management (IAM)* enables organizations to securely control access to AWS Cloud services and resources for their users. Using IAM, organizations can create and manage AWS users and groups and use permissions to allow and deny their access to AWS resources.

### **AWS Key Management Service (KMS)**

*AWS Key Management Service (KMS)* is a managed service that makes it easy for organizations to create and control the encryption keys used to encrypt their data and uses Hardware Security Modules (HSMs) to protect the security of your keys. AWS KMS is integrated with several other AWS Cloud services to help protect data stored with these services.

### **AWS Directory Service**

*AWS Directory Service* allows organizations to set up and run Microsoft Active Directory on the AWS Cloud or connect their AWS resources with an existing on-premises Microsoft Active Directory. Organizations can use it to manage users and groups, provide single sign-on to applications and services, create and apply Group Policies, domain join Amazon EC2 instances, and simplify the deployment and management of cloud-based Linux and Microsoft Windows workloads.

### **AWS Certificate Manager**

*AWS Certificate Manager* is a service that lets organizations easily provision, manage, and deploy Secure Sockets Layer/Transport Layer Security (SSL/TLS) certificates for use with AWS Cloud services. It removes the time-consuming manual process of purchasing, uploading, and renewing SSL/TLS certificates. With AWS Certificate Manager, organizations



can quickly request a certificate, deploy it on AWS resources such as Elastic Load Balancing or Amazon CloudFront distributions, and let AWS Certificate Manager handle certificate renewals.

## **AWS Web Application Firewall (WAF)**

*AWS Web Application Firewall (WAF)* helps protect web applications from common attacks and exploits that could affect application availability, compromise security, or consume excessive resources. AWS WAF gives organizations control over which traffic to allow or block to their web applications by defining customizable web security rules.

## **Application Services**

AWS provides a variety of managed services to use with applications. The following section explores the application services at a high level.

### **Amazon API Gateway**

*Amazon API Gateway* is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale. Organizations can create an API that acts as a “front door” for applications to access data, business logic, or functionality from back-end services, such as workloads running on Amazon EC2, code running on AWS Lambda, or any web application. Amazon API Gateway handles all the tasks involved in accepting and processing up to hundreds of thousands of concurrent API calls, including traffic management, authorization and access control, monitoring, and API version management.

### **Amazon Elastic Transcoder**

*Amazon Elastic Transcoder* is media transcoding in the cloud. It is designed to be a highly scalable and cost-effective way for developers and businesses to convert (or transcode) media files from their source formats into versions that will play back on devices like smartphones, tablets, and PCs.

### **Amazon Simple Notification Service (Amazon SNS)**

*Amazon Simple Notification Service (Amazon SNS)* is a web service that coordinates and manages the delivery or sending of messages to recipients. In Amazon SNS, there are two types of clients—publishers and subscribers—also referred to as producers and consumers. Publishers communicate asynchronously with subscribers by producing and sending a message to a topic, which is a logical access point and communication channel. Subscribers consume or receive the message or notification over one of the supported protocols when they are subscribed to the topic.

### **Amazon Simple Email Service (Amazon SES)**

*Amazon Simple Email Service (Amazon SES)* is a cost-effective email service that organizations can use to send transactional email, marketing messages, or any other type of content to their customers. Amazon SES can also be used to receive messages and deliver them to an Amazon S3 bucket, call custom code via an AWS Lambda function, or publish notifications to Amazon SNS.

## **Amazon Simple Workflow Service (Amazon SWF)**

*Amazon Simple Workflow Service (Amazon SWF)* helps developers build, run, and scale background jobs that have parallel or sequential steps. Amazon SWF can be thought of as a fully managed state tracker and task coordinator on the cloud. In common architectural patterns, if your application's steps take more than 500 milliseconds to complete, it is vitally important to track the state of processing and to provide the ability to recover or retry if a task fails. Amazon SWF helps organizations achieve this reliability.

## **Amazon Simple Queue Service (Amazon SQS)**

*Amazon Simple Queue Service (Amazon SQS)* is a fast, reliable, scalable, fully managed message queuing service. Amazon SQS makes it simple and cost effective to decouple the components of a cloud application. With Amazon SQS, organizations can transmit any volume of data, at any level of throughput, without losing messages or requiring other services to be always available.

# Summary

The term “cloud computing” refers to the on-demand delivery of IT resources via the Internet with pay-as-you-go pricing. Instead of buying, owning, and maintaining data centers and servers, organizations can acquire technology such as compute power, storage, databases, and other services on an as-needed basis. With cloud computing, AWS manages and maintains the technology infrastructure in a secure environment and businesses access these resources via the Internet to develop and run their applications. Capacity can grow or shrink instantly and businesses pay only for what they use.

Cloud computing introduces a revolutionary shift in how technology is obtained, used, and managed, and how organizations budget and pay for technology services. While each organization experiences a unique journey to the cloud with numerous benefits, six advantages become apparent time and time again. Understanding these advantages allows architects to shape solutions that deliver continuous benefits to organizations.

AWS provides a highly available technology infrastructure platform with multiple locations worldwide. These locations are composed of regions and Availability Zones. This enables organizations to place resources and data in multiple locations around the globe. Helping to protect the confidentiality, integrity, and availability of systems and data is of the utmost importance to AWS, as is maintaining the trust and confidence of organizations around the world.

AWS offers a broad set of global compute, storage, database, analytics, application, and deployment services that help organizations move faster, lower IT costs, and scale applications. Having a broad understanding of these services allows solutions architects to design effective distributed applications and systems on the AWS platform.

# Exam Essentials

**Understand the global infrastructure.** AWS provides a highly available technology infrastructure platform with multiple locations worldwide. These locations are composed of regions and Availability Zones. Each region is located in a separate geographic area and has multiple, isolated locations known as Availability Zones.

**Understand regions.** An AWS region is a physical geographic location that consists of a cluster of data centers. AWS regions enable the placement of resources and data in multiple locations around the globe. Each region is completely independent and is designed to be completely isolated from the other regions. This achieves the greatest possible fault tolerance and stability. Resources aren't replicated across regions unless organizations choose to do so.

**Understand Availability Zones.** An Availability Zone is one or more data centers within a region that are designed to be isolated from failures in other Availability Zones. Availability Zones provide inexpensive, low-latency network connectivity to other zones in the same region. By placing resources in separate Availability Zones, organizations can protect their website or application from a service disruption impacting a single location.

**Understand the hybrid deployment model.** A hybrid deployment model is an architectural pattern providing connectivity for infrastructure and applications between cloud-based resources and existing resources that are not located in the cloud.

# Review Questions

1. Which of the following describes a physical location around the world where AWS clusters data centers?
  - A. Endpoint
  - B. Collection
  - C. Fleet
  - D. Region
2. Each AWS region is composed of two or more locations that offer organizations the ability to operate production systems that are more highly available, fault tolerant, and scalable than would be possible using a single data center. What are these locations called?
  - A. Availability Zones
  - B. Replication areas
  - C. Geographic districts
  - D. Compute centers
3. What is the deployment term for an environment that extends an existing on-premises infrastructure into the cloud to connect cloud resources to internal systems?
  - A. All-in deployment
  - B. Hybrid deployment
  - C. On-premises deployment
  - D. Scatter deployment
4. Which AWS Cloud service allows organizations to gain system-wide visibility into resource utilization, application performance, and operational health?
  - A. AWS Identity and Access Management (IAM)
  - B. Amazon Simple Notification Service (Amazon SNS)
  - C. Amazon CloudWatch
  - D. AWS CloudFormation
5. Which of the following AWS Cloud services is a fully managed NoSQL database service?
  - A. Amazon Simple Queue Service (Amazon SQS)
  - B. Amazon DynamoDB
  - C. Amazon ElastiCache
  - D. Amazon Relational Database Service (Amazon RDS)
6. Your company experiences fluctuations in traffic patterns to their e-commerce website

based on flash sales. What service can help your company dynamically match the required compute capacity to the spike in traffic during flash sales?

- A. Auto Scaling
- B. Amazon Glacier
- C. Amazon Simple Notification Service (Amazon SNS)
- D. Amazon Virtual Private Cloud (Amazon VPC)

7. Your company provides an online photo sharing service. The development team is looking for ways to deliver image files with the lowest latency to end users so the website content is delivered with the best possible performance. What service can help speed up distribution of these image files to end users around the world?

- A. Amazon Elastic Compute Cloud (Amazon EC2)
- B. Amazon Route 53
- C. AWS Storage Gateway
- D. Amazon CloudFront

8. Your company runs an Amazon Elastic Compute Cloud (Amazon EC2) instance periodically to perform a batch processing job on a large and growing filesystem. At the end of the batch job, you shut down the Amazon EC2 instance to save money but need to persist the filesystem on the Amazon EC2 instance from the previous batch runs. What AWS Cloud service can you leverage to meet these requirements?

- A. Amazon Elastic Block Store (Amazon EBS)
- B. Amazon DynamoDB
- C. Amazon Glacier
- D. AWS CloudFormation

9. What AWS Cloud service provides a logically isolated section of the AWS Cloud where organizations can launch AWS resources in a virtual network that they define?

- A. Amazon Simple Workflow Service (Amazon SWF)
- B. Amazon Route 53
- C. Amazon Virtual Private Cloud (Amazon VPC)
- D. AWS CloudFormation

10. Your company provides a mobile voting application for a popular TV show, and 5 to 25 million viewers all vote in a 15-second timespan. What mechanism can you use to decouple the voting application from your back-end services that tally the votes?

- A. AWS CloudTrail
- B. Amazon Simple Queue Service (Amazon SQS)
- C. Amazon Redshift
- D. Amazon Simple Notification Service (Amazon SNS)