

Chapter 1

Introduction to Cloud Computing and AWS

The cloud is where much of the serious technology innovation and growth happens these days, and Amazon Web Services (AWS), more than any other, is the platform of choice for business and institutional workloads. If you want to be successful as an AWS solutions architect, you'll first need to understand what the cloud really is and how Amazon's end of it works.

TO MAKE SURE YOU'VE GOT THE BIG PICTURE, THIS CHAPTER WILL EXPLORE THE BASICS:

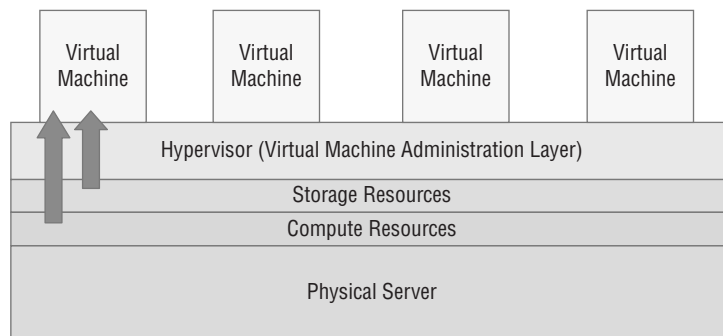
- ✓ **What makes cloud computing different from other applications and client-server models**
- ✓ **How the AWS platform provides secure and flexible virtual networked environments for your resources**
- ✓ **How AWS provides such a high level of service reliability**
- ✓ **How to access and manage your AWS-based resources**
- ✓ **How to migrate existing on-premises resources to your AWS account**
- ✓ **Where you can go for documentation and help with your AWS deployments**



Cloud Computing and Virtualization

The technology that lies at the core of all cloud operations is virtualization. As illustrated in Figure 1.1, virtualization lets you divide the hardware resources of a single physical server into smaller units. That physical server could therefore host multiple virtual machines (VMs) running their own complete operating systems, each with its own memory, storage, and network access.

FIGURE 1.1 A virtual machine host



Virtualization's flexibility makes it possible to provision a virtual server in a matter of seconds, run it for exactly the time your project requires, and then shut it down. The resources released will become instantly available to other workloads. The usage density you can achieve lets you squeeze the greatest value from your hardware and makes it easy to generate experimental and sandboxed environments.

Cloud Computing Architecture

Major cloud providers like AWS have enormous server farms where hundreds of thousands of servers and disk drives are maintained along with the network cabling necessary

to connect them. A well-built virtualized environment could provide a virtual server using storage, memory, compute cycles, and network bandwidth collected from the most efficient mix of available sources it can find.

A cloud computing platform offers on-demand, self-service access to pooled compute resources where your usage is metered and billed according to the volume you consume. Cloud computing systems allow for precise billing models, sometimes involving fractions of a penny for an hour of consumption.

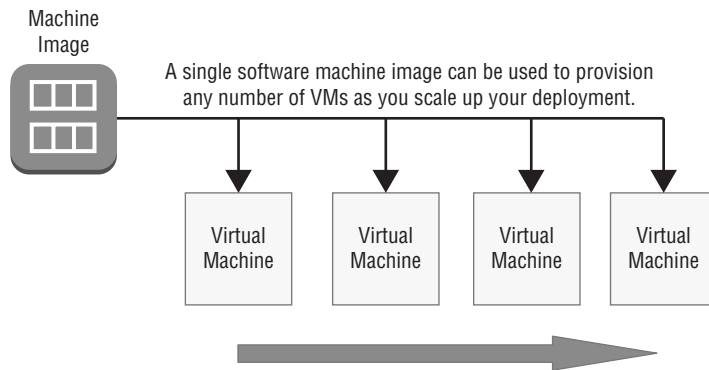
Cloud Computing Optimization

The cloud is a great choice for so many serious workloads because it's scalable, elastic, and often a lot cheaper than traditional alternatives. Effective deployment provisioning will require some insight into those three features.

Scalability

A scalable infrastructure can efficiently meet unexpected increases in demand for your application by automatically adding resources. As Figure 1.2 shows, this most often means dynamically increasing the number of virtual machines (or *instances* as AWS calls them) you've got running.

FIGURE 1.2 Copies of a machine image are added to new VMs as they're launched.



AWS offers its autoscaling service through which you define a machine image that can be instantly and automatically replicated and launched into multiple instances to meet demand.

Elasticity

The principle of elasticity covers some of the same ground as scalability—both address how the system manages changing demand. However, though the images used in a scalable environment let you ramp up capacity to meet rising demand, an elastic infrastructure will

automatically *reduce* capacity when demand drops. This makes it possible to control costs, since you'll run resources only when they're needed.

Cost Management

Besides the ability to control expenses by closely managing the resources you use, cloud computing transitions your IT spending from a capital expenditure (capex) framework into something closer to operational expenditure (opex).

In practical terms, this means you no longer have to spend \$10,000 up front for every new server you deploy—along with associated electricity, cooling, security, and rack space costs. Instead, you're billed much smaller incremental amounts for as long as your application runs.

The cloud paradigm also makes it easy to scale resources up and down to meet changing demand. For an online e-commerce business, for instance, that might mean deploying extra resources only during periods of peak demand and then automatically decommissioning them as demand falls.

That doesn't necessarily mean your long-term cloud-based opex costs will always be less than you'd pay over the lifetime of a comparable datacenter deployment. But it does mean you won't have to expose yourself to risky speculation about your long-term needs. If, sometime in the future, changing demand calls for new hardware, AWS will be able to deliver it within a minute or two.

To help you understand the full implications of cloud compute spending, AWS provides a free Pricing Calculator at <http://calculator.aws/#>. This calculator helps you perform proper “apples-to-apples” comparisons between your current datacenter costs and what an identical operation would cost you on AWS.

The AWS Cloud

Keeping up with the steady stream of new services showing up on the AWS Console can be overwhelming. But as a solutions architect, your main focus should be on the core service categories. This section briefly summarizes each of the core categories (as shown in Table 1.1) and then does the same for key individual services. You'll learn much more about all of these (and other) services through the rest of the book, but it's worth focusing on these short definitions, because they lie at the foundation of everything else you're going to learn.

TABLE 1.1 AWS service categories

Category	Function
Compute	Services replicating the traditional role of local physical servers for the cloud, offering advanced configurations such as autoscaling, load balancing, containers, and even serverless architectures (a method for delivering server functionality with a very small footprint)
Networking	Application connectivity, access control, and enhanced remote connections

Category	Function
Storage	Various kinds of storage platforms designed to fit a range of both immediate accessibility and long-term backup needs
Database	Managed data solutions for use cases requiring multiple data formats: relational, NoSQL, or caching
Application management	Monitoring, auditing, and configuring AWS account services and running resources
Security and identity	Services for managing authentication and authorization, data and connection encryption, and integration with third-party authentication management systems

Table 1.2 describes the functions of some core AWS services, organized by category.

TABLE 1.2 Core AWS services (by category)

Category	Service	Function
Compute	Elastic Compute Cloud (EC2)	EC2 server instances provide virtual versions of the servers you would run in your local datacenter. EC2 instances can be provisioned with the CPU, memory, storage, and network interface profile to meet any application need, from a simple web server to one part of a cluster of instances, providing an integrated multi-tiered fleet architecture. Since EC2 instances are virtual, they're resource-efficient and deploy nearly instantly.
	Lambda	Serverless application architectures like the one provided by Amazon's Lambda service allow you to provide responsive public-facing services without the need for a server that's actually running 24/7. Instead, network events (like consumer requests) can trigger the execution of a pre-defined code-based operation. When the operation (which can currently run for as long as 15 minutes) is complete, the Lambda event ends, and all resources automatically shut down.
	Auto Scaling	Copies of running EC2 instances can be defined as image templates and automatically launched (or <i>scaled up</i>) when client demand can't be met by existing instances. As demand drops, unused instances can be terminated (or <i>scaled down</i>).

TABLE 1.2 Core AWS services (by category) *(continued)*

Category	Service	Function
	Elastic Load Balancing	Incoming network traffic can be directed between multiple web servers to ensure that a single web server isn't overwhelmed while other servers are underused or that traffic isn't directed to failed servers.
	Elastic Container Service	Compute workloads using container technologies like Docker and Kubernetes can be provisioned, automated, deployed, and administered using full integration with your other AWS account resources. Kubernetes workloads have their own environment: Amazon Elastic Kubernetes Service (EKS).
	Elastic Beanstalk	Beanstalk is a managed service that abstracts the provisioning of AWS compute and networking infrastructure. You are required to do nothing more than push your application code, and Beanstalk automatically launches and manages all the necessary services in the background.
Networking	Virtual Private Cloud (VPC)	VPCs are highly configurable networking environments designed to host your EC2 (and RDS) instances. You use VPC-based tools to secure and, if desired, isolate your instances by closely controlling inbound and outbound network access.
	Direct Connect	By purchasing fast and secure network connections to AWS through a third-party provider, you can use Direct Connect to establish an enhanced direct tunnel between your local datacenter or office and your AWS-based VPCs.
	Route 53	Route 53 is the AWS DNS service that lets you manage domain registration, record administration, routing protocols, and health checks, which are all fully integrated with the rest of your AWS resources.
	CloudFront	CloudFront is Amazon's distributed global content delivery network (CDN). When properly configured, a CloudFront distribution can store cached versions of your site's content at edge locations around the world so that they can be delivered to customers on request with the greatest efficiency and lowest latency.
Storage	Simple Storage Service (S3)	S3 offers highly versatile, reliable, and inexpensive object storage that's great for data storage and backups. It's also commonly used as part of larger AWS production processes, including through the storage of script, template, and log files.

Category	Service	Function
	S3 Glacier	A good choice for when you need large data archives stored cheaply over the long term and can live with retrieval delays measuring in the hours. Glacier's life cycle management is closely integrated with S3.
	Elastic Block Store (EBS)	EBS provides the persistent virtual storage drives that host the operating systems and working data of an EC2 instance. They're meant to mimic the function of the storage drives and partitions attached to physical servers.
	Storage Gateway	Storage Gateway is a hybrid storage system that exposes AWS cloud storage as a local, on-premises appliance. Storage Gateway can be a great tool for migration and data backup and as part of disaster recovery operations.
Database	Relational Database Service (RDS)	RDS is a managed service that builds you a stable, secure, and reliable database instance. You can run a variety of SQL database engines on RDS, including MySQL, Microsoft SQL Server, Oracle, and Amazon's own Aurora.
	DynamoDB	DynamoDB can be used for fast, flexible, highly scalable, and managed nonrelational (NoSQL) database workloads.
Application management	CloudWatch	CloudWatch can be set to monitor process performance and resource utilization and, when preset thresholds are met, either send you a message or trigger an automated response.
	CloudFormation	This service enables you to use template files to define full and complex AWS deployments. The ability to script your use of any AWS resources makes it easier to automate, standardizing and speeding up the application launch process.
	CloudTrail	CloudTrail collects records of all your account's API events. This history is useful for account auditing and troubleshooting purposes.
	Config	The Config service is designed to help you with change management and compliance for your AWS account. You first define a desired configuration state, and Config evaluates any future states against that ideal. When a configuration change pushes too far from the ideal baseline, you'll be notified.

TABLE 1.2 Core AWS services (by category) (*continued*)

Category	Service	Function
Security and identity	Identity and Access Management (IAM)	You use IAM to administer user and programmatic access and authentication to your AWS account. Through the use of users, groups, roles, and policies, you can control exactly who and what can access and/or work with any of your AWS resources.
	Key Management Service (KMS)	KMS is a managed service that allows you to administrate the creation and use of encryption keys to secure data used by and for any of your AWS resources.
	Directory Service	For AWS environments that need to manage identities and relationships, Directory Service can integrate AWS resources with identity providers like Amazon Cognito and Microsoft AD domains.
Application integration	Simple Notification Service (SNS)	SNS is a notification tool that can automate the publishing of alert <i>topics</i> to other services (to an SQS Queue or to trigger a Lambda function, for instance), to mobile devices, or to recipients using email or SMS.
	Simple Workflow (SWF)	SWF lets you coordinate a series of tasks that must be performed using a range of AWS services or even nondigital (meaning human) events.
	Simple Queue Service (SQS)	SQS allows for event-driven messaging within distributed systems that can decouple while coordinating the discrete steps of a larger process. The data contained in your SQS messages will be reliably delivered, adding to the fault-tolerant qualities of an application.
	API Gateway	This service enables you to create and manage secure and reliable APIs for your AWS-based applications.

AWS Platform Architecture

AWS maintains datacenters for its physical servers around the world. Because the centers are so widely distributed, you can reduce your own services' network transfer latency by hosting your workloads geographically close to your users. It can also help you manage compliance with regulations, requiring you to keep data within a particular legal jurisdiction. This can help meet any data sovereignty requirements you might face.

Datacenters exist within AWS regions, of which there are currently 21—not including private U.S. government AWS GovCloud regions—although this number is constantly growing. It’s important to always be conscious of the region you have selected when you launch new AWS resources; pricing and service availability can vary from one to the next. Table 1.3 shows a list of all 24 (nongovernment) regions, along with each region’s name and core endpoint address. Note that accessing and authenticating to the two Chinese regions requires unique protocols.

TABLE 1.3 A list of publicly accessible AWS regions

Region name	Region	Endpoint
US East (N. Virginia)	us-east-1	us-east-1.amazonaws.com
US East (Ohio)	us-east-2	us-east-2.amazonaws.com
US West (N. California)	us-west-1	us-west-1.amazonaws.com
US West (Oregon)	us-west-2	us-west-2.amazonaws.com
Africa (Cape Town)	af-south-1	af-south-1.amazon.aws.com
Asia Pacific (Hong Kong)	ap-east-1	ap-east-1.amazonaws.com
Asia Pacific (Mumbai)	ap-south-1	ap-south-1.amazonaws.com
Asia Pacific (Seoul)	ap-northeast-2	ap-northeast-2.amazonaws.com
Asia Pacific (Osaka-Local)	ap-northeast-3	ap-northeast-3.amazonaws.com
Asia Pacific (Singapore)	ap-southeast-1	ap-southeast-1.amazonaws.com
Asia Pacific (Sydney)	ap-southeast-2	ap-southeast-2.amazonaws.com
Asia Pacific (Tokyo)	ap-northeast-1	ap-northeast-1.amazonaws.com
Asia Pacific (Jakarta)	ap-southeast-3	ap-southeast-3.amazon.aws.com
Canada (Central)	ca-central-1	ca-central-1.amazonaws.com
China (Beijing)	cn-north-1	cn-north-1.amazonaws.com.cn
China (Ningxia)	cn-northwest-1	cn-northwest-1.amazonaws.com.cn

TABLE 1.3 A list of publicly accessible AWS regions (*continued*)

Region name	Region	Endpoint
EU (Frankfurt)	eu-central-1	eu-central-1.amazonaws.com
EU (Ireland)	eu-west-1	eu-west-1.amazonaws.com
EU (London)	eu-west-2	eu-west-2.amazonaws.com
EU (Paris)	eu-west-3	eu-west-3.amazonaws.com
EU (Stockholm)	eu-north-1	eu-north-1.amazonaws.com
Europe (Milan)	eu-south-1	eu-south-1.amazonaws.com
Middle East (Bahrain)	me-south-1	me-south-1.amazonaws.com
South America (São Paulo)	sa-east-1	sa-east-1.amazonaws.com



Endpoint addresses are used to access your AWS resources remotely from within application code or scripts. Prefixes like `ec2`, `api-gateway`, or `cloudformation` are often added to the endpoints to specify a particular AWS service. Such an address might look like this: `cloudformation.us-east-2.amazonaws.com`. You can see a complete list of endpoint addresses and their prefixes at <http://docs.aws.amazon.com/general/latest/gr/rande.html>.

Because low-latency access is so important, certain AWS services are offered from designated edge network locations. These services include Amazon CloudFront, Amazon Route 53, AWS Firewall Manager, AWS Shield, and AWS WAF. For a complete and up-to-date list of available locations, see <http://aws.amazon.com/about-aws/global-infrastructure/regional-product-services>.

Physical AWS datacenters are exposed within your AWS account as availability zones. There might be half a dozen availability zones within a region, like `us-east-1a` and `us-east-1b`, each consisting of one or more datacenters.

You organize your resources from a region within one or more virtual private clouds (VPCs). A VPC is effectively a network address space within which you can create network subnets and associate them with availability zones. When configured properly, this architecture can provide effective resource isolation and durable replication.

AWS Reliability and Compliance

AWS has a lot of the basic regulatory, legal, and security groundwork covered before you even launch your first service.

AWS has invested significant planning and funds into resources and expertise relating to infrastructure administration. Its heavily protected and secretive datacenters, layers of redundancy, and carefully developed best-practice protocols would be difficult or even impossible for a regular enterprise to replicate.

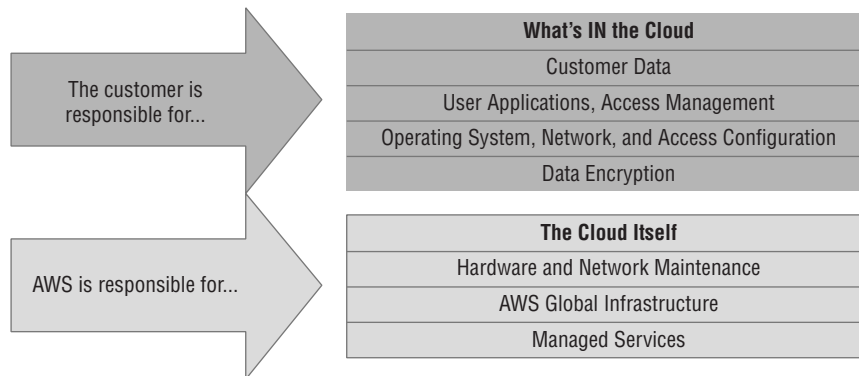
Where applicable, resources on the AWS platform are compliant with dozens of national and international standards, frameworks, and certifications, including ISO 9001, FedRAMP, NIST, and GDPR. (See <http://aws.amazon.com/compliance/programs> for more information.)

The AWS Shared Responsibility Model

Of course, those guarantees cover only the underlying AWS platform. The way you decide to use AWS resources is your business—and therefore your responsibility. So, it’s important to be familiar with the AWS Shared Responsibility Model.

AWS guarantees the secure and uninterrupted operation of its “cloud.” That means its physical servers, storage devices, networking infrastructure, and managed services. AWS customers, as illustrated in Figure 1.3, are responsible for whatever happens *within* that cloud. This covers the security and operation of installed operating systems, client-side data, the movement of data across networks, end-user authentication and access, and customer data.

FIGURE 1.3 The AWS Shared Responsibility Model



Of course, precisely defining the “cloud”—meaning, what AWS is responsible for—will depend on the particular cloud model you’re using. For infrastructure as a service (IaaS) resources like EC2, AWS leaves all software and networking configurations—including the OS itself—in your hands. By contrast, AWS takes care of just about everything except your application code for platform as a service (PaaS) services like Elastic Beanstalk.

The AWS Service Level Agreement

By “guarantee,” AWS doesn’t mean that service disruptions or security breaches will *never* occur. Drives may stop spinning, major electricity systems may fail, and natural disasters may happen. But when something does go wrong, AWS will provide service credits to reimburse customers for their *direct* losses whenever uptimes fall below a defined threshold. Of course, that won’t help you recover customer confidence or lost business.

The exact percentage of the guarantee will differ according to service. The service level agreement (SLA) rate for AWS EC2, for instance, is set to 99.99 percent—meaning that you can expect your EC2 instances, ECS containers, and EBS storage devices to be available for all but around four minutes of each month.

The important thing to remember is that it’s not *if* things will fail but *when*. Build your applications to be geographically dispersed and fault-tolerant so that when things do break, your users will barely notice.

Working with AWS

Whatever AWS services you choose to run, you’ll need a way to manage them all. The browser-based management console is an excellent way to introduce yourself to a service’s features and learn how it will perform in the real world. And, as you’ll soon see, AWS provides some practical and useful tools for handling the many administration tasks necessary for keeping your accounts happy and healthy.

There are few AWS administration tasks that you can’t do from the console, which provides plenty of useful visualizations and helpful documentation. But as you become more familiar with the way things work, and especially as your AWS deployments become more complex, you’ll probably find yourself getting more of your serious work done away from the console.

So let’s explore some of those tools.

AWS Organizations

You can consolidate the management of multiple AWS accounts using AWS Organizations. Why bother? Because it’s not uncommon for larger organizations to have so many projects and resources running on AWS that they need some degree of resource isolation to keep a

handle on them. It can also be hugely convenient to maintain some level of bird’s-eye-view supervision over the whole thing—and to channel billing through a single payer.

AWS Organizations lets you create new accounts and invite existing accounts into an organization. Once your accounts are connected, the organization administration can create global access policies much the same way as you might for individual users within a single account.

Account security for an organization is even more critical than for a single account since a breach could potentially impact all the resources in all member accounts. You should therefore be especially careful to enforce all the normal security best practices, including the use of multifactor authentication (MFA), strong passwords, and properly securing root accounts.

AWS Organizations allows you to:

- Share resources through a unified AWS Single Sign-on configuration that applies global permissions to all your accounts.
- Apply IAM rules globally through service control policies (SCPs).
- Create and manage accounts—along with account users and groups—programmatically.
- Audit, monitor, and secure all your environments for compliance and functional purposes. A single account’s CloudTrail service can be configured to watch events from your entire organization.

AWS Organizations has replaced the old and deprecated Consolidated Billing tool.

AWS Control Tower

The functionality of AWS Organizations can be extended through “landing zones” generated by Control Tower. Landing zones streamline the onboarding of new accounts, automatically applying your organization’s governance policies and incorporating them into your cloud infrastructure.

You’ll be prompted to configure your landing zone to operate in one or multiple AWS regions and to be controlled by the owners of specifically defined management and log archive accounts. Once you do, you’ll be able to select blueprints for imposing security, compliance, and operations best practices over the governed resources.

Beyond policy creation, Control Tower landing zones simplify ongoing monitoring of your resource and policies and allows closer management of your infrastructure.

AWS Service Catalog

Organizations often wish to provide their teams with effective and consistent sets of tools. People working with business analytics, marketing, or research projects, for example, might benefit from easy access to a predefined set of AWS services and even prebuilt machine images from the AWS Marketplace. You can define customized bundles of resources through a CloudFormation template that points to just the tools you want.

AWS Service Catalog will take your template and use it to make exactly resources you approve available to exactly the users you designate. You can set your own availability and consumption limits, and Service Catalog handles security and general best practices.

AWS License Manager

AWS License Manager is a service that helps you manage your software licenses. Through the License Manager dashboard, the service allows you to avoid licensing violations by tracking usage, monitoring compliance, and enforcing applicable rules. You can use License Manager to manage licenses for both on-premises and cloud-based software.

You can use License Manager itself without incurring costs.

AWS Artifact

AWS Artifact is a digital repository that allows customers to download compliance-related information about their AWS accounts and services. You can use AWS Artifact to access compliance reports and security audit information, such as AWS Service Organization Control (SOC) reports, Payment Card Industry Data Security Standard (PCI DSS) reports, and ISO 27001 certifications. AWS Artifact also makes it easy for you to share compliance-related information with auditors and other stakeholders.

The AWS CLI

The AWS Command-Line Interface (CLI) lets you run complex AWS operations from your local command line. Once you get the hang of how it works, you'll discover that it can make things much simpler and more efficient.

As an example, suppose you need to launch a half-dozen EC2 instances to make up a microservices environment. Each instance is meant to play a separate role and therefore will require a subtly different provisioning process. Clicking through window after window to launch the instances from the console can quickly become tedious and time-consuming, especially if you find yourself repeating the task every few days. But the whole process can alternatively be incorporated into a simple script that you can run from your local terminal shell or PowerShell interface using the AWS CLI.

Installing and configuring the AWS CLI on Linux, Windows, or macOS machines is not hard at all, but the details might change depending on your platform. For the most up-to-date instructions, see <http://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html>. Exercise 1.1 shows how to use it.

EXERCISE 1.1

Use the AWS CLI

Install (if necessary) and configure the AWS CLI on your local system and demonstrate it's running properly by listing the buckets that currently exist in your account. For extra practice, create an S3 bucket and then copy a simple file or document from your machine to your new bucket. From the browser console, confirm that the file reached its target.

To get you started, here are some basic CLI commands:

```
aws s3 ls
aws s3 mb <bucketname>
aws s3 cp /path/to/file.txt s3://bucketname
```

AWS SDKs

If you want to incorporate access to your AWS resources into your application code, you'll need to use an AWS software development kit (SDK) for the language you're working with. AWS currently offers SDKs for nine languages, including Java, .NET, and Python, and a number of mobile SDKs that include Android and iOS. There are also toolkits available for IntelliJ, Azure DevOps, Visual Studio, and Visual Studio Code.

You can see a full overview of AWS developer tools at <http://aws.amazon.com/tools>.

Technical Support and Online Resources

Things won't always go smoothly for you—on AWS just as everywhere else in your life. Sooner or later, you'll need some kind of technical or account support. There's a variety of types of support, and you should understand what's available.

One of the first things you'll be asked to decide when you create a new AWS account is which support plan you'd like. Your business needs and budget will determine the way you answer this question.

Support Plans

The Basic plan is free with every account and gives you access to customer service, along with documentation, whitepapers, and the support forum. Customer service covers billing and account support issues.

The Developer plan starts at \$29/month or 3 percent of monthly AWS charges and adds access for one account holder to a Cloud Support associate along with limited general guidance and “system impaired” response.

For \$100/month or 10 percent of monthly AWS charges (up to \$10,000), the Business plan will deliver faster guaranteed response times to unlimited users for help with “impaired” systems, personal guidance and troubleshooting, and a support API.

The greater of \$5,500 or 10 percent of monthly AWS charges will get you Enterprise On-Ramp service, which adds fast response for business-critical system down events and proactive guidance and consultative review based on your applications.

Finally, Enterprise support plans cover all of the other features, plus direct access to AWS solutions architects for operational and design reviews, your own technical account manager, and access to online self-paced labs. For complex, mission-critical deployments, those benefits can make a big difference. But they’ll cost you at least \$15,000 each month.

You can read more about AWS support plans at <http://aws.amazon.com/premiumsupport/plans>.

Other Support Resources

There’s plenty of self-serve support available outside of the official support plans:

- re:Post, the AWS community help forum site, is open to anyone with a valid AWS account (<http://repost.aws>).
- Extensive and well-maintained AWS documentation is available at <http://docs.aws.amazon.com>.
- The AWS Well-Architected page (<http://aws.amazon.com/architecture/well-architected>) is a hub that links to some valuable whitepapers and documentation addressing best practices for cloud deployment design.
- There are also plenty of third-party companies offering commercial support for your AWS deployments.

Migrating Existing Resources to AWS

AWS offers a set of tools to help you migrate your existing applications and infrastructure resources to the cloud. These tools include the AWS Migration Hub, the AWS Application Migration Service, and the AWS Database Migration Service. Each of these tools has its own set of features and benefits that can help you move your resources to AWS quickly and securely.

AWS Migration Hub

AWS Migration Hub is a service that helps organizations track their progress as they migrate their workloads to the AWS cloud. The Migration Hub dashboard provides a central place to view information about your existing inventory and the status of all your migration projects, making it easy to track progress and identify any issues that may come up.

The Migration Hub can be used to coordinate the operations of all migration-related AWS tools and to assess the feasibility of migrating your workloads to AWS. It can also help transform and refactor your specific workloads as cloud-native applications.

AWS Application Migration Service

The Application Migration Service—which replaced the now-deprecated AWS Server Migration Service—is a tool that automates the testing and transfer of AWS-bound migrations of your non-cloud application servers. In an ideal case, the entire lift and shift process can be virtually hands-free.

You get things started by installing the AWS Replication Agent on each of the servers you want migrated. The agents will initiate replication in the background without impacting your ongoing server operations. Once your application software has reached the cloud, the service will test everything to ensure it's fully operational. Once that's confirmed, the final seamless cutover will be executed and your application will launch on AWS.

Note that CloudEndure is the AWS tool you would use to automate migrations to AWS GovCloud or China Region.

AWS Database Migration Service

AWS Database Migration Service (AWS DMS) has the same basic goals as Application Migration Service but for databases. The big-picture idea is to make your non-cloud data available to your applications in the cloud with comparable or better performance. The process is, in some ways, even simpler. Instead of installing Replication Agents, all you need to provide is access, along with your existing database endpoints.

However data can take many forms. Which database engine are you currently using? Oracle? Microsoft SQL Server? And how would you like your data used once it's in the cloud?

The Database Migration Service gives you lots of choices. You can create what they call a homogeneous migration (Oracle-to-Oracle, MariaDB-to-MariaDB, etc.). But you also could convert, say, a relational database to a data lake that lives in Amazon S3, or move an Oracle database to Amazon's own Aurora.

AWS Application Discovery Service

AWS Application Discovery Service helps you plan application migration projects by identifying which applications are running in your on-premises datacenter, how they are interconnected, and how they use your resources. This information can help you select the right migration strategy and tools for the job.

Application Discovery Service inventories your datacenter, collects information about your applications, and then creates a configuration model of your applications and their dependencies. This model can be used with the Migration Hub to plan application migration projects and track their progress.

Normally, you'll need to install Application Discovery Service Discovery Agents on your on-premises servers. The agents will then get to work gathering data about those resources and their dependencies. The data collected by Application Discovery Service is retained in a data store. You can optionally export this data into a comma separate values (CSV) file for your own planning and records.

Summary

Cloud computing is built on the ability to efficiently divide physical resources into smaller but flexible virtual units. Those units can be “rented” by businesses on a pay-as-you-go basis and used to satisfy just about any networked application and/or workflow's needs in an affordable, scalable, and elastic way.

Amazon Web Services provides reliable and secure resources that are replicated and globally distributed across a growing number of regions and availability zones. AWS infrastructure is designed to be compliant with best-practice and regulatory standards—although the Shared Responsibility Model leaves you in charge of what you place *within* the cloud.

The growing family of AWS services covers just about any digital needs you can imagine, with core services addressing compute, networking, database, storage, security, and application management and integration needs.

You manage your AWS resources from the management console, with the AWS CLI, or through code generated using an AWS SDK. Additional account-wide administration services include AWS Organizations and AWS Control Tower.

Technical and account support is available through support plans and through documentation and help forums. AWS also makes whitepapers and user and developer guides available for free as Kindle books. You can access them at this address: www.amazon.com/Amazon-Web-Services/e/B007R6MVQ6.

Migrating entire applications or individual databases and other resources to AWS is facilitated through the AWS Migration Hub. App servers can be migrated using the AWS Application Migration Service, databases through AWS Database Migration Service, and local application resources can be inventoried in preparation for a migrating using the AWS Application Discovery Service.

Exam Essentials

Understand AWS platform architecture. AWS divides its servers and storage devices into globally distributed regions and, within regions, into availability zones. These divisions permit replication to enhance availability but also permit process and resource isolation for security and compliance purposes. You should design your own deployments in ways that take advantage of those features.

Understand how to use AWS administration tools. Though you will use your browser to access the AWS administration console at least from time to time, most of your serious work will probably happen through the AWS CLI and, from within your application code, through an AWS SDK.

Know how to choose a support plan. Understanding which support plan level is right for a particular customer is an important element in building a successful deployment. Become familiar with the various options.

Understand how to migrate existing applications and resources to AWS. Moving databases, archives, or full-stack applications will often require more than a simple upload. AWS provides a number of valuable tools to help your migration.

Review Questions

1. Your developers want to run fully provisioned EC2 instances to support their application code deployments but prefer not to have to worry about manually configuring and launching the necessary infrastructure. Which of the following should they use?
 - A. AWS Lambda
 - B. AWS Elastic Beanstalk
 - C. Amazon EC2 Auto Scaling
 - D. Amazon Route 53
2. Some of your application's end users are complaining of delays when accessing your resources from remote geographic locations. Which of these services would be the most likely to help reduce the delays?
 - A. Amazon CloudFront
 - B. Amazon Route 53
 - C. Elastic Load Balancing
 - D. Amazon Glacier
3. Which of the following is the best use-case scenario for Elastic Block Store?
 - A. You need a cheap and reliable place to store files your application can access.
 - B. You need a safe place to store backup archives from your local servers.
 - C. You need a source for on-demand compute cycles to meet fluctuating demand for your application.
 - D. You need persistent storage for the filesystem run by your EC2 instance.
4. You need to integrate your company's local user access controls with some of your AWS resources. Which of the following can help you control the way your local users access your AWS services and administration console? (Choose two.)
 - A. AWS Identity and Access Management (IAM)
 - B. Key Management Service (KMS)
 - C. AWS Directory Service
 - D. Simple WorkFlow (SWF)
 - E. Amazon Cognito
5. The data consumed by the application you're planning will require more speed and flexibility than you can get from a closely defined relational database structure. Which AWS database service should you choose?
 - A. Relational Database Service (RDS)
 - B. Amazon Aurora
 - C. Amazon DynamoDB
 - D. Key Management Service (KMS)

6. You've launched an EC2 application server instance in the AWS Ireland region and you need to access it from the web. Which of the following is the correct endpoint address that you should use?
 - A. `compute.eu-central-1.amazonaws.com`
 - B. `ec2.eu-central-1.amazonaws.com`
 - C. `elasticcomputecloud.eu-west-2.amazonaws.com`
 - D. `ec2.eu-west-1.amazonaws.com`
7. When working to set up your first AWS deployment, you keep coming across the term *availability zone*. What exactly is an availability zone?
 - A. An isolated physical datacenter within an AWS region
 - B. A region containing multiple isolated datacenters
 - C. A single network subnet used by resources within a single region
 - D. A single isolated server room within a datacenter
8. As you plan your multi-tiered, multi-instance AWS application, you need a way to effectively organize your instances and configure their network connectivity and access control. Which tool will let you do that?
 - A. Load Balancing
 - B. Amazon Virtual Private Cloud (VPC)
 - C. Amazon CloudFront
 - D. AWS endpoints
9. You want to be sure that the application you're building using EC2 and S3 resources will be reliable enough to meet the regulatory standards required within your industry. What should you check?
 - A. Historical uptime log records
 - B. The AWS Program Compliance Tool
 - C. The AWS service level agreement (SLA)
 - D. The AWS Compliance Programs documentation page
 - E. The AWS Shared Responsibility Model
10. Your organization's operations team members need a way to access and administer your AWS infrastructure via your local command line or shell scripts. Which of the following tools will let them do that?
 - A. AWS Config
 - B. AWS CLI
 - C. AWS SDK
 - D. The AWS Console

11. While building a large AWS-based application, your company has been facing configuration problems they can't solve on their own. As a result, they need direct access to AWS support for both development and IT team leaders. Which support plan should you purchase?
 - A. Business
 - B. Developer
 - C. Basic
 - D. Enterprise

12. You've got a complex, multi-tiered application running on local servers that you want to migrate to the cloud. Which of these tools will provide you with the specific tools you'll need to move the application with the least risk and the least disruption?
 - A. AWS Application Migration Service
 - B. AWS Migration Hub
 - C. AWS Application Discovery Service
 - D. AWS Lift and Shift