



training and
certification

**AWS Cloud Practitioner Essentials
Student Guide
Version 2.0.2**

100-ACPEXX-20-EN-SG

© 2019 Amazon Web Services, Inc. or its affiliates. All rights reserved.

This work may not be reproduced or redistributed, in whole or in part, without prior written permission from Amazon Web Services, Inc. Commercial copying, lending, or selling is prohibited.

Corrections or feedback on the course, please email us at:

aws-course-feedback@amazon.com.

For all other questions, contact us at:

<https://aws.amazon.com/contact-us/aws-training/>.

All trademarks are the property of their owners.

Contents

Module-01 : Introduction to the AWS Cloud	4
Module-02 : Getting started with the cloud	38
Module-03 : Building in the cloud	105
Module-04 : Security	200
Module-05 : Pricing models and cloud application support	247
Module-06 : Architecture	281

AWS Cloud Practitioner Essentials



Course goals



Goals

- Value proposition
- Global infrastructure
- Key services
- Security and compliance
- Architecture
- Pricing
- Support

Audience

- Sales
- Legal
- Marketing
- Business analysts
- Project managers
- Other IT-related professionals

Course modules



1. Introduction to the AWS Cloud
2. Getting started with the cloud
3. Building in the cloud
4. Secure your cloud applications
5. Support your cloud applications
6. Architecture

Logistics



- Facility:
 - Emergency exits
 - Fire alarm protocol
 - Security
- Breaks and lunch
- Cell phones

Module 1: Introduction to the AWS Cloud



Module goals

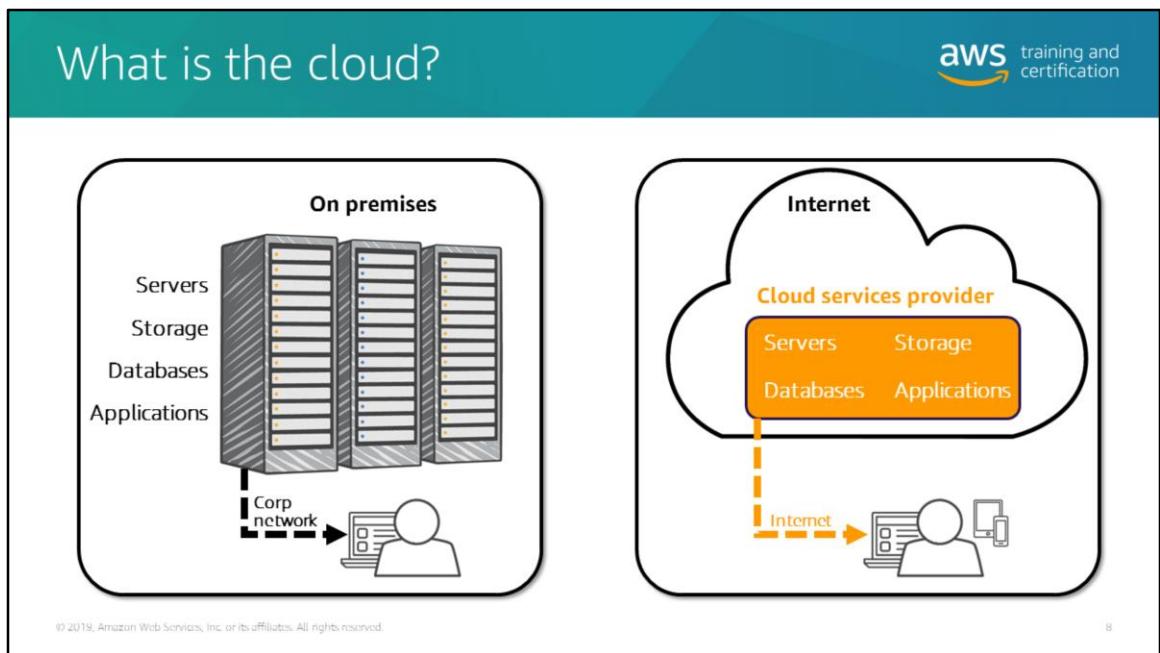


- Define the cloud
- Compare cloud vs on premises
- State the benefits of the AWS Cloud
- Identify AWS service categories
- Describe AWS physical architecture
- Interact with AWS

What is the AWS Cloud?

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.





8

In a traditional on-premises IT environment, all the physical components required to run business systems were owned, operated, maintained, and housed by the company. A user would connect and log on to the corporate network to access resources. This would include things like corporate applications, file sharing, and storage.

The cloud provides access to the same or similar resources through the internet, hosted by a cloud services platform. The term “cloud” comes from a time when internet resources or connectivity were often represented by a cloud in diagrams that illustrated connectivity and data flow.

How does it work?

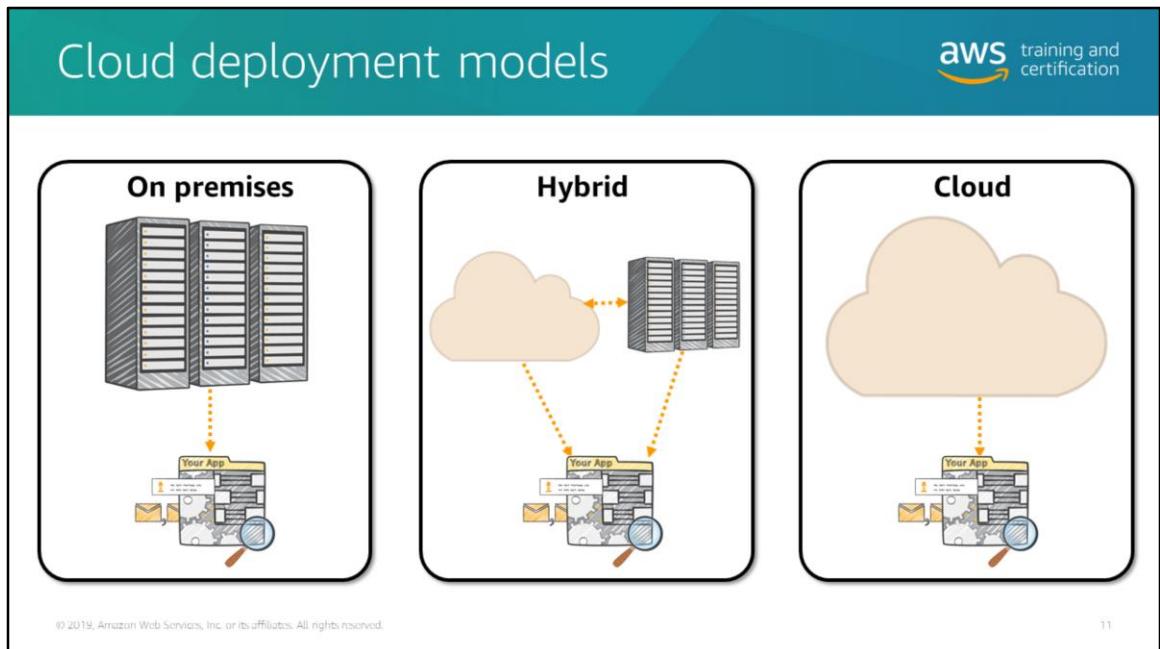
aws training and certification

- AWS owns and maintains the network-connected hardware
- You provision and use what you need

The diagram illustrates the AWS cloud computing model. On the left, there is a graphic of a world map with several interlocking gears of different colors (blue, grey, red, green) representing the global reach and interconnected nature of AWS services. An orange arrow points from this graphic to a screenshot of the AWS Management Console interface. The console shows a search bar at the top, followed by a 'Recently visited services' section with links to EC2, AWS Budgets, and S3. Below this is a 'Build a solution' section with various quick-launch options like 'Launch a virtual machine', 'Build a web app', and 'Connect an IoT device'. To the right of the console screenshot is a grid of service icons enclosed in a dashed orange border. The grid is organized into four columns: Storage (Amazon S3 icon), Database (Amazon RDS icon), Business Applications (Amazon Lambda icon), Compute (Amazon Lambda icon), Networking & Content Delivery (Amazon CloudFront icon), and Internet of Things (Amazon IoT icon). A user icon with an upward-pointing arrow is positioned below the grid, indicating interaction with the services.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cloud computing provides a simple way to access servers, storage, databases, and a broad set of application services over the internet. A cloud services platform such as Amazon Web Services (AWS) owns and maintains the network-connected hardware (like servers) that you need to run these application services, while you provision resources such as virtual servers called *instances* and manage them through a web application.



On premises/private:

Deployment of on-premises resources using virtualization and resource management tools is sometimes called a “private cloud.” On-premises deployment lacks many of the benefits of cloud computing, but is sometimes preferred for its ability to provide dedicated resources. This is like using legacy IT infrastructure with new technologies (such as virtualization and application management) to try to mimic a true cloud environment.

Hybrid:

A hybrid deployment connects infrastructure and applications between cloud-based resources and on-premises resources. Organizations use this to extend their infrastructure into the cloud while connecting cloud resources to internal systems.

For more information on hybrid deployment, see
<https://aws.amazon.com/enterprise/hybrid/>.

Cloud:

A cloud-based application is fully deployed in the cloud, and all parts of the application run 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 (as we discussed previously with IaaS) or can use higher-level services (such as SaaS) that provide abstraction from the management, architecting, and scaling requirements of core infrastructure.

What are the benefits of the AWS Cloud?

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Trade capital expense for variable expense



Data center investment
based upon forecast



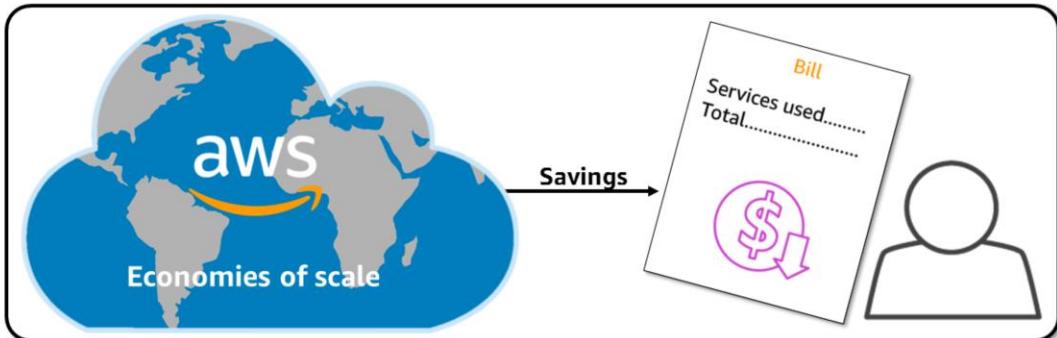
Pay only for the amount
you consume

Trade capital expense for variable expense: Instead of investing heavily in data centers (purchasing expense, running costs, and equipment within) and servers before you know how you're going to use them, you can pay only when you consume different resources and pay only for the amount you consume.

Massive economies of scale



Because of aggregate usage from all customers, AWS can achieve higher economies of scale and pass savings on to customers



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

14

Benefit from massive economies of scale: By using cloud computing, you can achieve a lower variable cost than you can 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 pay-as-you-go prices.

Stop guessing capacity

The diagram is divided into three main sections. On the left, under the heading 'Overestimated server capacity', there is an illustration of four server racks. The top two racks are grey, and the bottom two are light blue. A blue thermometer icon is positioned next to the bottom two racks, showing a low reading. In the center, under the heading 'Underestimated server capacity', there is an illustration of four server racks. The top two racks are red, and the bottom two are green. A red thermometer icon is positioned next to the top two racks, showing a high reading. To the right, under the heading 'Scaling on demand', there is a line graph titled 'Application Demand' on the y-axis and 'Time' on the x-axis. The graph shows a fluctuating blue line representing demand over time. Orange bars represent 'Instances Running', which scale up and down in response to the demand line.

Overestimated server capacity

Underestimated server capacity

Scaling on demand

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

15

Stop guessing capacity: Eliminate guessing about your infrastructure capacity needs. When you make a capacity decision before deploying an application, you often end up either sitting on expensive idle resources or dealing with limited capacity. With cloud computing, these problems go away. You can access as much or as little as you need and scale up and down as required with only a few minutes' notice.

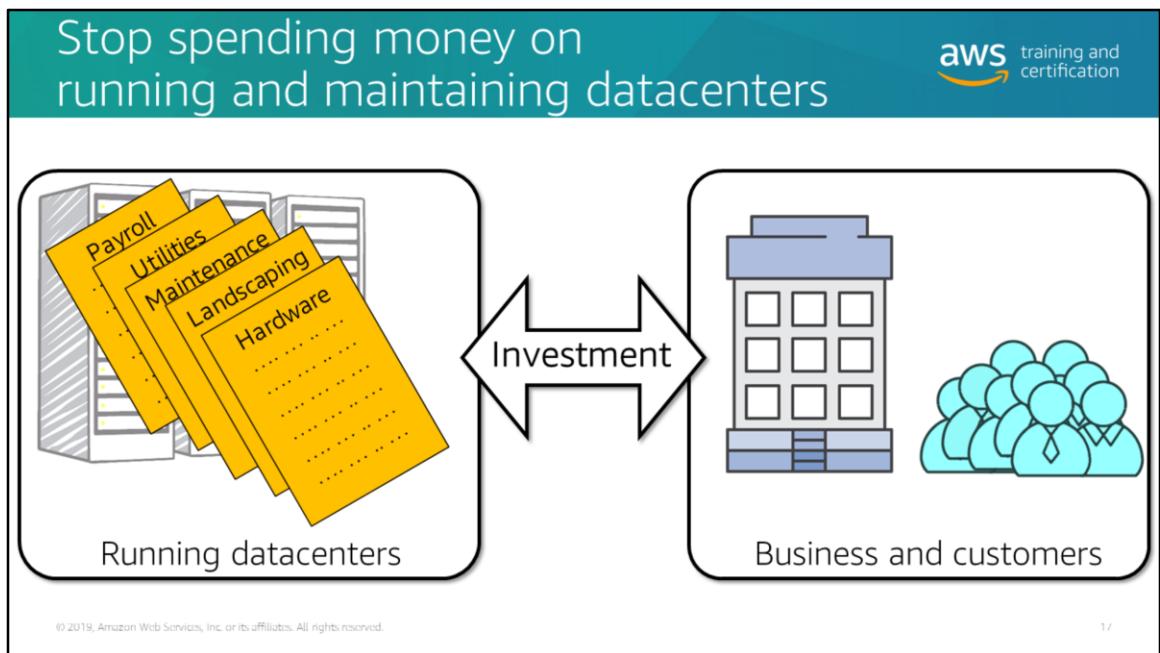
Increase speed and agility

The diagram illustrates the concept of increasing speed and agility by comparing two scenarios. On the left, a whiteboard titled 'Purchasing Request' lists a sequence of steps: Management Approval, Finance Approval, PO Received, Item Ordered, Order Pulled, Order Shipped, Order Delivered, and Order Unpacked. A red vertical line labeled 'waiting' is positioned next to the 'Order Pulled' step, with several small red dots above it, indicating a long delay. Below this box, the text reads: Weeks between wanting resources and having resources. On the right, a large blue button labeled 'Launch' has a white hand cursor icon pointing to its center. To the right of the button is a downward-pointing arrow. Below this box, the text reads: Minutes between wanting resources and having resources.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

1b

Increase speed and agility: In a cloud computing environment, new IT resources are only ever a click away, which means you reduce the time it takes to make those resources available to your developers from weeks to just minutes. This results in a dramatic increase in agility for the organization because the cost and time it takes to experiment and develop is significantly lower.

**Stop spending money on running and maintaining data centers:**

Focus on projects that differentiate your business instead of focusing on the infrastructure. Cloud computing lets you focus on your own customers instead of the heavy lifting of racking, stacking, and powering servers.

The screenshot shows the AWS Management Console homepage. At the top, there's a search bar and navigation links for 'Services' and 'Resource Groups'. Below the search bar, the 'AWS services' section lists recently visited services like EC2, S3, and Lambda. It also features a 'Build a solution' section with options like 'Launch a virtual machine', 'Build a web app', 'Connect an IoT device', and 'Start a development project'. To the right, a sidebar lists various AWS regions: US East (N. Virginia), US East (Ohio), US West (N. California), US West (Oregon), Asia Pacific (Mumbai), Asia Pacific (Osaka-Local), Asia Pacific (Seoul), **Asia Pacific (Singapore)**, Asia Pacific (Sydney), Asia Pacific (Tokyo), Canada (Central), EU (Frankfurt), EU (Ireland), EU (London), EU (Paris), and South America (São Paulo). The background of the page is a world map with colored dots representing AWS locations. Four specific regions are highlighted with callout boxes: North America (US West (Oregon)), Europe (EU (Paris)), Asia Pacific (Singapore), and South America (South America (São Paulo)).

Go global in minutes

aws training and certification

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

18

Go global in minutes: Easily deploy your application in multiple regions around the world with just a few clicks. This means you can provide a lower latency and better experience for your customers simply and at minimal cost.

We'll cover AWS Regions and Availability Zones in more depth in the next module.

AWS security



-  Keep your data safe
-  Meet compliance requirements
-  Save money
-  Scale quickly

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

19

Keep your data safe

The AWS infrastructure puts strong safeguards in place to help protect customer privacy. All data is stored in highly secure AWS data centers.

Meet compliance requirements

AWS manages dozens of compliance programs in its infrastructure. This means that segments of your compliance have already been completed.

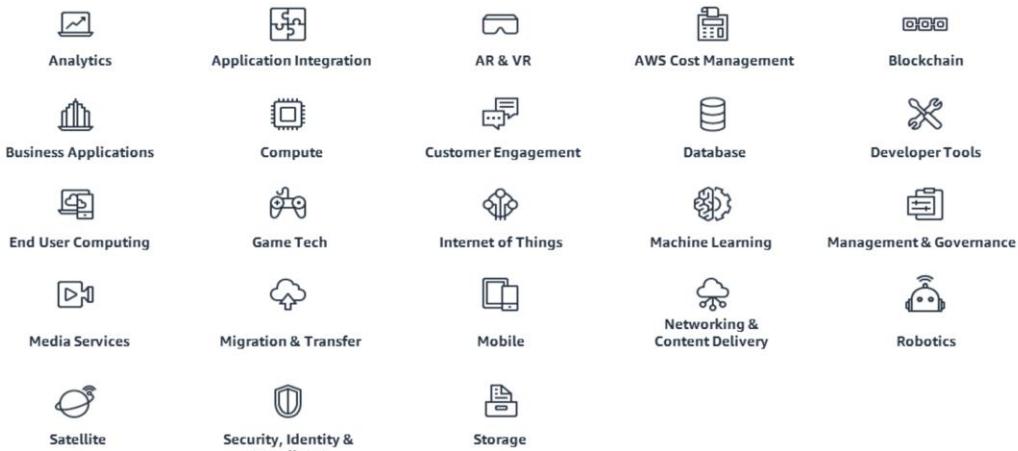
Save money

Cut costs by using AWS data centers. Maintain the highest standard of security without having to manage your own facility.

Scale quickly

Security scales with your AWS cloud usage. No matter the size of your business, the AWS infrastructure is designed to keep data safe.

AWS service categories



The grid displays 20 service categories, each with an icon and a brief description:

- Analytics**: Services supporting business intelligence.
- Application Integration**: Services supporting distributed applications.
- AR & VR**: The fastest and easiest way to create augmented reality (AR), virtual reality (VR), and 3D experiences.
- AWS Cost Management**: Tools to help you to access, organize, understand, control, and optimize your AWS costs and usage.
- Blockchain**: Create and manage scalable blockchain networks and managed ledger databases.
- Business Applications**: Services supporting distributed applications.
- Compute**: Services supporting distributed applications.
- Customer Engagement**: Services supporting customer engagement.
- Database**: Services for managing and storing data.
- Developer Tools**: Tools to help developers build, test, and deploy software.
- End User Computing**: Services supporting end-user computing.
- Game Tech**: Services supporting game technology.
- Internet of Things**: Services supporting the Internet of Things.
- Machine Learning**: Services supporting machine learning.
- Management & Governance**: Services supporting management and governance.
- Media Services**: Services supporting media delivery.
- Migration & Transfer**: Services for migrating and transferring data.
- Mobile**: Services supporting mobile applications.
- Networking & Content Delivery**: Services for networking and content delivery.
- Robotics**: Services supporting robotics.
- Satellite**: Services supporting satellite operations.
- Security, Identity & Compliance**: Services supporting security, identity, and compliance.
- Storage**: Services for managing storage.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

20

AWS offers a broad set of global cloud-based services on-demand, available in seconds, with pay-as-you-go pricing. From data warehousing to deployment tools, directories to content delivery, over 140 AWS services are available. AWS services are organized into categories which may contain just one or a whole family of services.

For your reference, a brief description of each service category is listed:

- **Analytics**
Services supporting business intelligence
- **Application integration**
Services supporting distributed applications
- **AR and VR**
The fastest and easiest way to create augmented reality (AR), virtual reality (VR), and 3D experiences
- **AWS Cost Management**
Tools to help you to access, organize, understand, control, and optimize your AWS costs and usage
- **Blockchain**
Create and manage scalable blockchain networks and managed ledger databases

- **Business Applications**
Productivity applications
- **Compute**
Virtual server hosting, container management, and serverless computing
- **Customer engagement**
Services supporting customer engagement
- **Database**
Purpose-built databases for all your application needs
- **Developer tools**
Host code and automatically build, test, and deploy your applications to AWS
- **End User Computing**
Provision virtual, cloud-based Microsoft Windows desktops for your users
- **Game Tech**
Support for game development and multiplayer game hosting
- **Internet of Things (IoT)**
Connect devices and collect, store, and analyze data from them
- **Machine learning**
Machine learning in the hands of every developer and data scientist
- **Management & Governance**
Complete control for your cloud environment
- **Media services**
Build video workflows in the cloud
- **Migration & Transfer**
Features to assist with migration of data, applications and databases
- **Mobile**
The fastest way to build apps that scale
- **Networking and content delivery**
Content delivery network, virtual private cloud, direct connections, load balancing, and DNS
- **Robotics**
Develop, test, and deploy intelligent robotics applications at scale
- **Satellite**
Fully managed ground stations as a service
- **Security, Identity, & Compliance**
Secure your environment and achieve compliance
- **Storage**
A reliable, scalable, and secure place for your data

AWS global infrastructure

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



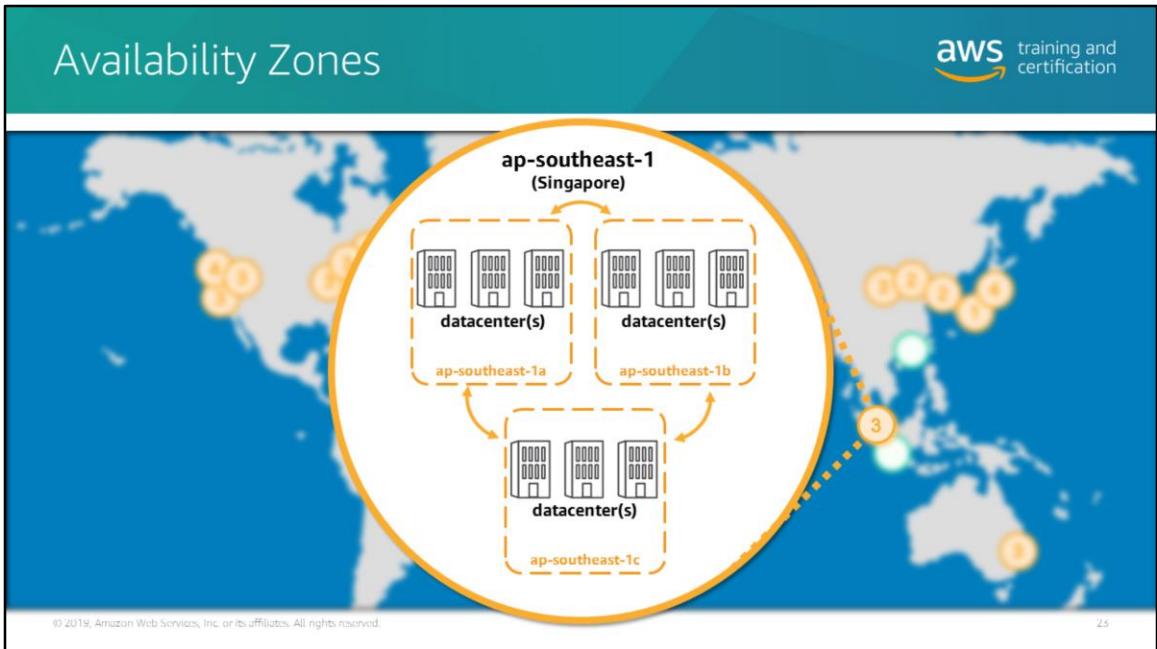


Amazon's cloud computing resources are hosted in multiple locations worldwide, which are called *AWS Regions*.

Each Region is designed to be completely isolated from the other regions to achieve the greatest possible fault tolerance and stability.

When you view your resources, you'll only see the resources tied to the Region you have specified. This is because Regions are isolated from each other, and we don't replicate resources across Regions automatically.

The AWS Cloud spans 20 geographic Regions around the world, with announced plans for five more Regions in Bahrain, Cape Town, Hong Kong SAR, Jakarta, and Milan. For more information, see <https://aws.amazon.com/about-aws/global-infrastructure/>



Each Region is a separate geographic area that has multiple locations isolated from each other, known as *Availability Zones*. Each Availability Zone is isolated, but the Availability Zones in a Region are connected through low-latency links. Where natural disasters or fault lines are a consideration, AWS isolates its zones so that they are not easily affected at the same time. For example, where earthquakes are a problem, AWS would not build two zones on the same fault line. When you launch an *instance* (which is a virtual server in the cloud), you can select a zone or let AWS choose one for you. If you distribute your instances across multiple zones, you can design your application so that if an instance fails, an instance in another zone can handle requests.

In the example on the slide, we have zoomed in on the Region named ap-southeast-1. Everything within that Region is located in Singapore. That Region consists of three distinct zones, referred to as ap-southeast-1a, b, and c. If you were to host a resource in Singapore, hosting it within two of those Availability Zones would provide sufficient high availability in a wide variety of failure scenarios.

For more information, visit:

- <http://aws.amazon.com/about-aws/global-infrastructure/regional-product-services/>

- <http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html>

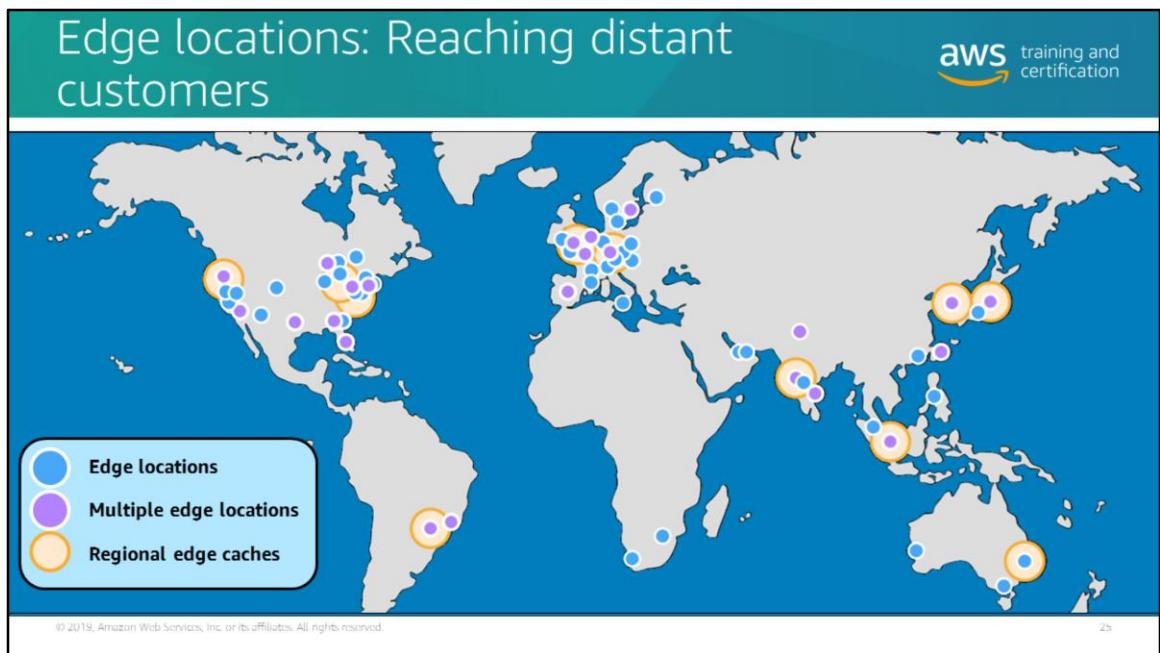
Selecting a region

Determine the right region for your services, applications, and data based on these factors

- Data governance, legal requirements
- Proximity to customers (latency)
- Services available within the region
- Costs (vary by region)

To ensure high performance for your customers, select the region closest to them, thereby minimizing latency.

- Sometimes, local laws will require that certain information be kept within the geographical boundaries of a country. Such laws may restrict the regions in which you can offer content or services.
- New AWS services tend to roll out gradually across regions. Check the Products and Services by Region chart (<http://aws.amazon.com/about-aws/global-infrastructure/regional-product-services/>) to determine whether a region offers the services required by your system.
- CloudPing (test latency between your location and all AWS regions): <http://www.cloudping.info/>.



To deliver content to end users with lower latency, AWS uses a global network of 166 points of presence (155 edge locations and 11 regional edge caches) in 65 cities across 29 countries.

Edge locations are located in North America, Europe, Asia, Australia, South America, Africa, and the Middle East. Edge locations cache copies of your content for faster delivery to users at any location. They support AWS services like Amazon Route 53 and Amazon CloudFront. We'll cover those services later in this course.

Regional edge caches, used by default with CloudFront, are used when you have content that is not accessed frequently enough to remain in an edge location. Regional edge caches absorb this content and provide an alternative to fetching that content from the origin server.

For more information, visit: <https://aws.amazon.com/cloudfront/details>

AWS management interfaces

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Three ways to interact with AWS

AWS Management Console
Easy-to-use graphical interface



Command Line Interface (AWS CLI)
Access to services by discrete command



Software Development Kits (SDKs)
Access services in your code

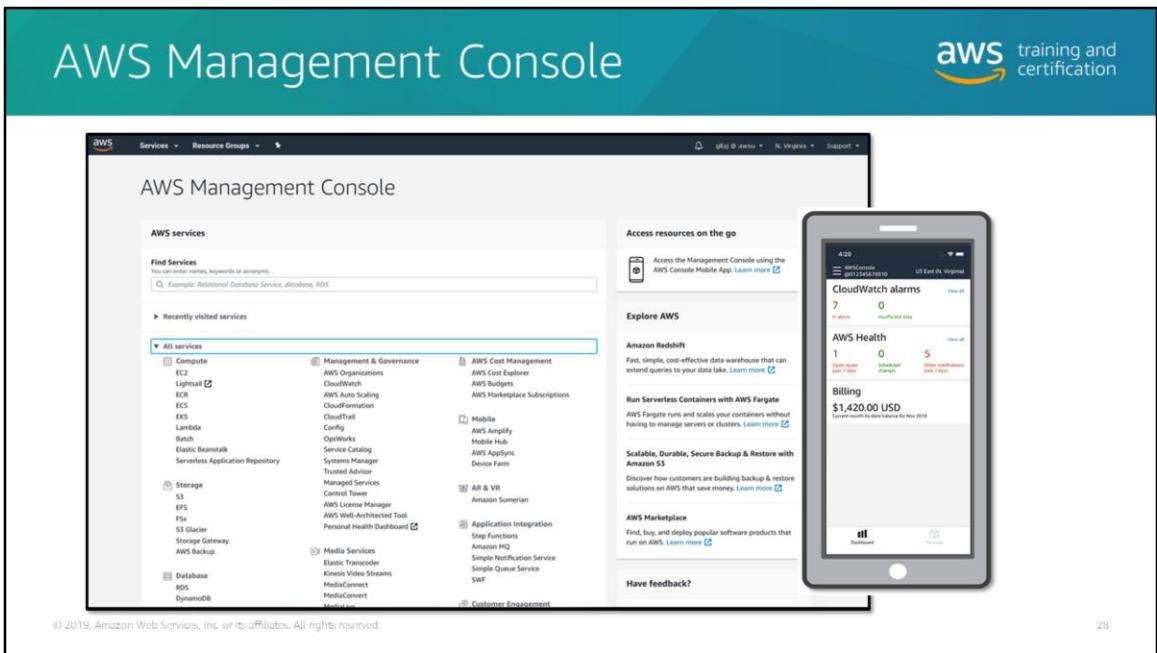


© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

2 /

AWS users can create and manage resources on the platform in three ways. All three options are built on a common REST-like API that serves as the foundation of AWS.

- 1. AWS Management Console:** The console provides a rich graphical interface to a majority of the features offered by AWS. (Note: From time to time, new features may not have all of their capabilities baked into the console when the feature initially launches.)
- 2. AWS Command Line Interface (AWS CLI):** The CLI provides a suite of utilities that can be launched from a command program in Linux, Mac or Windows.
- 3. Software Development Kits (SDKs):** AWS provides packages that enable accessing AWS in a large variety of popular programming languages. This not only makes it easy to use AWS in your existing applications but also enables creating applications to deploy and monitor complex systems entirely through code.



Administer your AWS account

The console facilitates cloud management for all aspects of your AWS account, including monitoring your monthly spending by service, managing security credentials, and even setting up new IAM Users.

Finding services in the console

There are several ways for you to locate and navigate to the services you need. On the AWS Console Home page, use the search functionality, select services from the Recently Visited Services section, or expand the All Services section to browse through the list of all the services offered by AWS.

At any time, you can also select the Services menu in the top-level navigation bar, which includes the search functionality and the list of all services, either grouped or arranged alphabetically.

Learn more About AWS

The AWS Console Home page features various types of resources to help you learn about the services and features AWS has to offer and get started with building your solutions faster.

The Build a Solution section features various simple automated wizards and

workflows that help you create the resources you need for the solution you are seeking.

The Learn to Build section shows you various aggregated learning and training resources, organized by solution type and use case. These resources include tutorials, videos, self-paced labs, project guides, and documentation.

Pin service shortcuts

Personalize your console experience by creating shortcuts to the services you visit most often. Clicking the "pin" icon in the global navigation bar (top-level toolbar) will allow you to drag and drop service links onto the toolbar to create your shortcuts.

Resource Groups

With Resource Groups, you can view collections of resources that share common tags. Streamline your use of the console by creating a resource group for each application, service, or collection of related resources that you work with regularly. Quickly navigate to each saved resource group using the AWS menu. Resource Groups are specific to each identity, so each user in an account can create unique Resource Groups for frequently accessed resources and common tasks. Users can also use a URL to share Resource Group definitions with others in the same account.

Tag Editor

Use the Tag Editor to easily manage tags for all resource types that support tags in any Region. Apply tag keys and values to multiple resources at once. The Tag Editor supports global tag search and bulk editing, so it's easy to find all resources with a particular tag or make tag changes across multiple resources with just a few clicks.

Manage AWS resources from your mobile device

With the AWS Management Console mobile app, you can quickly and easily view your existing resources, including CloudWatch alarms, and perform operational tasks from your mobile device. Download our mobile app from Amazon Appstore, Google Play, or iTunes.

AWS CLI



- Open source tool for interacting with AWS services
- Environments
 - Linux
 - MacOS
 - Windows



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

29

The CLI is an open-source tool built on top of the AWS SDK for Python (Boto) that provides commands for interacting with AWS services. With minimal configuration (just one tool to download and configure), you can start using all of the functionality provided by the console from your favorite terminal program.

Linux shells – Use common shell programs such as Bash, Zsh, and tsch to run commands in Linux, macOS, or Unix.

Windows command line – On Microsoft Windows, run commands in either PowerShell or the Windows Command Processor.

AWS SDKs

The slide displays a collection of language-specific AWS SDKs arranged in two columns. Each entry consists of a small icon followed by the language name. The languages listed are: JavaScript, Go, Python, Node.js, PHP, C++, .NET, Java, Ruby, and IoT.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

51

AWS manages infrastructure as code by using the AWS SDKs and the APIs that underlie them. The concept of infrastructure as code (IaC) is powerful and disruptive and sets the cloud apart from the old IT world. These language-specific SDKs contain APIs that allow you to easily incorporate the wide range of AWS Cloud services into your code without writing the functions yourself. There are extensive notes and documentation on how to use the SDKs listed on the slide. Some of the examples of documentation provided for you include guides on how to get started, developer guides, API references, and community forums or developer blogs.

For more information about these SDKs, visit <https://aws.amazon.com/tools/>

Key takeaways



- With a pay-as-you-go pricing, cloud services platform AWS delivers:
 - Compute power
 - Storage
 - Database services
 - Other resources
- Regions and Availability Zones are more highly available, fault tolerant, and scalable than traditional datacenter infrastructures.
- AWS supports three different management interfaces to access your account:
 - Web-based AWS Management Console (mobile app as well)
 - CLI
 - SDKs

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

52

For more information, visit <https://aws.amazon.com>.

Module 2: Getting started with the cloud



Module goals



- Get started with AWS services
- Build your infrastructure
- Store your data
- Secure your data

Get started with AWS services

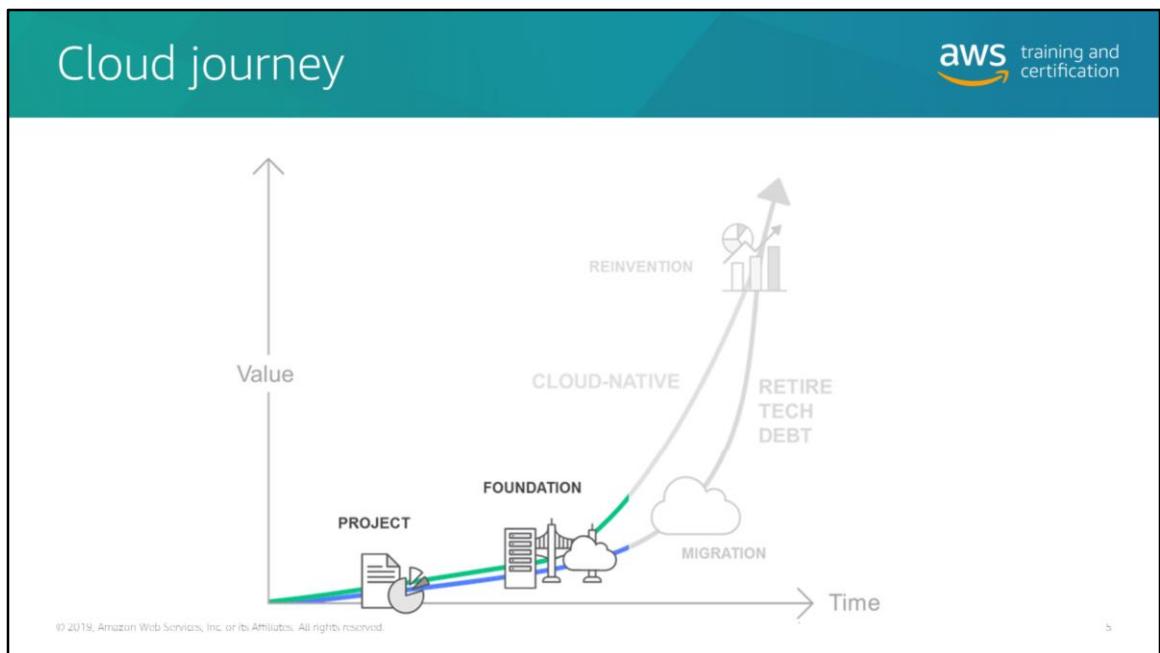
© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



The screenshot shows the AWS products landing page. At the top, there's a teal header with the AWS logo and "aws training and certification". Below the header, the main title "AWS products" is displayed. The page features a dark background with several service highlights. One prominent highlight is "AWS Deep Learning Containers", which allows users to quickly set up deep learning environments with optimized Docker images. Other visible highlights include "Amazon Lightsail" (a low-cost, predictable price), "Amazon EC2 M5ad & R5ad Instances" (10% lower cost), "Amazon S3 Glacier Deep Archive" (secure, durable object storage), and "110,000+ Databases Migrated to AWS" (saving time and costs). A "Customer News" section mentions Volkswagen's plans to build the Volkswagen Industrial Cloud. At the bottom, there's a "Explore Our Products" button and a copyright notice: "© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved."

AWS offers a broad set of global cloud-based products including compute, storage, databases, analytics, networking, mobile, developer tools, management tools, IoT, security and enterprise applications. These services help organizations move faster, lower IT costs, and scale. Within each product is a broad set of infrastructure services that you can utilize for your business requirements.

AWS provides a customizable platform for virtually every use case. With just a few mouse clicks, you can provision new services, without upfront capital expense, and build what infrastructure suits your business needs.



Throughout this course we are going to be exploring what AWS products/services are available to you that might be the right solution for your needs. We will begin with the project and foundational tools that are necessary to start your journey. As we continue on, we will incorporate products/services that will assist you with your migration and finally we will discuss the reinvention of your business needs along with the cloud.

Build your infrastructure

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



To begin your build, let's discuss our featured services that can be used in your infrastructure.

What is Amazon EC2?

The diagram illustrates the transition from physical hardware to cloud computing. On the left, three dark grey server racks are labeled "On-premises servers". To their right is a list of server types, each preceded by a checkmark. On the far right is a grid of 16 smaller server icons, with a diagonal banner across it reading "Amazon EC2 instances".

- ✓ Application server
- ✓ Web server
- ✓ Database server
- ✓ Game server
- ✓ Mail server
- ✓ Media server
- ✓ Catalog server
- ✓ File server
- ✓ Computing server
- ✓ Proxy server

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon EC2 provides secure, resizable compute capacity in the cloud. It's designed to make web-scale cloud computing easier for developers.

Compute capacity means functionality traditionally provided by virtual or on-premises, physical servers. You get the same functionality as you would from a physical server but with the benefits of hosting it in the cloud.

Benefits of Amazon EC2

aws training and certification

- Elasticity

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

8

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Benefits of Amazon EC2

The screenshot shows the AWS Cloud9 interface. On the left, a callout box lists two benefits: 'Elasticity' and 'Control'. On the right, a screenshot of the AWS Management Console EC2 service shows an instance named 'i-0cb158fb655'. A context menu is open over the instance, with 'Stop' highlighted. The AWS logo is visible in the top right corner.

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Benefits of Amazon EC2

Family	Type	vCPUs	Memory (GB)	Instance Storage	EBS-Optimized Available	Network Performance	IPv6 Support
Compute optimized	c5d.large	2	4	1 x 50 (SSD)	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5d.xlarge	4	8	1 x 100 (SSD)	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5d.2xlarge	8	16	1 x 200 (SSD)	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5d.4xlarge	16	32	1 x 400 (SSD)	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5d.8xlarge	36	72	1 x 900 (SSD)	Yes	10 Gigabit	Yes
Compute optimized	c5d.16xlarge	72	144	2 x 900 (SSD)	Yes	25 Gigabit	Yes
Compute optimized	c5.large	2	4	EBS only	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5.xlarge	4	8	EBS only	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5.2xlarge	8	16	EBS only	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5.4xlarge	16	32	EBS only	Yes	Up to 10 Gigabit	Yes
Compute optimized	c5.8xlarge	36	72	EBS only	Yes	10 Gigabit	Yes
Compute optimized	c5.16xlarge	72	144	EBS only	Yes	25 Gigabit	Yes
Compute optimized	c4.large	2	3.75	EBS only	Yes	Moderate	Yes

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

10

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Benefits of Amazon EC2

The diagram illustrates the architecture of an Amazon EC2 setup. It shows a flow from external domains (www.example.com and media.example.com) through CloudFront distributions and Amazon S3 buckets, down to logs in an Amazon EBS snapshot, and finally to an Auto Scaling group within an Availability Zone. The Auto Scaling group contains an EC2 instance security group with a web app server and a security group. The instance is connected to root and data volumes.

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

11

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Benefits of Amazon EC2



- Elasticity
- Control
- Flexibility
- Integrated
- Reliable



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

12

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

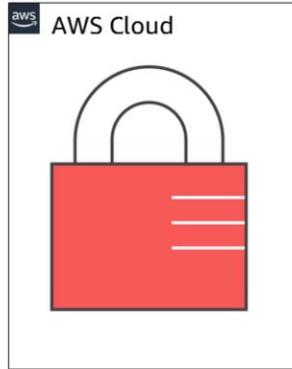
Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Benefits of Amazon EC2



- Elasticity
- Control
- Flexibility
- Integrated
- Reliable
- Secure



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Benefits of Amazon EC2



- Elasticity
- Control
- Flexibility
- Integrated
- Reliable
- Secure
- Inexpensive



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

14

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

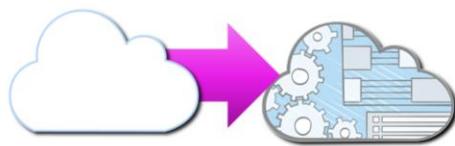
Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Benefits of Amazon EC2



- Elasticity
- Control
- Flexibility
- Integrated
- Reliable
- Secure
- Inexpensive
- Easy



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

15

Elastic web-scale computing

Amazon EC2 enables you to increase or decrease capacity within minutes, not hours or days. You can commission one, hundreds, or even thousands of server instances simultaneously. You can also use Amazon EC2 Auto Scaling to maintain availability of your Amazon EC2 fleet and automatically scale your fleet up and down depending on your needs to maximize performance and minimize cost. To scale multiple services, you can use AWS Auto Scaling.

Complete control

You have complete control of your instances, including root access, and the ability to interact with them as you would any machine. You can stop any instance while retaining the data on the boot partition and then subsequently restart the same instance using web service APIs. Instances can be rebooted remotely using web service APIs, and you also have access to their console output.

Flexible cloud hosting services

You have the choice of multiple instance types, operating systems, and software packages. Amazon EC2 allows you to select a configuration of memory, CPU, instance storage, and the boot partition size that is optimal for your choice of operating

system and application. For example, choice of operating systems includes numerous Linux distributions and Microsoft Windows Server.

Integrated

Amazon EC2 is integrated with most AWS services (including Amazon S3, Amazon RDS, and Amazon VPC) to provide a complete, secure solution for computing, query processing, and cloud storage across a wide range of applications.

Reliable

Amazon EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned. The service runs within Amazon's proven network infrastructure and data centers.

Secure

Security is our highest priority. As an AWS customer, you will benefit from a data center and network architecture built to meet the requirements of the most security-sensitive organizations. Amazon EC2 works in conjunction with Amazon VPC to provide security and robust networking functionality for your compute resources.

Inexpensive

Using Amazon EC2 lets you take advantage of Amazon's scale—it enables you to pay a very low rate for the compute capacity that you actually consume. For more information, see <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/instance-purchasing-options.html>.

Easy to get started

To get started with Amazon EC2, use the AWS Management Console, the AWS CLI, or the AWS SDKs. AWS is free to get started.

Choosing the right Amazon EC2 instances



- EC2 Instance types are optimized for different use cases, workloads & come in multiple sizes. This allows you to optimally scale resources to your workload requirements.
- AWS utilizes Intel® Xeon® processors for EC2 Instances providing customers with high performance and value.
- Consider the following when choosing your instances: core count, memory size, storage size & type, network performance, I/O requirements & CPU technologies.
- Hurry Up & Go Idle - A larger compute instance can save you time and money, therefore paying more per hour for a shorter amount of time can be less expensive.

EC2 instances powered by Intel Technologies											
EC2 instance type	Compute optimized		General purpose			Memory optimized			Storage optimized		
	C5	C4	M5	M4	T2	X1	X1e	R4	I11	I3	D2
Intel processor	Xeon Platinum 8175M	Xeon E5 2666 v3	Xeon Platinum 8175M	Xeon E5 2686 v4	Xeon Family	Xeon E7 8880 v3	Xeon E7 8880 v3	Xeon E5 2686 v4	Xeon E5 2686 v4	Xeon E5 2686 v4	Xeon E5 2676 v3
Intel processor technology	Skylake	Haswell	Skylake	Broadwell-Haswell	Yes	Haswell	Haswell	Broadwell	Broadwell	Broadwell	Haswell
Intel AVX	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intel AVX2	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes
Intel AVX-512	Yes	-	Yes	-	-	-	-	-	-	-	-
Intel turbo boost	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Storage	EBS-only	EBS-only	EBS-only	EBS-only	EBS-only	SSD EBS-Opt	SSD EBS-Opt	-	HDD	SSD	HDD

C5: Compute-optimized instances

aws training and certification

25% price/performance improvement over C4

Based on 3.0 GHz Intel Xeon Scalable Processors (Skylake)

- Up to 72 vCPUs and 144 GiB of memory (2:1 Memory:vCPU ratio)
- 25 Gbps NW bandwidth
- Support for Intel AVX-512

NETFLIX "We saw significant performance improvement on Amazon EC2 C5, with up to a 140% performance improvement in industry standard CPU benchmarks over C4."

GRAIL "We are eager to migrate onto the AVX-512 enabled c5.18xlarge instance size... We expect to decrease the processing time of some of our key workloads by more than 30%."

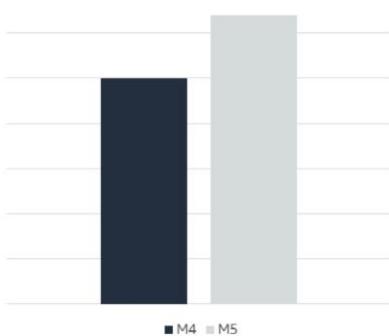
C5 instances are designed for compute-heavy applications, such as batch processing, distributed analytics, high-performance computing (HPC), ad serving, highly scalable multiplayer gaming, and video encoding. The new instances offer a 25 percent price and performance improvement over the C4 instances, with over 50 percent improvement for some workloads. They also have additional memory per vCPU, and (for code that can make use of the new [AVX-512](#) instructions) twice the performance for vector and floating point workloads.

Over the years, we have worked nonstop to provide our customers with the best possible networking, storage, and compute performance with a long-term focus on offloading many types of work to dedicated hardware designed and built by AWS. The C5 instance type incorporates the latest generation of our hardware offloads and also takes another big step forward with the addition of a new hypervisor that runs in close connection with our hardware. The new hypervisor allows us to give you access to all of the processing power provided by the host hardware while also making performance even more consistent and further raising the bar on security.

M5: Next-gen general purpose instances



14% price/performance improvement With M5



- Powered by 2.5 GHz Intel Xeon Scalable Processors (Skylake)
- New larger instance size—m5.24xlarge with 96 vCPUs and 384 GiB of memory (4:1 Memory:vCPU ratio)
- Improved network and EBS performance on smaller instance sizes
- Support for Intel AVX-512 offering up to twice the performance for vector and floating point workloads



The general-purpose (M) instances go all the way back to 2006 when we launched the m1.small. We continued to evolve along this branch of our family tree, launching the M2 (2009), M3 (2012), and the M4 (2015) instances. Our customers use the general-purpose instances to run web and app servers, host enterprise applications, support online games, and build cache fleets.

Available now, M5 instances are the next generation of Amazon EC2 General Purpose instances, powered by 2.5 GHz Intel Xeon Platinum 8000 series (Skylake-SP) processors. With enhanced networking and a new larger instance size that provides up to 96 vCPUs and 384 GiB of memory, M5 instances have up to 50 percent more vCPUs, 50 percent more memory, and 25 percent more network bandwidth than M4, making them ideal for web and application servers, backend enterprise applications, gaming servers, caching fleets, and application development environments.

The screenshot shows the AWS CloudFormation interface with the title "What's your platform?". Under "Step 1: Choose an Amazon Machine Image (AMI)", it says: "An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs." A search bar at the top right contains the placeholder "Search for an AMI by entering a search term e.g. 'Windows'". Below the search bar is a sidebar with "Quick Start" sections: "My AMIs", "AWS Marketplace", "Community AMIs", and a "Free tier only" section which is currently collapsed. The main area displays three AMI options:

- Amazon Linux 2 AMI (HVM), SSD Volume Type - ami-0d1000aff9a9bad89**
Amazon Linux 2 comes with five years support. It provides Linux kernel 4.14 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1, and the latest software packages through extras.
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Select button (64-bit)
- Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type - ami-a0cfeed8**
The Amazon Linux AMI is an EBS-backed, AWS-supported image. The default image includes AWS command line tools, Python, Ruby, Perl, and Java. The repositories include Docker, PHP, MySQL, PostgreSQL, and other packages.
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Select button (64-bit)
- Red Hat Enterprise Linux 7.5 (HVM), SSD Volume Type - ami-28e07e50**
Red Hat Enterprise Linux version 7.5 (HVM), EBS General Purpose (SSD) Volume Type
Root device type: ebs Virtualization type: hvm ENA Enabled: Yes
Select button (64-bit)

At the bottom left is the copyright notice "© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved." and at the bottom right is the page number "22".

An Amazon Machine Image (AMI) provides the information required to launch an instance. You must specify a source AMI when you launch an instance. You can launch multiple instances from a single AMI when you need multiple instances with the same configuration. You can use different AMIs to launch instances when you need instances with different configurations.

An AMI includes:

- A template for the root volume for the instance (for example, an operating system, an application server, or applications).
- Launch permissions that control which AWS accounts can use the AMI to launch instances.
- A block device mapping that specifies the volumes to attach to the instance when it's launched.

For more information about AMIs, visit

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AMIs.html>.

You can select an AMI provided by AWS, our user community, or the AWS Marketplace, or you can select one of your own AMIs.

Knowledge check



What are the benefits of using Amazon EC2 instances compared to physical servers in your infrastructure? (Select all that apply)

- A. Resizable
- B. ~~The ability to hot-add additional RAM~~
- C. Automatic automated backups
- D. Pay only for the capacity you use
- E. The ability to have different storage requirements

Knowledge check



What are the benefits of using Amazon EC2 instances compared to physical servers in your infrastructure? (Select all that apply)

- A. Resizable
- B. The ability to hot-add additional RAM
- C. Automatic automated backups
- D. Pay only for the capacity you use
- E. The ability to have different storage requirements

A, D, E are correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

24

Store your data

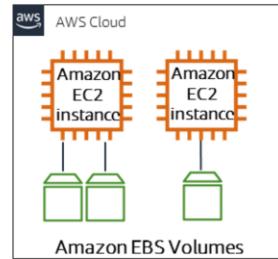
© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Amazon Elastic Block Store (Amazon EBS)



- Persistent block storage for instances



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

2b

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Different drive types are available depending upon your specific needs.

Solid State Drives (SSD)

Provisioned IOPS SSD (io1) volumes

General Purpose SSD (gp2) volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) volumes

Cold HDD (sc1) volumes

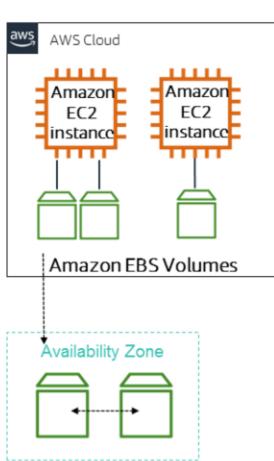
With Amazon EBS, you can scale your usage up or down within minutes, all while paying a low price for only what you provision.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to

different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted EBS volumes at no additional cost. The encryption occurs on the EC2 side, so the data moving between the EC2 instance and the EBS volume inside AWS data centers will be encrypted both in transit and at rest with EBS volume encryption.

Amazon Elastic Block Store (Amazon EBS)



- Persistent block storage for instances
- Protected through replication

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

2 /

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Different drive types are available depending upon your specific needs.

Solid State Drives (SSD)

Provisioned IOPS SSD (io1) volumes

General Purpose SSD (gp2) volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) volumes

Cold HDD (sc1) volumes

With Amazon EBS, you can scale your usage up or down within minutes, all while paying a low price for only what you provision.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to

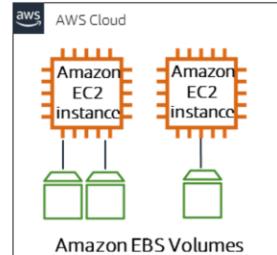
different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted EBS volumes at no additional cost. The encryption occurs on the EC2 side, so the data moving between the EC2 instance and the EBS volume inside AWS data centers will be encrypted both in transit and at rest with EBS volume encryption.

Amazon Elastic Block Store (Amazon EBS)



- Persistent block storage for instances
- Protected through replication
- Different drive types



Solid State Drives (SSD)

Provisioned IOPS SSD (io1) Volumes
General Purpose SSD (gp2) Volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) Volumes
Cold HDD (sc1) Volumes

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

28

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Different drive types are available depending upon your specific needs.

Solid State Drives (SSD)

Provisioned IOPS SSD (io1) volumes
General Purpose SSD (gp2) volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) volumes
Cold HDD (sc1) volumes

With Amazon EBS, you can scale your usage up or down within minutes, all while paying a low price for only what you provision.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to

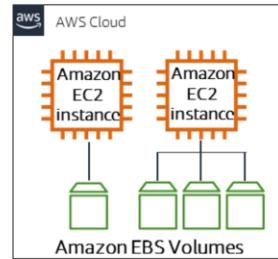
different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted EBS volumes at no additional cost. The encryption occurs on the EC2 side, so the data moving between the EC2 instance and the EBS volume inside AWS data centers will be encrypted both in transit and at rest with EBS volume encryption.

Amazon Elastic Block Store (Amazon EBS)



- Persistent block storage for instances
- Protected through replication
- Different drive types
- Scale up or down in minutes



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

29

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Different drive types are available depending upon your specific needs.

Solid State Drives (SSD)

Provisioned IOPS SSD (io1) volumes

General Purpose SSD (gp2) volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) volumes

Cold HDD (sc1) volumes

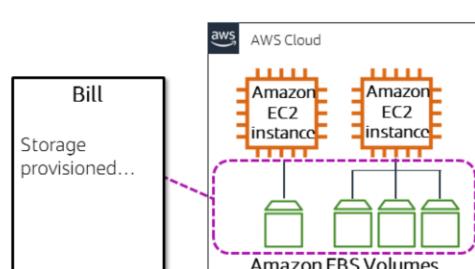
With Amazon EBS, you can scale your usage up or down within minutes, all while paying a low price for only what you provision.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to

different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted EBS volumes at no additional cost. The encryption occurs on the EC2 side, so the data moving between the EC2 instance and the EBS volume inside AWS data centers will be encrypted both in transit and at rest with EBS volume encryption.

Amazon Elastic Block Store (Amazon EBS)



- Persistent block storage for instances
- Protected through replication
- Different drive types
- Scale up or down in minutes
- Pay for only what you provision

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Different drive types are available depending upon your specific needs.

Solid State Drives (SSD)

Provisioned IOPS SSD (io1) volumes

General Purpose SSD (gp2) volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) volumes

Cold HDD (sc1) volumes

With Amazon EBS, you can scale your usage up or down within minutes, all while paying a low price for only what you provision.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to

different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted EBS volumes at no additional cost. The encryption occurs on the EC2 side, so the data moving between the EC2 instance and the EBS volume inside AWS data centers will be encrypted both in transit and at rest with EBS volume encryption.

Amazon Elastic Block Store (Amazon EBS)

The diagram illustrates the Amazon EBS architecture. It shows two Amazon EC2 instances connected to four Amazon EBS volumes. A dashed purple box encloses the instances and volumes, labeled "Storage provisioned...". To the right, a separate box shows a vertical stack of five snapshots labeled "Monday's snapshot" through "Friday's snapshot".

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Different drive types are available depending upon your specific needs.

Solid State Drives (SSD)

Provisioned IOPS SSD (io1) volumes

General Purpose SSD (gp2) volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) volumes

Cold HDD (sc1) volumes

With Amazon EBS, you can scale your usage up or down within minutes, all while paying a low price for only what you provision.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to

different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted EBS volumes at no additional cost. The encryption occurs on the EC2 side, so the data moving between the EC2 instance and the EBS volume inside AWS data centers will be encrypted both in transit and at rest with EBS volume encryption.

Amazon Elastic Block Store (Amazon EBS)

The diagram shows the AWS Cloud environment. Two Amazon EC2 instances are connected to a group of four Amazon EBS Volumes. A callout from a box labeled "Bill" says "Storage provisioned...". A purple dashed line encloses the EC2 instances and the EBS volumes. Below this, a vertical stack of five green containers represents snapshots, labeled from top to bottom: "Monday's snapshot", "Tuesday's snapshot", "Wednesday's snapshot", "Thursday's snapshot", and "Friday's snapshot". Arrows point from the EC2 instances to the EBS volumes, and from the EBS volumes to the snapshots.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5.2

Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure, offering high availability and durability. Different drive types are available depending upon your specific needs.

Solid State Drives (SSD)

Provisioned IOPS SSD (io1) volumes

General Purpose SSD (gp2) volumes

Hard Disk Drives (HDD)

Throughput Optimized HDD (st1) volumes

Cold HDD (sc1) volumes

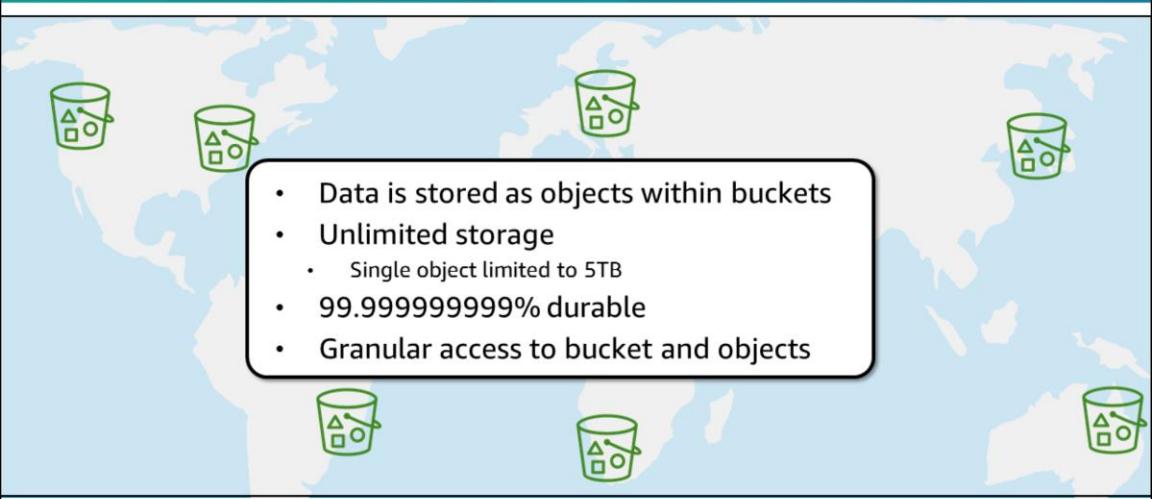
With Amazon EBS, you can scale your usage up or down within minutes, all while paying a low price for only what you provision.

To provide an even higher level of data durability, Amazon EBS gives you the ability to create point-in-time snapshots of your volumes, and AWS allows you to recreate a new volume from a snapshot at any time. Share snapshots or even copy snapshots to

different AWS Regions for even greater disaster recovery (DR) protection. You can, for example, encrypt and share your snapshots from Virginia to Tokyo.

You could also have encrypted EBS volumes at no additional cost. The encryption occurs on the EC2 side, so the data moving between the EC2 instance and the EBS volume inside AWS data centers will be encrypted both in transit and at rest with EBS volume encryption.

What is Amazon S3?



- Data is stored as objects within buckets
- Unlimited storage
 - Single object limited to 5TB
- 99.99999999% durable
- Granular access to bucket and objects

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5.5

Amazon S3 stores data as objects in resources called *buckets*. You can store as many objects as you want in a bucket, and write, read, and delete objects in your bucket. Objects can be up to 5 terabytes in size.

You can control access to both the bucket and the objects—for example, controlling who can create, delete, and retrieve objects in the bucket. You can also view access logs for the bucket and its objects, and [choose the AWS Region](#) where a bucket is stored to optimize for latency, minimize costs, or address regulatory requirements.

Amazon S3 runs on the world's largest global cloud infrastructure, and is designed to deliver 99.99999999% durability.

Amazon S3 core functionality

aws training and certification

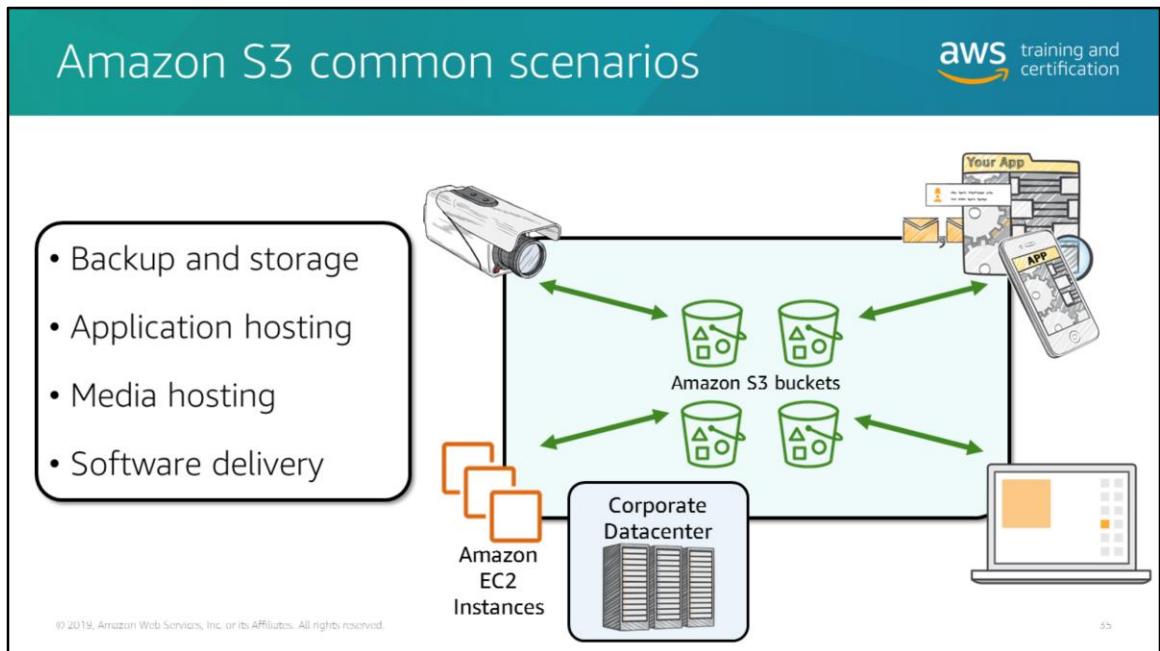
- Fast, durable, highly available key-based access to objects
- Object storage built to store and retrieve data
- Not a file system

The diagram shows a flow from a local application to an Amazon S3 bucket. On the left, a box labeled 'Your App' contains a magnifying glass icon over a document, with a tooltip: 'My App interacts with the AWS services via their APIs'. A horizontal dashed arrow points from the app to the right, labeled 'CLI sends GET request via S3 API →'. On the right, a green icon of a bucket is labeled 'Amazon S3 bucket'. A return arrow labeled '← Object returned' points back from the bucket to the app.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

54

The term “bucket” is a hint to the different approach that Amazon S3 takes with file storage. Instead of a “folder,” you create a “bucket,” which is a large space that can hold any number of objects.



Backup and storage – Provide data backup and storage services for others

Application hosting – Provide services that deploy, install, and manage web applications

Media hosting – Build a redundant, scalable, and highly available infrastructure that hosts video, photo, or music uploads and downloads

Software delivery – Host your software applications that customers can download

Not just a storage bucket



Requester pays



Versioning



Hosting static websites



Object lifecycle management

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5b

Requester Pays Buckets

With Requester Pays buckets, the requester instead of the bucket owner pays the cost of the request and the data download from the bucket. The bucket owner always pays the cost of storing data.

Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures.

Hosting Static Websites

You can host a static website on Amazon Simple Storage Service (Amazon S3). On a static website, individual webpages include static content. They might also contain client-side scripts. By contrast, a dynamic website relies on server-side processing, including server-side scripts such as PHP, JSP, or ASP.NET. Amazon S3 does not support server-side scripting. AWS also has resources for hosting dynamic websites.

Object Lifecycle Management

Lifecycle management includes expiring objects and archiving objects (transitioning objects to the Amazon S3 Glacier storage class). While managing your object through their lifecycle, you can move objects to the storage class best suited for its needs and data access patterns as well. You can use the two types of lifecycle actions such as transition (define when objects transition to another storage class) and expiration actions (define when your objects expire and then Amazon S3 deletes those expired objects).

What is Amazon S3 Glacier?



- Low-cost data archiving and long-term backup
- Can configure lifecycle archiving of Amazon S3 content to Amazon Glacier
- Retrieval Options:
 - Standard: 3- to 5-hours
 - Bulk: 5-12 hours
 - Expedited: 1 – 5 minutes



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5 /

Amazon S3 Glacier is a secure, durable, and extremely low-cost cloud storage service for data archiving and long-term backup. It is designed to deliver 99.999999999% durability, and provides comprehensive security and compliance capabilities that can help meet even the most stringent regulatory requirements. Amazon S3 Glacier provides query-in-place functionality, allowing you to run powerful analytics directly on your archive data at rest. To keep costs low yet suitable for varying retrieval needs, Amazon S3 Glacier provides three options for access to archives, from a few minutes to several hours.

For more information, see: <https://aws.amazon.com/glacier/>.

Expedited retrievals allow you to quickly access your data when occasional urgent requests for a subset of archives are required. For all but the largest archives (250MB+), data accessed using Expedited retrievals are typically made available within 1 – 5 minutes. There are two types of Expedited retrievals: On-Demand and Provisioned. On-Demand requests are like EC2 On-Demand instances and are available the vast majority of the time. Provisioned requests are guaranteed to be available when you need them.

Note that Amazon S3 also has a Glacier Deep Archive storage class, which is an even more cost-effective way to store important, infrequently accessed data in Amazon S3. Amazon S3 Glacier Deep Archive has a retrieval time of within 12 hours. More information about different Amazon S3 storage classes and pricing can be found online at <https://aws.amazon.com/s3/pricing/>.

Amazon S3 Glacier use cases

aws training and certification

-  Media asset workflows
-  Healthcare information archiving
-  Regulatory and compliance archiving
-  Scientific data storage
-  Digital preservation
-  Magnetic tape replacement

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

58

Media asset workflows

Media assets such as video and news footage require durable storage and can grow to many petabytes over time. Amazon S3 Glacier allows you to archive older media content affordably, then move it to Amazon S3 for distribution when needed.

Healthcare information archiving

Hospital systems need to retain petabytes of patient records (LIS, PACS, EHR, etc.) for decades to meet regulatory requirements. Amazon S3 Glacier helps you reliably archive patient record data securely at a very low cost.

Regulatory and compliance archiving

Many enterprises, like those in financial services and healthcare, must retain regulatory and compliance archives for extended durations. Amazon S3 Glacier Vault Lock helps you set compliance controls to meet your compliance objectives, such as SEC Rule 17a-4(f).

Scientific data storage

Research organizations generate, analyze, and archive vast amounts of data. With Amazon S3 Glacier, you avoid the complexities of hardware and facility management

and capacity planning.

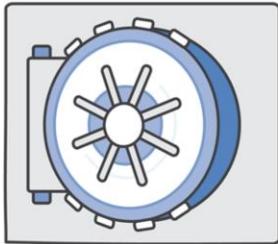
Digital preservation

Libraries and government agencies face data-integrity challenges in their digital preservation efforts. Unlike traditional systems, which can require laborious data verification and manual repair, Amazon S3 Glacier performs regular, systematic data integrity checks and is built to be automatically self-healing.

Magnetic tape replacement

On-premises or offsite tape libraries can lower storage costs but require large upfront investments and specialized maintenance. Amazon S3 Glacier has no upfront cost and eliminates the cost and burden of maintenance.

Amazon S3 Glacier vault lock policy



- Deploy and enforce compliance controls on individual Amazon Glacier vaults
- Vault becomes immutable once locked

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

59

Vault Lock allows you to easily deploy and enforce compliance controls on individual Amazon S3 Glacier vaults via a lockable policy (the Vault Lock policy). Once locked, the Vault Lock policy becomes immutable and Amazon S3 Glacier will enforce the prescribed controls to help achieve your compliance objectives. To learn more, see the Amazon S3 Glacier developer's guide.

Vault Lock supports the SEC 17a-4(f)(2)(i) and CFTC 1.31(c) requirement for notifying regulators.

Amazon S3 storage classes		
Storage class	Features	
S3 Standard	<ul style="list-style-type: none"> ≥3 availability zones 	
S3 Standard - Infrequent Access (IA)	<ul style="list-style-type: none"> Retrieval fee associated with objects Most suitable for infrequently accessed data 	
S3 Intelligent-Tiering	<ul style="list-style-type: none"> Automatically moves objects between tiers based on access patterns ≥3 availability zones 	
S3 One Zone-IA	<ul style="list-style-type: none"> 1 availability zone Costs 20% less than S3 Standard-IA 	
S3 Glacier	<ul style="list-style-type: none"> Not available for real-time access Must restore objects before you can access them Restoring objects can take 1 minute - 12 hours 	
S3 Glacier Deep Dive	<ul style="list-style-type: none"> Lowest cost storage for long term retention (7-10 years) ≥3 availability zones Retrieval time within 12 hours 	

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Each object in Amazon S3 has a storage class associated with it.

All the storage classes are designed for durability of 99.999999999% (11 9's) of objects across multiple availability zones (except S3 One Zone-IA, that's in a single availability zone).

S3 Standard is ideal for performance-sensitive use cases and frequently used data. Standard is the default storage class in S3.

S3 Intelligent-Tiering is designed to optimize costs by automatically moving data to the most cost-effective access tier, without performance impact or operational overhead.

S3 Infrequent Access (IA) is optimized for long-lived and less frequently accessed data, such as backups and older data that are accessed less but still require high performance.

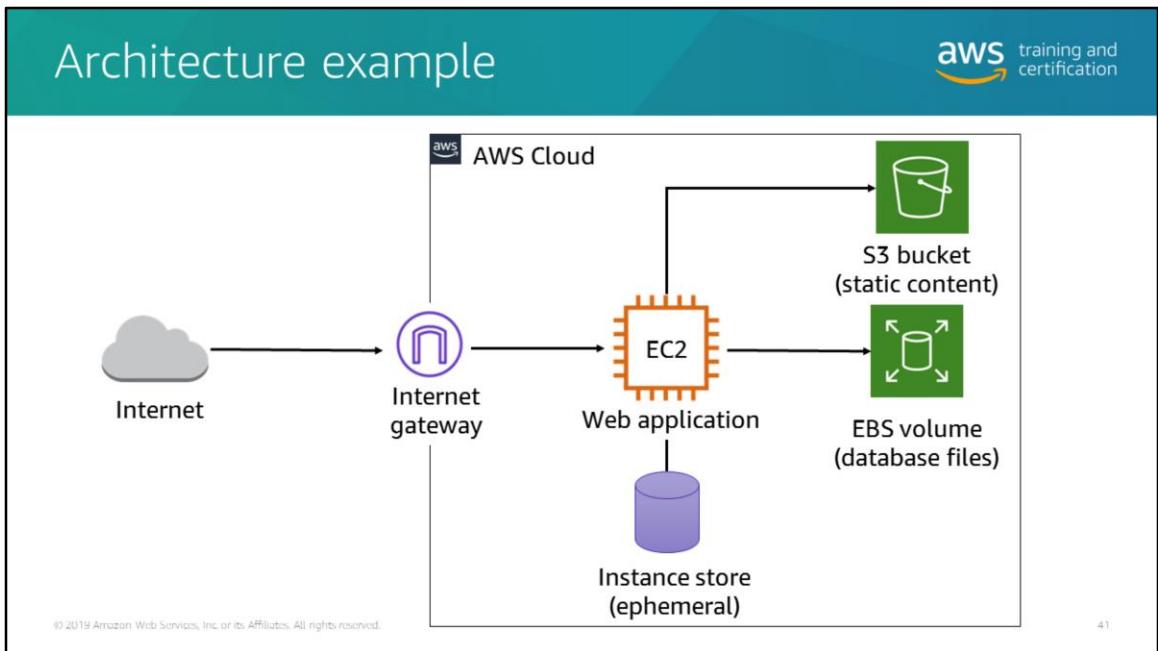
S3 One Zone-Infrequent Access (IA) stores data in a single AZ and costs 20% less than S3 Standard-IA. S3 One Zone-IA is ideal for customers who want a lower-cost option for infrequently accessed data but do not require the availability and resilience of S3 Standard or S3 Standard-IA.

S3 Glacier is suitable for archiving data where access is infrequent and a retrieval time of several hours is acceptable. Archived objects are not available for real-time access: they must be restored before they can be accessed. The Glacier storage class is very low-cost.

S3 Glacier Deep Archive is Amazon S3's lowest-cost storage class and supports long-term

retention and digital preservation for data that may be accessed once or twice in a year. It is designed for customers to retain data sets for 7-10 years or longer to meet regulatory compliance requirements.

- <http://docs.aws.amazon.com/AmazonS3/latest/dev/storage-class-intro.html>
- <https://aws.amazon.com/s3/storage-classes/>



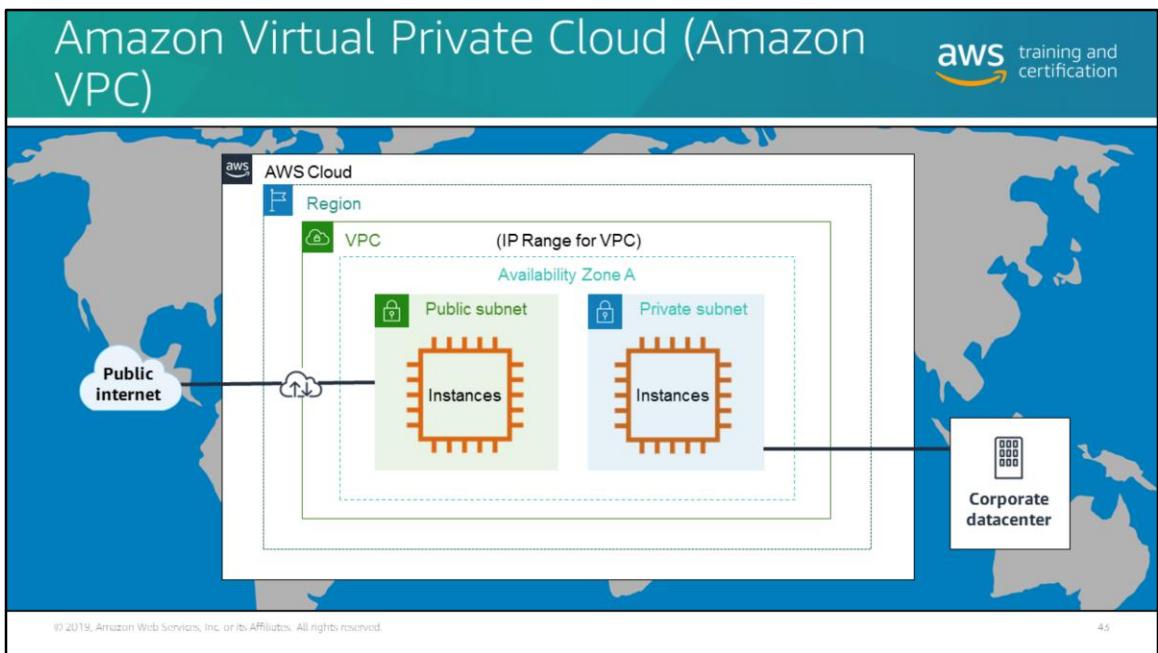
This is a simple architecture that highlights some of the storage options available in AWS. You might build a solution like this if you wished to migrate a small existing application to the cloud with minimal architectural changes. The EC2 instance is a web server and runs the business application and a database. The instance store is attached to the physical server hosting the EC2 instance and provides fast storage for the operating system and the application but should not be used for any persistent data. An S3 bucket holds static content for the website (such as HTML and images) that might be accessed by other instances or services. The EBS volume provides resilient storage for the database files.

There are many ways this architecture could be improved using other AWS services, which we will discuss later in this course.

Secure your data

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.





Virtual private clouds (VPCs) are isolated portions of the AWS Cloud in which customers deploy their AWS infrastructure—for example, they may deploy an Amazon EC2 instance or an Amazon Relational Database Service (Amazon RDS) instance.

VPCs are virtual networks. They support multiple subnets, routing, and fine-grained security mechanisms. You have complete control over your virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways. You can use both IPv4 and IPv6 in your VPC for secure and easy access to resources and applications.

In the simplified diagram above, after selecting a Region, you create a VPC and then specify the full IP address range for all resources that will be hosted within that VPC. The VPC can include resources in any or all Availability Zones within the Region. You can then create subnets within the network you specified for the VPC, choosing whether they'll allow connections to the public internet or remain private.

An Amazon VPC is a virtual network dedicated to your AWS account. It is logically isolated from other virtual networks in the AWS Cloud. You can launch your AWS

resources, such as Amazon EC2 instances, into your VPC. You can specify an IP address range for the VPC, add subnets, associate security groups, and configure route tables.

If you don't specify a subnet when you create a resource, it's launched into your default VPC. You can launch instances into your default VPC without needing to know anything about Amazon VPC. You can create your own VPC and configure it as you need. This is known as a *non-default VPC*. Subnets that you create in your non-default VPC and additional subnets that you create in your default VPC are called *non-default subnets*.

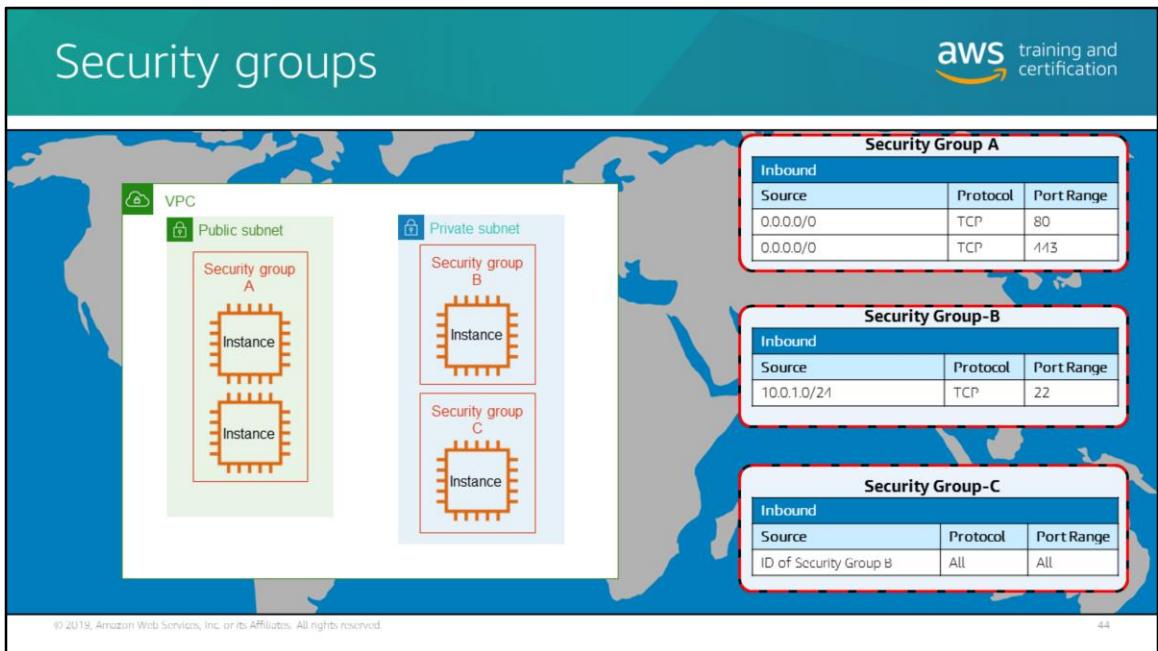
VPCs exist within a single AWS Region with up to five VPCs per Region.

Configurable features for a VPC include:

- Networks (for VPC and subnets).
- TCP/IP routing.
- Internet gateways.
- Security settings.

Security settings include:

- Network access control lists (network ACL) — Act as a firewall for associated subnets, controlling both inbound and outbound traffic at the subnet level.
- Security groups — Act as a firewall for associated Amazon EC2 instances, controlling both inbound and outbound traffic at the instance level (detailed in the next section).



A *security group* acts as a virtual firewall for your instance to control inbound and outbound traffic. Security groups act at the instance level, not the subnet level. Therefore, each resource in a subnet in your VPC could be assigned to a different set of security groups. If you don't specify a particular group at launch time, the instance is automatically assigned to the default security group for the VPC.

For each security group, you allow *rules* that control the inbound traffic to instances, and a separate set of rules that control the outbound traffic. This section describes the basic things you need to know about security groups for your VPC and their rules.

For more information on Security Groups, see
https://docs.aws.amazon.com/vpc/latest/userguide/VPC_SecurityGroups.html

In this diagram, we are creating two security groups. Security Group A has been configured to allow inbound web traffic from any possible IP address.

The default setting is to allow inbound traffic from any other resource placed in the default security group, and allow outbound to all possible IPs.

Security group details

The slide features a world map in the background, showing continents in light blue and oceans in white. In the top right corner, there is the AWS training and certification logo.

- Only “allow” rules; no “deny” rules
- Default values:
 - No inbound traffic allowed
 - All outbound traffic allowed
- Stateful:
 - Allows responses from allowed inbound traffic

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

45

You can specify allow rules, but not deny rules.

You can specify separate rules for inbound and outbound traffic.

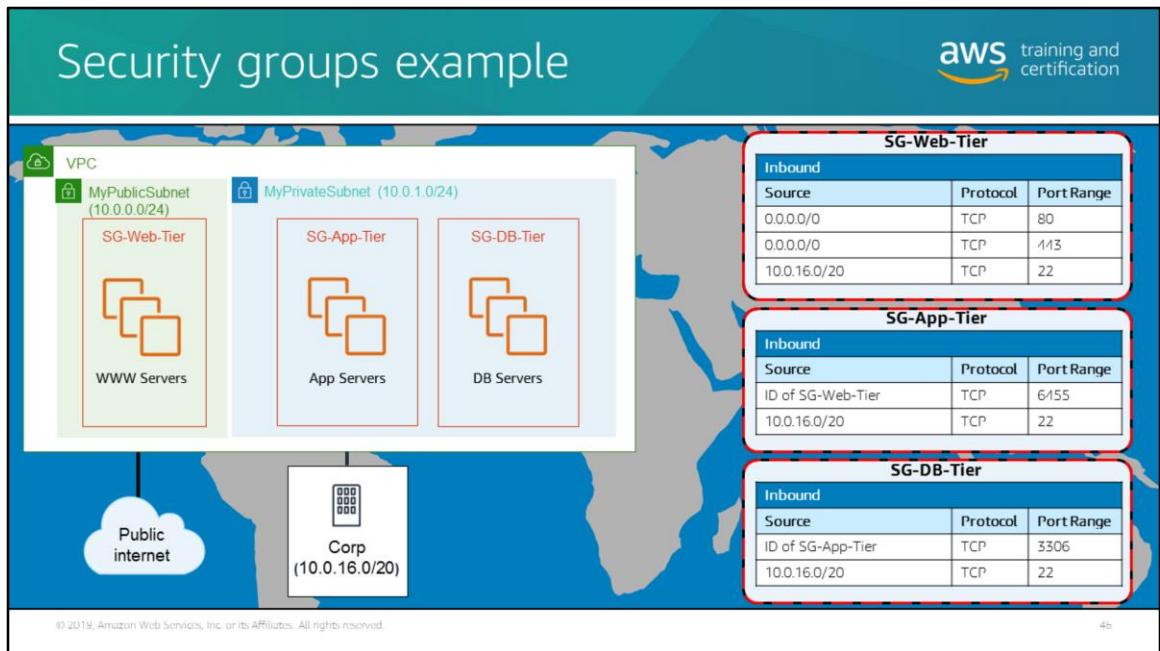
When you create a security group, it has no inbound rules. Therefore, no inbound traffic originating from another host to your instance is allowed until you add inbound rules to the security group.

By default, a security group includes an outbound rule that allows all outbound traffic. You can remove the rule and add outbound rules that allow specific outbound traffic only. If your security group has no outbound rules, no outbound traffic originating from your instance is allowed.

Security groups are stateful — if you send a request from your instance, the response traffic for that request is allowed to flow in, regardless of inbound security group rules. Responses to allowed inbound traffic are allowed to flow out, regardless of outbound rules.

For more information on limits, see

<https://docs.aws.amazon.com/vpc/latest/userguide/amazon-vpc-limits.html>



Here is an example of a classic AWS multi-tier security group example. In this architecture, notice that multiple different security group rules have been created to accommodate this multi-tiered web architecture.

If you start at the web tier, there is a rule set up to accept traffic from anywhere on the internet on port 80/443 by selecting the source of 0.0.0.0/0.

Next if you move to the app tier, there is a security group that only accepts traffic from the web tier, and similarly, the database tier can only accept traffic from the app tier.

Finally, there has also been a rule created to allow administration remotely from the corporate network over SSH port 22.

Knowledge check



Which of the following best describes the types of data for which Amazon S3 Glacier is best suited? (Choose two)

- A. Frequently erased within 30 days
- B. Is available after a three to five-hour restore period
- C. Is infrequently or rarely accessed
- D. Requires block storage

B and C.

Amazon S3 Glacier is optimized for long-term archival object storage and is not suited to data that needs immediate access or short lived data that is erased within 90 days.

Knowledge check



Which of the following best describes the types of data for which Amazon S3 Glacier is best suited? (Choose two)

- A. Frequently erased within 30 days
- B. Is available after a three to five-hour restore period
- C. Is infrequently or rarely accessed
- D. Requires block storage

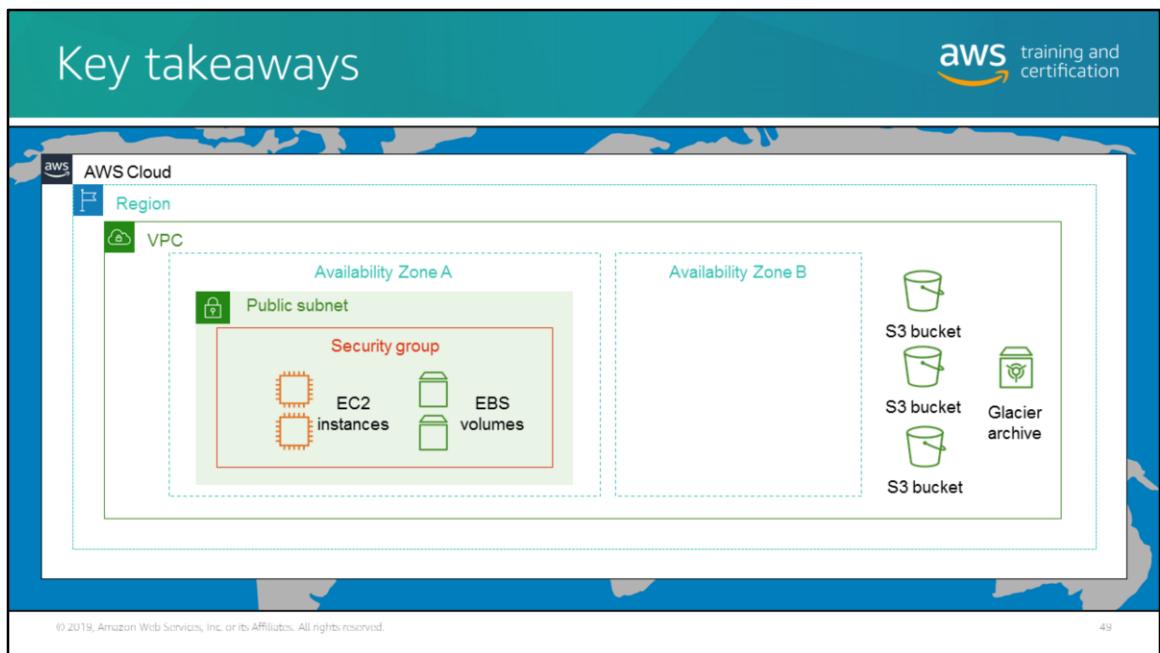
B and C are correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

48

B and C.

Amazon S3 Glacier is optimized for long-term archival object storage and is not suited to data that needs immediate access or short lived data that is erased within 90 days.



In this module, we introduced AWS products and services. We also discussed how these services can assist you with your AWS Cloud infrastructure and how you can use the tools AWS provides to build according to your business needs. Some of the services and details we referred to included Amazon EC2 , instance types, and Amazon EBS, which provides block storage for Amazon EC2 (providing your server functionality in the cloud). Finally, we talked about storing and securing your data with services such as Amazon S3 (general-purpose storage) and Amazon Glacier (archiving and long-term storage) and using networking components, such as VPCs and subnets. We also discussed securing your hosted instances with tools such as security groups.

Lab 1: Amazon Simple Storage Service



- Create a bucket in Amazon S3
- Add an object to your bucket
- Manage access permissions on an object
- Create a bucket policy
- Use bucket versioning
- Host a static website

Module 3: Building in the cloud



Module goals

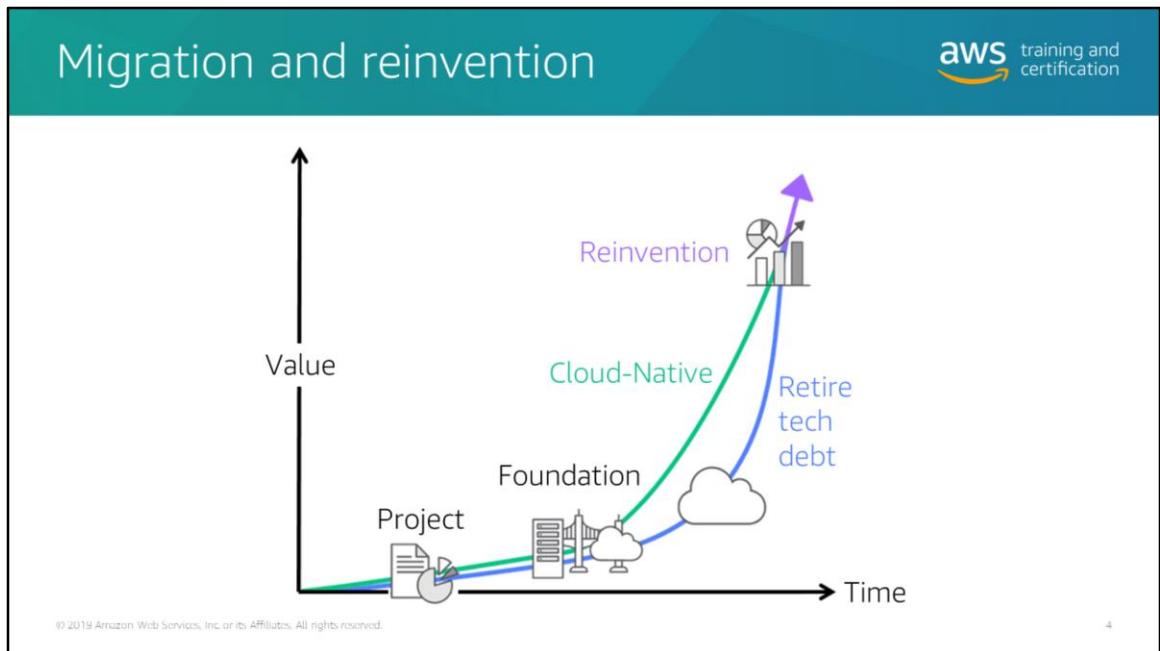


- Go beyond servers and storage
- Monitor AWS resources
- Manage demand efficiently
- Deploy database services
- Automate deployment
- Connect and share data
- Deliver content faster

Go beyond servers and storage

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.





So far you have learned about some of the services that make up the foundation of AWS—enough to get started building simple cloud applications. In this lesson, we will discuss additional services that will take you to the next level in your journey to designing cloud-native architectures. Cloud-native applications are designed to maximize the benefits the cloud offers, including agility, scalability, and cost-efficiency.

The path to cloud adoption is unique for every enterprise. Use the stages of adoption described here as a useful way to understand some of the steps involved. A customer's journey to the cloud typically involves these four phases:

- **Project**
In the project phase, you are running projects to get familiar with and experience the benefits from the cloud.
- **Foundation**
After experiencing the benefits of the cloud, you then build the foundation to scale your cloud adoption. Building the foundation includes creating a landing zone (a pre-configured, secure, multi-account AWS environment), Cloud Center of

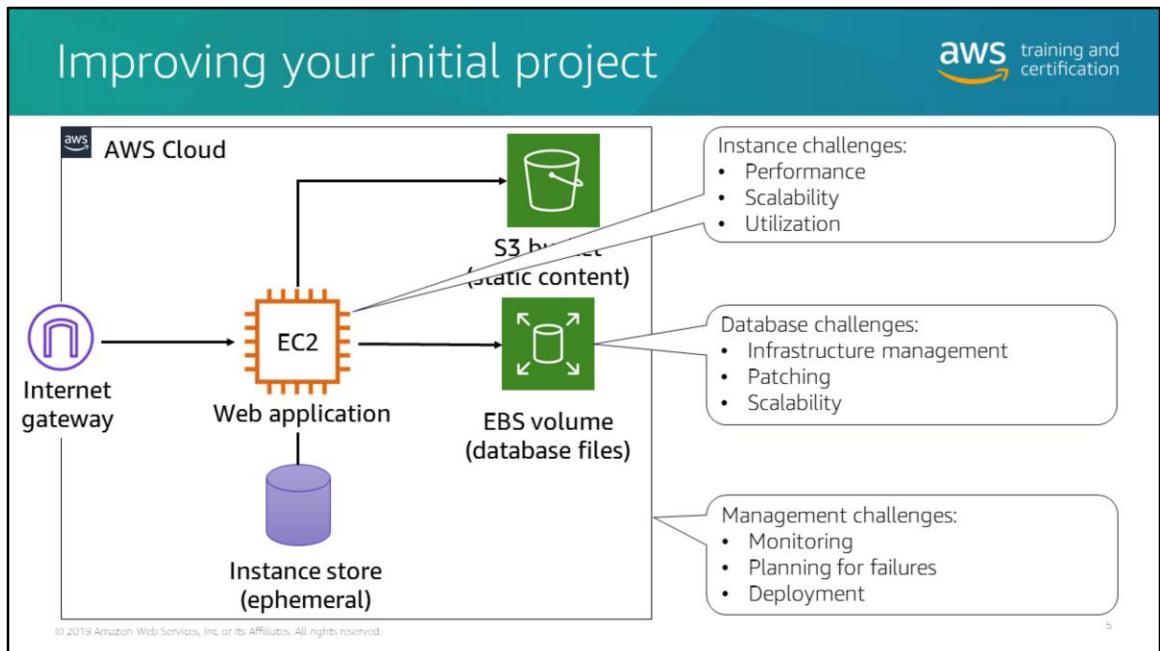
Excellence (CCoE), an operations model, in addition to facilitating security and compliance readiness.

- **Migration**

In this stage, you migrate existing applications, including mission-critical applications or entire data centers, to the cloud as you scale your adoption across a growing portion of your IT portfolio.

- **Reinvention**

Now that your operations are in the cloud, you can focus on reinvention. This is achieved by taking advantage of the flexibility and capabilities of AWS to transform your business by speeding time to market and increasing the attention on innovation.



Let's examine the sample architecture we looked at previously. It's a good architecture, but it has its limitations. Over time, you may find that your needs are outgrowing your initial project. AWS offers services that you can use to address some of these pain points and evolve your initial project into a foundation for further growth. Some of the specific challenges discussed are:

Limitations of a single EC2 instance:

- **Performance** – This instance runs three major applications: a web server, a production application, and a database. Distributing these roles across more compute resources improves efficiency and performance.
- **Scalability** – Can scale out to meet demand and scale in during less busy times improves utilization and efficiency.
- **Utilization** – The workload for this application varies over time, and spikes in traffic are difficult to manage. You want the ability to balance the load across your compute resources.

Limitations of a legacy database:

- **Infrastructure management** – You must spend time provisioning adequate compute and storage resources.

- **Patching** – You must maintain and back up the database application regularly.
- **Scalability** – It is challenging to scale the database to keep up with growing application demands.

Management challenges

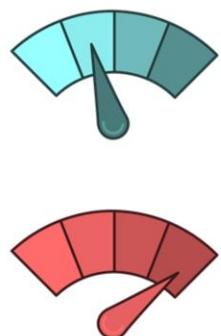
- Collecting information about the environment and monitoring activity
- Planning for failures and automating failover mechanisms where possible for continuity of operations
- As you grow your environment, deploying all the disparate services in one common and meaningful way

Monitor AWS resources

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



What is Amazon CloudWatch?



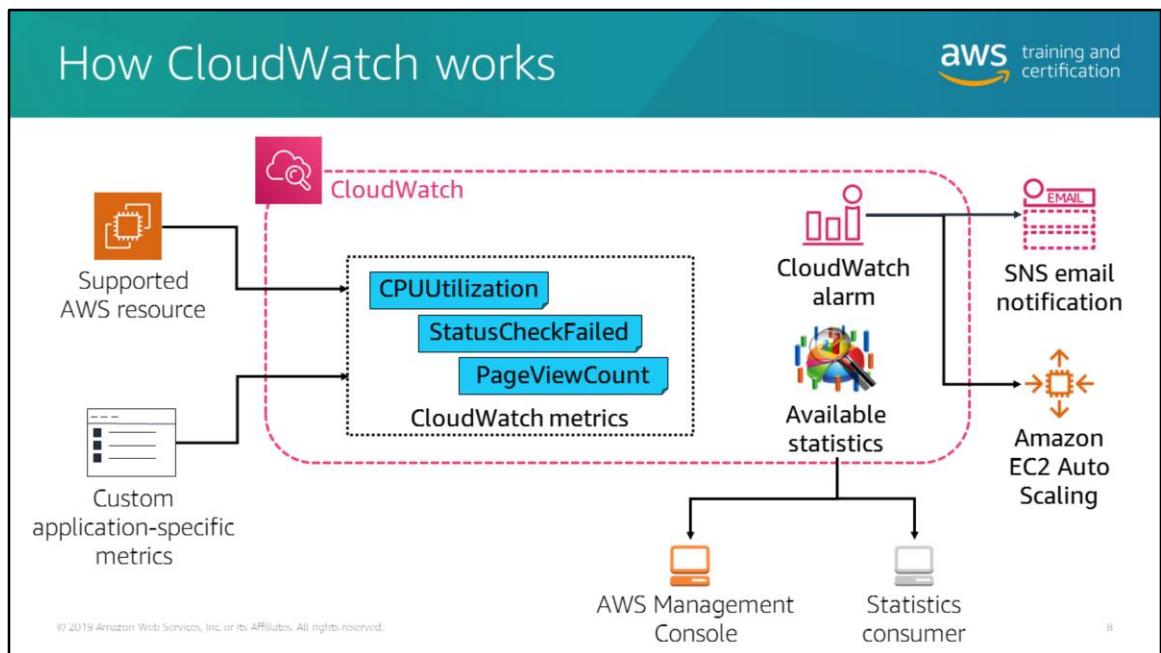
- Monitors:
 - AWS resources
 - Applications running on AWS
- Collects and tracks:
 - Standard metrics
 - Custom metrics
- Alarms:
 - Send notifications
 - Automatically make changes based on rules you define

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

/

Amazon CloudWatch is a monitoring and management service built for developers, system operators, site reliability engineers (SRE), and IT managers. CloudWatch monitors your AWS resources and the applications you run on AWS in real time. You can use CloudWatch to collect and track metrics, which are variables you can measure for your resources and applications.

CloudWatch alarms send notifications or automatically make changes to the resources you are monitoring based on rules that you define. For example, you can monitor the CPU usage and disk reads and writes of your Amazon EC2 instances and then use this data to determine whether you should launch additional instances to handle increased load. You can also use this data to stop under-used instances to save money. In addition to monitoring the AWS built-in metrics, you can monitor your own custom metrics. With CloudWatch, you gain system-wide visibility into resource utilization, application performance, and operational health. There is no upfront commitment or minimum fee; you simply pay for what you use. You are charged at the end of the month for what you use.



CloudWatch is a metrics repository. AWS services put metrics into the repository, and you retrieve statistics based on the metrics. Use the CloudWatch console to view the graphical representation of the statistics.

CloudWatch benefits



-  Access all your metrics from a single platform
-  Maintain visibility across your applications, infrastructure, and services
-  Reduce mean time to resolution (MTTR) and improve total cost of ownership (TCO)
-  Drive insights to optimize applications and operational resources
-  Pay as you go

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

9

Access all your metrics from a single platform

Modern applications are distributed (that is, they run on microservices architectures) and generate lots of data as metrics, logs, and more. You need a way to easily collect, access, and correlate these data points from individual sources in silos (server, network, database, etc.) to effectively monitor applications and infrastructure resources. CloudWatch enables you to collect metrics and logs from all your AWS resources, applications, and services that run on AWS and on-premises servers. Collecting metrics and logs helps you break down data silos so that you can easily gain system-wide visibility.

Maintain visibility across your applications, infrastructure, and services

Gaining visibility across your distributed stack means correlating and visualizing metrics and logs to quickly pinpoint and resolve issues. With CloudWatch, you can visualize key metrics like CPU utilization and memory. You can also correlate a log pattern to quickly get the context and go from diagnosing the problem to understanding the root cause.

Reduce mean time to resolution (MTTR) and improve total cost of ownership (TCO)

CloudWatch enables you to set high-resolution alarms and take automated actions. This means freeing up important resources to focus on adding business value. For example, you can get alerted on Amazon EC2 instances and set up EC2 Auto Scaling to add or remove instances. You can also execute automated responses to detect and shut down unused EC2 resources, reducing billing overages and improving resource optimization.

Drive insights to optimize applications and operational resources

To optimize performance and resource utilization, you need a unified operational view, real-time granular data, and historical reference. With CloudWatch, you get enhanced monitoring with 1-second granularity and up to 15 months of metrics storage and retention. You also have access to native CloudWatch features, such as Metric Math, to perform calculations on your metric data. For example, you can aggregate usage across an entire fleet of EC2 instances to derive operational and utilization insights.

Pay as you go

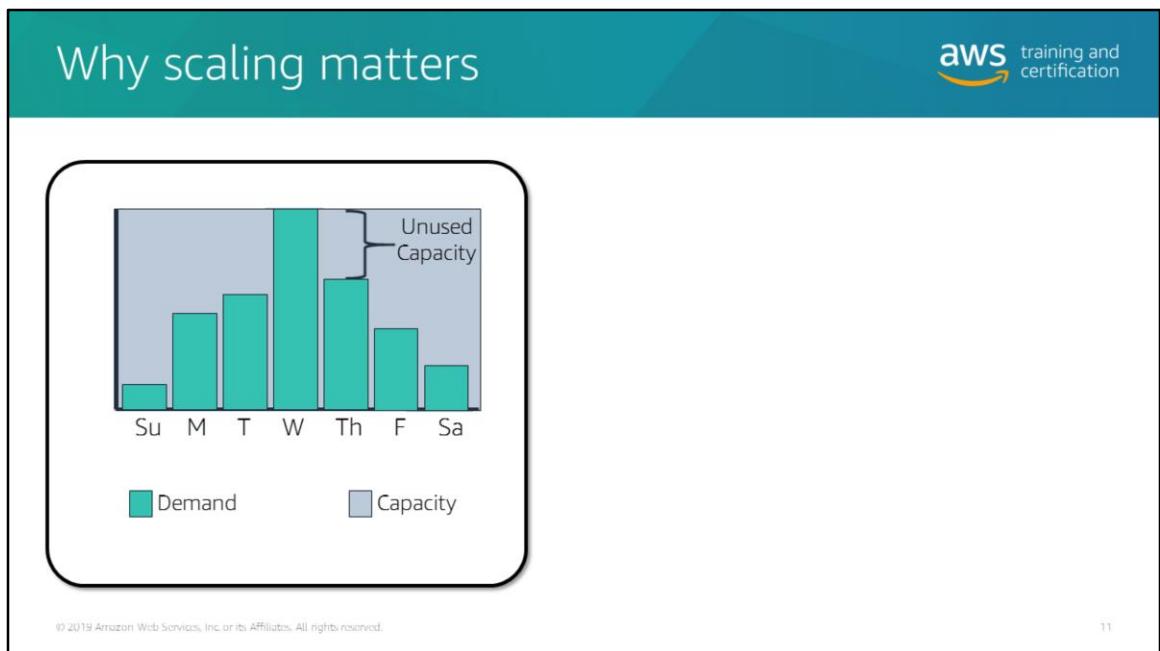
With CloudWatch, there is no upfront commitment or minimum fee; you pay for what you use. You are charged at the end of the month for your usage. For more information, see <https://aws.amazon.com/cloudwatch/pricing/>. You can estimate your monthly bill using the AWS Simple Monthly Calculator.

Manage demand efficiently

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



10

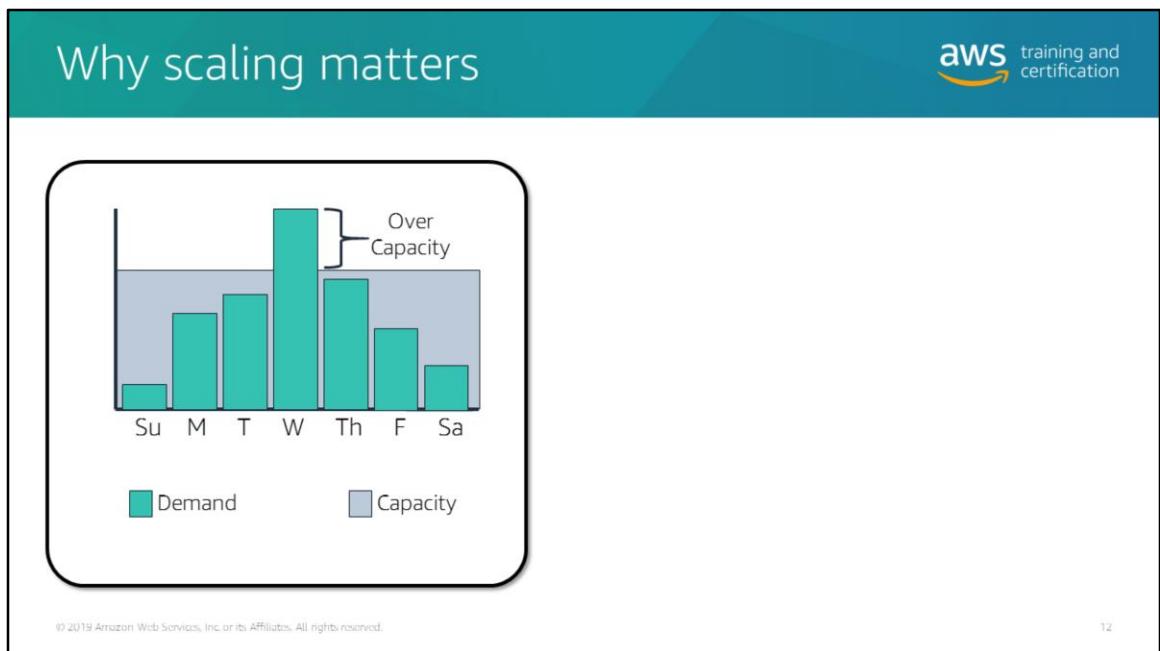


In a traditional data center environment, the scalability of your system is bound by your hardware. For example, this slide shows an example of a basic web application running on AWS. This application enables employees to search for conference rooms that they might want to use for meetings. During the beginning and end of the week, usage of this application is minimal. During the middle of the week, more employees are scheduling meetings, so the demand on the application increases significantly.

In a data center, anticipating periods of heavy use requires spinning up enough physical servers to handle the anticipated load. But what happens to those servers the rest of the time? They sit idle in the data center.

In the cloud, because computing power is a programmatic resource, we can take a more flexible approach to the issue of scaling. By adding EC2 Auto Scaling to this application, you can add new instances to the application only when necessary and terminate them when they're no longer needed. Because EC2 Auto Scaling uses EC2 instances, you pay for only the instances you use, when you use them. You now have a cost-effective architecture that provides the best customer experience while reducing expenses.

What's required to implement such a system? Let's explore how several AWS services can be used together to create a scalable, on-demand architecture.



12

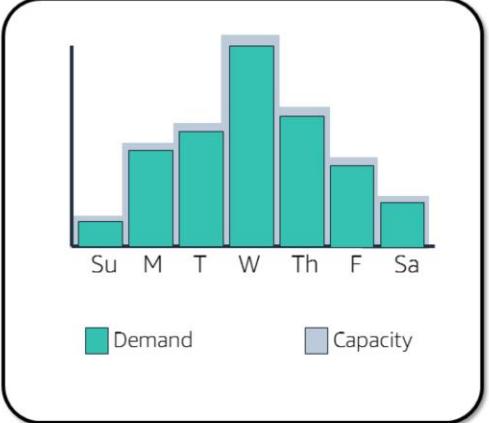
In a traditional data center environment, the scalability of your system is bound by your hardware. For example, this slide shows an example of a basic web application running on AWS. This application enables employees to search for conference rooms that they might want to use for meetings. During the beginning and end of the week, usage of this application is minimal. During the middle of the week, more employees are scheduling meetings, so the demand on the application increases significantly.

In a data center, anticipating periods of heavy use requires spinning up enough physical servers to handle the anticipated load. But what happens to those servers the rest of the time? They sit idle in the data center.

In the cloud, because computing power is a programmatic resource, we can take a more flexible approach to the issue of scaling. By adding EC2 Auto Scaling to this application, you can add new instances to the application only when necessary and terminate them when they're no longer needed. Because EC2 Auto Scaling uses EC2 instances, you pay for only the instances you use, when you use them. You now have a cost-effective architecture that provides the best customer experience while reducing expenses.

What's required to implement such a system? Let's explore how several AWS services can be used together to create a scalable, on-demand architecture.

Why scaling matters



The chart displays weekly usage patterns for a web application. The x-axis lists the days of the week: Su, M, T, W, Th, F, Sa. The left y-axis represents 'Demand' (teal bars), and the right y-axis represents 'Capacity' (light blue bars). Capacity is consistently higher than demand, peaking on Wednesday.

Day	Demand	Capacity
Su	Low	Medium
M	Medium	Medium-High
T	Medium-High	High
W	High	Very High
Th	Medium-High	High
F	Medium	Medium-Low
Sa	Low	Medium-Low

Amazon EC2 Auto Scaling adjusts capacity as needed

- Scale out for spikes
- Scale in during off-peak
- Replace unhealthy instances
- Pay only for what you use

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

In a traditional data center environment, the scalability of your system is bound by your hardware. For example, this slide shows an example of a basic web application running on AWS. This application enables employees to search for conference rooms that they might want to use for meetings. During the beginning and end of the week, usage of this application is minimal. During the middle of the week, more employees are scheduling meetings, so the demand on the application increases significantly.

In a data center, anticipating periods of heavy use requires spinning up enough physical servers to handle the anticipated load. But what happens to those servers the rest of the time? They sit idle in the data center.

In the cloud, because computing power is a programmatic resource, we can take a more flexible approach to the issue of scaling. By adding EC2 Auto Scaling to this application, you can add new instances to the application only when necessary and terminate them when they're no longer needed. Because EC2 Auto Scaling uses EC2 instances, you pay for only the instances you use, when you use them. You now have a cost-effective architecture that provides the best customer experience while reducing expenses.

What's required to implement such a system? Let's explore how several AWS services can be used together to create a scalable, on-demand architecture.

Dynamic scaling with Amazon EC2 Auto Scaling



Follow the demand curve for your applications

- Select a load metric for your application
- Set as conditional and/or scheduled
- Use with CloudWatch, optionally

Max	10
Min	2
Desired	6



Average Demand

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

14

Amazon EC2 Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost. The dynamic scaling capabilities of Amazon EC2 Auto Scaling automatically increase or decrease capacity based on load or other metrics. For example, if your CPU spikes above 80% (and you have an alarm set up) Amazon EC2 Auto Scaling can add a new instance dynamically. You can also set a condition to remove instances in the same increments when CPU utilization is low. If you have predictable load changes, you can set a schedule through Amazon EC2 Auto Scaling to plan your scaling activities.

Dynamic scaling with Amazon EC2 Auto Scaling



Follow the demand curve for your applications

- Select a load metric for your application
- Set as conditional and/or scheduled
- Use with CloudWatch, optionally

Max	10
Min	2
Desired	10



High Demand

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

15

Amazon EC2 Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost. The dynamic scaling capabilities of Amazon EC2 Auto Scaling automatically increase or decrease capacity based on load or other metrics. For example, if your CPU spikes above 80% (and you have an alarm set up) Amazon EC2 Auto Scaling can add a new instance dynamically. You can also set a condition to remove instances in the same increments when CPU utilization is low. If you have predictable load changes, you can set a schedule through Amazon EC2 Auto Scaling to plan your scaling activities.

Dynamic scaling with Amazon EC2 Auto Scaling



Follow the demand curve for your applications

- Select a load metric for your application
- Set as conditional and/or scheduled
- Use with CloudWatch, optionally

Max	10
Min	2
Desired	2



Low Demand

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

1b

Amazon EC2 Auto Scaling monitors your applications and automatically adjusts capacity to maintain steady, predictable performance at the lowest possible cost. The dynamic scaling capabilities of Amazon EC2 Auto Scaling automatically increase or decrease capacity based on load or other metrics. For example, if your CPU spikes above 80% (and you have an alarm set up) Amazon EC2 Auto Scaling can add a new instance dynamically. You can also set a condition to remove instances in the same increments when CPU utilization is low. If you have predictable load changes, you can set a schedule through Amazon EC2 Auto Scaling to plan your scaling activities.

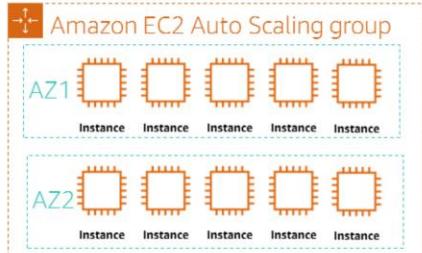
Fleet management with Amazon EC2 Auto Scaling



Replace impaired Amazon EC2 instances without intervention

- Monitor the health of running instances
- Replace impaired instances automatically
- Balance capacity across Availability Zones

Max	10
Min	2
Desired	10



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

1/

If your application runs on Amazon EC2 instances, then you have is called a *fleet*. Fleet management refers to the functionality that automatically replaces unhealthy instances and maintains your fleet at the desired capacity. Amazon EC2 Auto Scaling fleet management ensures that your application is able to receive traffic and that the instances themselves are working properly. When Amazon EC2 Auto Scaling detects a failed health check, it can replace the instance automatically.

In this example, the resources being scaled are Amazon EC2 instances. AWS Application Auto Scaling, a separate service is also available. AWS Application Auto Scaling provides application scaling for multiple resources across multiple services, such as Spot Fleets, Amazon ECS tasks, Amazon DynamoDB tables, and indexes, and Amazon Aurora Replicas. The Application Auto Scaling interface provides recommendations that enable you to optimize performance, costs, or balance between them. With AWS Application Auto Scaling, your applications have the right resources at the right time.

Both EC2 Auto Scaling and Application Auto Scaling are available at no additional charge. You pay only for the AWS resources needed to run your applications and any CloudWatch monitoring fees.

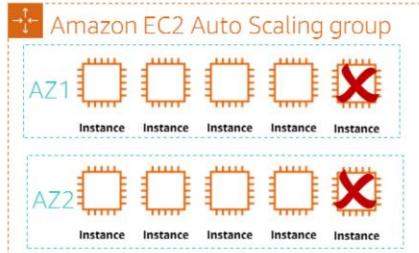
Fleet management with Amazon EC2 Auto Scaling



Replace impaired Amazon EC2 instances without intervention

- Monitor the health of running instances
- Replace impaired instances automatically
- Balance capacity across Availability Zones

Max	10
Min	2
Desired	10



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

18

If your application runs on Amazon EC2 instances, then you have is called a *fleet*. Fleet management refers to the functionality that automatically replaces unhealthy instances and maintains your fleet at the desired capacity. Amazon EC2 Auto Scaling fleet management ensures that your application is able to receive traffic and that the instances themselves are working properly. When Amazon EC2 Auto Scaling detects a failed health check, it can replace the instance automatically.

In this example, the resources being scaled are Amazon EC2 instances. AWS Application Auto Scaling, a separate service is also available. AWS Application Auto Scaling provides application scaling for multiple resources across multiple services, such as Spot Fleets, Amazon ECS tasks, Amazon DynamoDB tables, and indexes, and Amazon Aurora Replicas. The Application Auto Scaling interface provides recommendations that enable you to optimize performance, costs, or balance between them. With AWS Application Auto Scaling, your applications have the right resources at the right time.

Both EC2 Auto Scaling and Application Auto Scaling are available at no additional charge. You pay only for the AWS resources needed to run your applications and any CloudWatch monitoring fees.

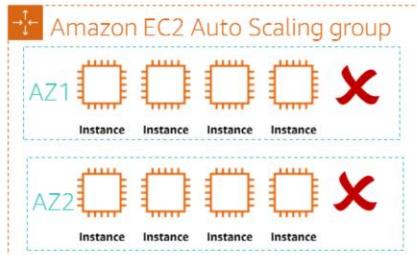
Fleet management with Amazon EC2 Auto Scaling



Replace impaired Amazon EC2 instances without intervention

- Monitor the health of running instances
- Replace impaired instances automatically
- Balance capacity across Availability Zones

Max	10
Min	2
Desired	10



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

19

If your application runs on Amazon EC2 instances, then you have is called a *fleet*. Fleet management refers to the functionality that automatically replaces unhealthy instances and maintains your fleet at the desired capacity. Amazon EC2 Auto Scaling fleet management ensures that your application is able to receive traffic and that the instances themselves are working properly. When Amazon EC2 Auto Scaling detects a failed health check, it can replace the instance automatically.

In this example, the resources being scaled are Amazon EC2 instances. AWS Application Auto Scaling, a separate service is also available. AWS Application Auto Scaling provides application scaling for multiple resources across multiple services, such as Spot Fleets, Amazon ECS tasks, Amazon DynamoDB tables, and indexes, and Amazon Aurora Replicas. The Application Auto Scaling interface provides recommendations that enable you to optimize performance, costs, or balance between them. With AWS Application Auto Scaling, your applications have the right resources at the right time.

Both EC2 Auto Scaling and Application Auto Scaling are available at no additional charge. You pay only for the AWS resources needed to run your applications and any CloudWatch monitoring fees.

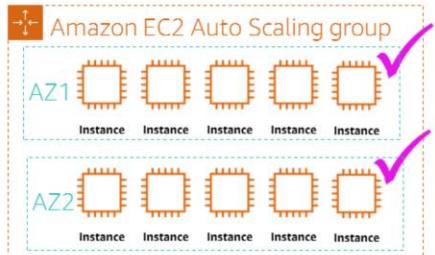
Fleet management with Amazon EC2 Auto Scaling



Replace impaired Amazon EC2 instances without intervention

- Monitor the health of running instances
- Replace impaired instances automatically
- Balance capacity across Availability Zones

Max	10
Min	2
Desired	10



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

20

If your application runs on Amazon EC2 instances, then you have is called a *fleet*. Fleet management refers to the functionality that automatically replaces unhealthy instances and maintains your fleet at the desired capacity. Amazon EC2 Auto Scaling fleet management ensures that your application is able to receive traffic and that the instances themselves are working properly. When Amazon EC2 Auto Scaling detects a failed health check, it can replace the instance automatically.

In this example, the resources being scaled are Amazon EC2 instances. AWS Application Auto Scaling, a separate service is also available. AWS Application Auto Scaling provides application scaling for multiple resources across multiple services, such as Spot Fleets, Amazon ECS tasks, Amazon DynamoDB tables, and indexes, and Amazon Aurora Replicas. The Application Auto Scaling interface provides recommendations that enable you to optimize performance, costs, or balance between them. With AWS Application Auto Scaling, your applications have the right resources at the right time.

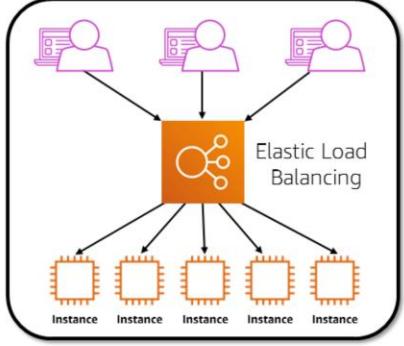
Both EC2 Auto Scaling and Application Auto Scaling are available at no additional charge. You pay only for the AWS resources needed to run your applications and any CloudWatch monitoring fees.

Elastic Load Balancing

aws training and certification

Automatically distribute traffic across multiple targets

-  High availability
-  Health checks
-  SSL/TLS termination
-  Operational monitoring



The diagram illustrates the function of an Elastic Load Balancer. At the top, three user icons (laptop, smartphone, tablet) have arrows pointing down to a central orange square labeled "Elastic Load Balancing". From the bottom of this central square, five arrows point down to five smaller orange squares at the bottom, each labeled "Instance".

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

21

High availability

Elastic Load Balancing automatically distributes traffic across multiple targets—Amazon EC2 instances, containers, and IP addresses—in a single Availability Zone or multiple Availability Zones.

Health checks

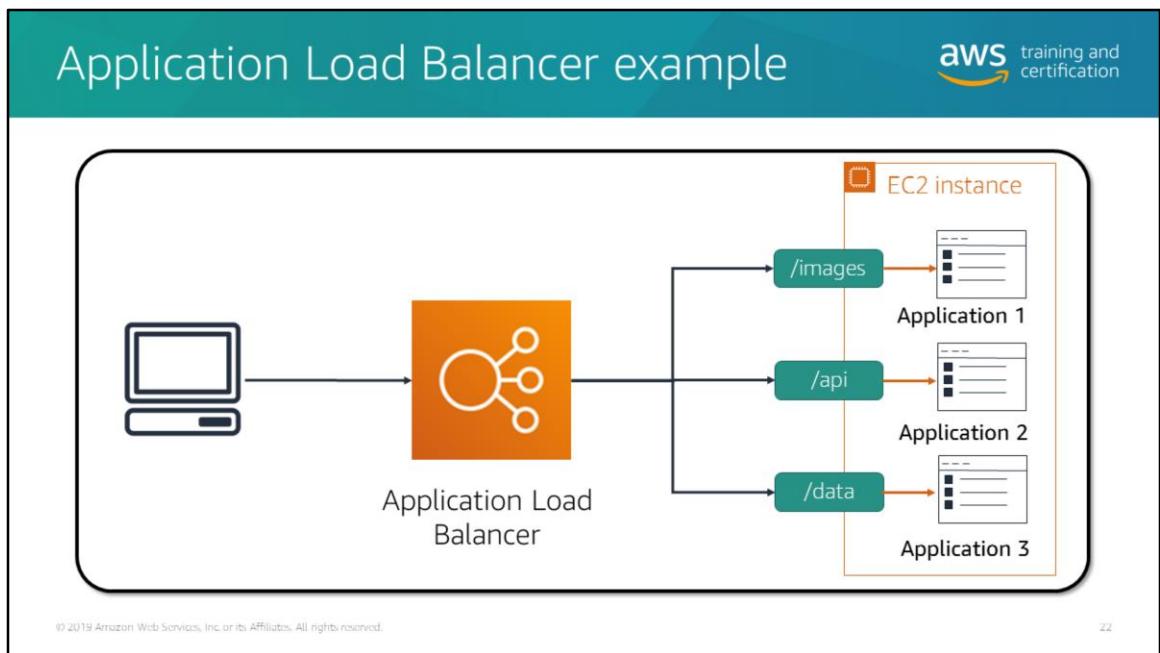
Elastic Load Balancing can detect unhealthy targets, stop sending traffic to them and then spread the load across the remaining healthy targets.

TLS termination

Elastic Load Balancing provides integrated certificate management and SSL decryption, allowing you the flexibility to centrally manage the SSL settings of the load balancer and offload CPU-intensive work from your application.

Operational monitoring

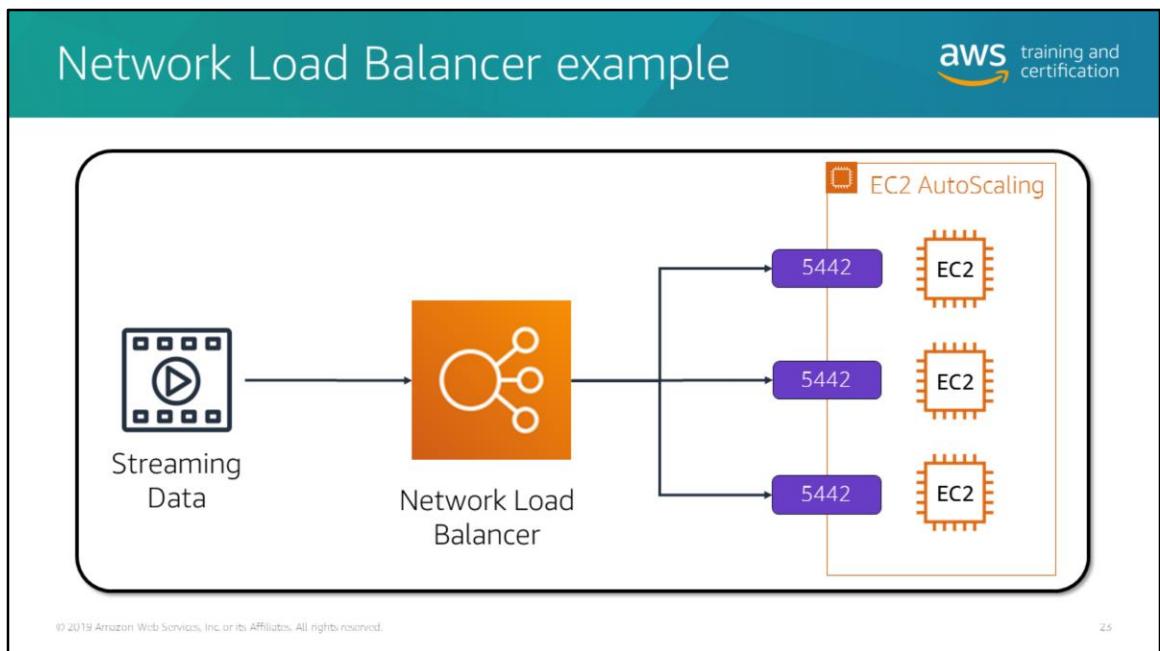
Elastic Load Balancing integrates with Amazon CloudWatch metrics and request tracing to monitor performance of your applications in real time.



The Application Load Balancer is ideal for advanced load balancing of HTTP and HTTPS traffic. The Application Load Balancer provides advanced request routing that supports modern application architectures, including microservices and container-based applications. An Application Load Balancer functions at the application layer, the seventh layer of the Open Systems Interconnection (OSI) model.

Application Load Balancers support a pair of open standard protocols (WebSocket and HTTP/2) and also provide additional visibility into the health of the target instances and containers. Websites and mobile apps, running in containers or on EC2 instances, will benefit from the use of Application Load Balancers.

You can use Application Load Balancers in many scenarios. One is the ability to use containers to host your microservices and route to those applications from a single load balancer. The Application Load Balancer enables you to route different requests to the same instance but differ the path based on the port. If you have different containers listening on various ports, you can set up routing rules to distribute traffic to only the desired backend application.



The Network Load Balancer automatically routes incoming web traffic across such a dynamically changing number of instances. Your load balancer acts as a single point of contact for all incoming traffic to the instances in your Auto Scaling group. You can automatically increase the size of your Auto Scaling group when demand goes up and decrease it when demand goes down. As the Auto Scaling group adds and removes Amazon EC2 instances, the Network Load Balancer makes sure that the traffic for your application is distributed across all of your instances.

Network Load Balancer is designed to handle tens of millions of requests per second while maintaining high throughput at ultra-low latency, with no effort on your part. Network Load Balancers operate at the connection level (Layer 4), routing connections to targets—Amazon EC2 instances, containers, and IP addresses—based on IP protocol data. The Network Load Balancer is API-compatible with the Application Load Balancer, including full programmatic control of Target Groups and Targets. The Network Load Balancer is ideal for load balancing of TCP traffic.

The Network Load Balancer is optimized to handle sudden and volatile traffic patterns while using a single static IP address per Availability Zone.

Knowledge check



You have an application composed of individual services. You need to route a request to a service based on the content of the request. What type of load balancer should you use?

- A. Auto Scaling Load Balancer
- B. Network Load Balancer
- C. Application Load Balancer
- D. Any type of load balancer

Knowledge check



You have an application composed of individual services. You need to route a request to a service based on the content of the request. What type of load balancer should you use?

- A. Auto Scaling Load Balancer
- B. Network Load Balancer
- C. Application Load Balancer
- D. Any type of load balancer

Knowledge check



You have an application composed of individual services. You need to route a request to a service based on the content of the request. What type of load balancer should you use?

- A. Auto Scaling Load Balancer
- B. Network Load Balancer
- C. Application Load Balancer
- D. Any type of load balancer

Knowledge check



You have an application composed of individual services. You need to route a request to a service based on the content of the request. What type of load balancer should you use?

- A. Auto Scaling Load Balancer
- B. Network Load Balancer
- C. Application Load Balancer
- D. Any type of load balancer

Knowledge check



You have an application composed of individual services. You need to route a request to a service based on the content of the request. What type of load balancer should you use?

- A. Auto Scaling Load Balancer
- B. Network Load Balancer
- C. Application Load Balancer
- D. Any type of load balancer

C is correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

28

Deploy database services

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



29

DIY vs. AWS database services



Databases on Amazon EC2

- Operating system access
- Need features of specific application



AWS Database Services

- Easy to set up, manage, maintain
- Push-button high availability
- Focus on performance
- Managed infrastructure

Amazon database services are easier to set up, manage, and maintain than running database software on Amazon EC2. They let you focus on tasks other than the day-to-day administration of the database platform. Alternatively, running your own database software will give you more control, flexibility, and choice. Depending on your application and your requirements, you might prefer one over the other.

Here are some of the advantages of using AWS database services (this list primarily refers to Amazon RDS, though some features are not available on all services):

- Focus on your business and applications, and have AWS take care of the management tasks, such as provisioning the database, performing backup and recovery tasks, and managing security patches, storage, and minor version updates.
- You need a highly available database solution and want to take advantage of the push-button, synchronous Multi-AZ deployment without having to manually set up and maintain a standby database.
- You would like to have synchronous replication to a standby instance for high availability for Oracle Database Standard Edition One (SE1) or Oracle Database Standard Edition Two (SE2).
- You don't want to manage backups and, most important, point-in-time recoveries

of your database.

- You would rather focus on high-level tasks, such as performance tuning and schema optimization, than on the daily administration of the database.
- You want to scale the instance type up or down based on your workload patterns without being concerned about licensing and the complexity involved.

Here are some of the advantages of hosting database software on EC2 instances:

- You need full control over the database, including SYS/SYSTEM user access, or you need access at the operating system level.
- You need to use commercial software features or options that are not currently supported by AWS. See the documentation for currently supported options.

What is Amazon Relational Database Service?



A database service that makes it easy to set up, operate, and scale a relational database in the cloud

Amazon RDS Engines

Amazon
Aurora

 PostgreSQL

 MariaDB

 ORACLE

 MySQL

 Microsoft SQL Server

- Easily scalable
- Automatic software patching
- Automated backups
- Database snapshots
- Multi-AZ deployments
- Automatic host replacement
- Encryption at rest and in transit

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

51

Amazon Relational Database Service (Amazon RDS) provides cost-efficient and resizable capacity while automating time-consuming administration tasks such as hardware provisioning, database setup, patching, and backups. It frees you to focus on your applications so that you can give them the fast performance, high availability, security, and compatibility they need.

Amazon RDS is available on several database instance types—optimized for memory, performance, or I/O—and provides you with six familiar database engines to choose from, including Amazon Aurora, PostgreSQL, MySQL, MariaDB, Oracle, and Microsoft SQL Server. You can use the AWS Database Migration Service to easily migrate or replicate your existing databases to Amazon RDS.

For detailed information about the benefits of Amazon RDS, see the product page online at <https://aws.amazon.com/rds/features/>

What is Amazon Aurora?



- Enterprise-class relational database
- MySQL- or PostgreSQL-compatible
- Up to 5X faster than standard MySQL databases
- Up to 3X faster than standard PostgreSQL databases
- Continuous backup to Amazon S3
- Up to 15 low-latency read replicas

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

52

Amazon Aurora is a great option for any enterprise application that can use a relational database. Compared to commercial databases, Amazon Aurora helps reduce your database costs by 90 percent or more while improving reliability and availability of the database. Amazon Aurora, as a fully managed service, helps you save time by automating time-consuming tasks such as provisioning, patching, backup, recovery, failure detection, and repair.

The service is a MySQL and PostgreSQL-compatible relational database built for the cloud, which combines the performance and availability of traditional enterprise databases with the simplicity and cost-effectiveness of open-source databases.

It's up to five times faster than standard MySQL databases and three times faster than standard PostgreSQL databases. It provides the security, availability, and reliability of commercial databases at 1/10 the cost. Amazon RDS fully manages Amazon Aurora by automating time-consuming administration tasks such as hardware provisioning, database setup, patching, and backups.

Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that automatically scales up to 64 TB per database instance. It delivers high performance and availability with up to 15 low-latency read replicas, point-in-time recovery, nearly continuous backup to Amazon S3, and replication across three Availability Zones.

To create your first Aurora database instance and start migrating your MySQL and PostgreSQL databases, use the Amazon RDS console.

Relational vs key-value databases																
	Relational (SQL)			Key-value (NoSQL)												
Data storage	Rows and columns			Key-value, document, graph												
Schemas	Fixed			Dynamic												
Querying	Using SQL			Focused on collection of documents												
Scalability	Vertical			Horizontal												
Example	<table border="1"> <thead> <tr> <th>ISBN</th> <th>Title</th> <th>Author</th> <th>Format</th> </tr> </thead> <tbody> <tr> <td>3111111223439</td> <td>Withering Depths</td> <td>Tark, Frank</td> <td>Paperback</td> </tr> <tr> <td>3122222223439</td> <td>Wily Willy</td> <td>Felton, Maria</td> <td>eBook</td> </tr> </tbody> </table>	ISBN	Title	Author	Format	3111111223439	Withering Depths	Tark, Frank	Paperback	3122222223439	Wily Willy	Felton, Maria	eBook	<pre>{ ISBN: 3111111223439, Title: "Withering Depths", Author: "Tark, Frank", Format: "Paperback" }</pre>		
ISBN	Title	Author	Format													
3111111223439	Withering Depths	Tark, Frank	Paperback													
3122222223439	Wily Willy	Felton, Maria	eBook													

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5.5

A SQL database stores data in rows and columns. Rows contain all the information about one entry, and columns are the attributes that separate the data points. A SQL database schema is fixed: columns must be locked before data entry. You can amend schemas if the database is altered entirely and taken offline. Data in SQL databases is queried using structure query language (SQL), which can allow for complex queries. SQL databases scale vertically by increasing hardware power. Relational databases are commonly used for traditional applications, ERP, CRM, and ecommerce.

NoSQL databases store data using one of many storage models, including key-value pairs, documents, and graphs. NoSQL schemas are dynamic, and information can be added rapidly. Each *row* doesn't have to contain data for each *column*. Data in NoSQL databases is queried by focusing on collections of documents. NoSQL databases scale horizontally by increasing servers. Key-value databases are commonly used for internet-scale applications, real-time bidding, shopping carts, and customer preferences.

What is Amazon DynamoDB?



Fast and flexible NoSQL database service for any scale



- Fully managed
- Low-latency queries
- Fine-grained access control
- Regional and global options

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

54

Amazon DynamoDB is a fast and flexible nonrelational database service for all applications that need consistent, single-digit millisecond latency at any scale. It is a fully managed cloud database and supports both document and key-value store models.

Fully managed

DynamoDB is a fully managed, nonrelational database service—you simply create a database table, set your target utilization for automatic scaling, and let the service handle the rest. You no longer need to worry about database management tasks, such as hardware or software provisioning, setup and configuration, software patching, operating a distributed database cluster, or partitioning data over multiple instances, as you scale. DynamoDB also provides point-in-time recovery, backup, and restore for all your tables, helping you meet your corporate and regulatory archival requirements.

Low-latency queries

Average service-side latencies are typically single-digit milliseconds. As your data volumes grow and application performance demands increase, DynamoDB uses automatic partitioning and SSD technologies to meet your throughput requirements.

and deliver low latencies at any scale.

Fine-grained access control

DynamoDB integrates with AWS Identity and Access Management (IAM) for fine-grained access control of users in your organization. You can assign unique security credentials to each user and control each user's access to services and resources.

Flexibility

DynamoDB supports storing, querying, and updating documents. By using the AWS SDK, you can write applications that store JSON documents directly into Amazon DynamoDB tables. This capability reduces the amount of new code to be written to insert, update, and retrieve JSON documents, and to perform powerful database operations, such as nested JSON queries, by using only a few lines of code.

Amazon DynamoDB use cases



- Serverless web applications
- Microservices data store
- Mobile backends
- Ad tech
- Gaming
- Internet of Things (IoT)

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

55

Serverless web applications

Build powerful web applications that automatically scale up and down. You don't need to maintain servers, and your applications have automated high availability.

Microservices data store

Build flexible and reusable microservices by using DynamoDB as a serverless data store for consistent and fast performance.

Mobile backends

Build personalized mobile apps with smooth experiences for your users. DynamoDB takes care of operational tasks so that you can focus on your applications.

Adtech

Create real-time bidding platforms and recommendation engines with the scalability, throughput, and availability of DynamoDB.

Gaming

Create responsive games for mobile, console, and desktop with DynamoDB. Store and query game data such as player state, high scores, or world dynamic content.

Internet of Things (IoT)

Analyze your devices by connecting your high-velocity, high-volume IoT data in DynamoDB to Amazon Redshift and Amazon QuickSight.

Other purpose-built database services



Amazon Redshift
Fast, scalable data warehouse



Amazon DocumentDB
MongoDB-compatible database



Amazon Neptune
Graph database

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5b

In addition to relational databases for transactional applications and key-value databases for internet-scale applications, the AWS fully managed database services include the following components:

- A Fast, scalable data warehouse
- A data warehouse for analytics
- A graph database for building applications with highly connected data

Amazon Redshift is a fast, scalable data warehouse that makes it simple and cost-effective to analyze all your data across your data warehouse and data lake. Amazon Redshift delivers 10 times faster performance than other data warehouses by using machine learning, massively parallel query execution, and columnar storage on high-performance disks. You can set up and deploy a new data warehouse in minutes. You can also run queries across petabytes of data in your Amazon Redshift data warehouse and exabytes of data in your data lake built on Amazon S3.

Amazon Redshift is an online analytical processing (OLAP) system as opposed to Amazon RDS databases, which are online transaction processing (OLTP). OLTP databases usually process a large number of small transactions and are often used to provide source data to data warehouses. OLAP systems usually process a small

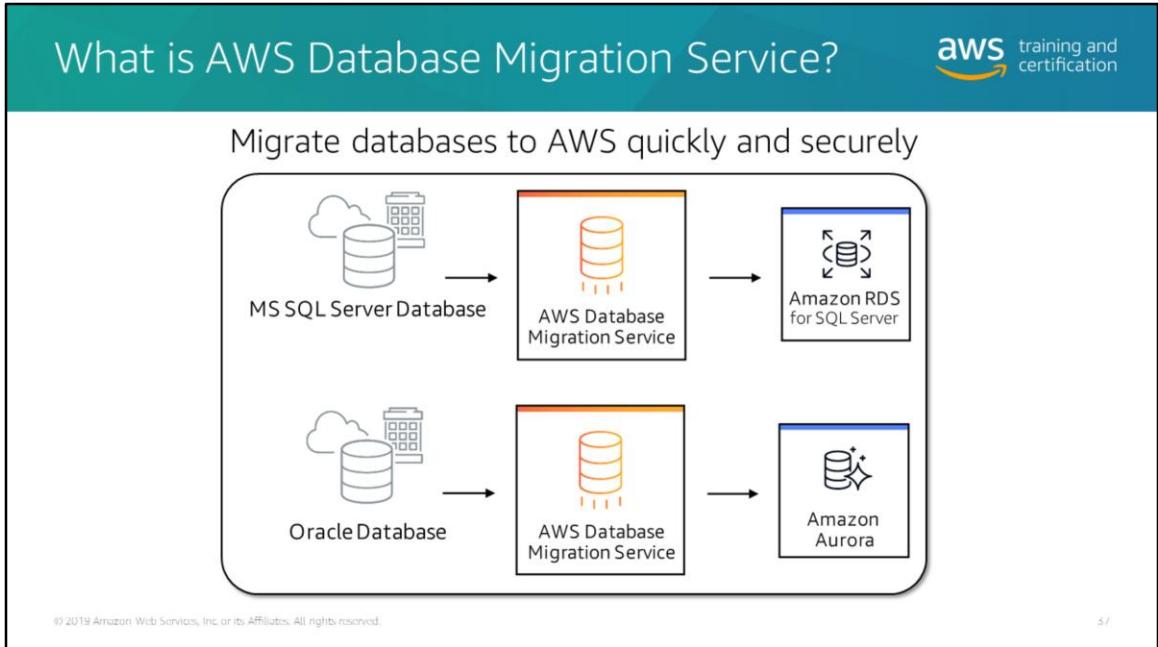
number of complex queries that help analyze data.

Amazon DocumentDB (with MongoDB compatibility) is a fast, scalable, highly available, and fully managed document database service that supports MongoDB (a cross-platform, document-oriented, NoSQL database) workloads. Amazon DocumentDB is designed to give you the performance, scalability, and availability you need when operating mission-critical MongoDB workloads at scale. Amazon DocumentDB implements the Apache 2.0 open source MongoDB 3.6 API by emulating the responses that a MongoDB client expects from a MongoDB server. This capability enables you to use your existing MongoDB drivers and tools with Amazon DocumentDB.

Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets. The core of Amazon Neptune is a purpose-built, high-performance graph database engine optimized for storing billions of relationships and querying the graph with milliseconds latency. Amazon Neptune supports popular graph models, such as Property Graph and W3C's RDF, and their respective query languages Apache TinkerPop and Gremlin-SPARQL, allowing you to easily build queries that efficiently navigate highly connected datasets.

Amazon Neptune is highly available, with read replicas, point-in-time recovery, continuous backup to Amazon S3, and replication across Availability Zones. Neptune is secure with support for encryption at rest. Because Neptune is fully managed, you no longer need to worry about database management tasks such as hardware provisioning, software patching, setup, configuration, or backups. Use cases include:

- Social networking
- Recommendation engines
- Fraud detection
- Knowledge graphs
- Life sciences
- Network/IT operations



The AWS Database Migration Service (AWS DMS) helps you migrate databases to AWS quickly and securely. The source database remains fully operational during the migration, reducing downtime to applications that rely on the database. The AWS DMS can migrate your data to and from most widely used commercial and open-source databases. The source database can be located on premises in Amazon EC2 or in Amazon RDS.

AWS DMS supports homogenous migrations, such as Oracle to Oracle, in addition to heterogeneous migrations between different database platforms, such as Oracle or Microsoft SQL Server to Amazon Aurora. With AWS DMS, you can continuously replicate your data with high availability and consolidate databases into a petabyte-scale data warehouse by streaming data to Amazon Redshift and Amazon S3. For more information on the supported source and target databases, see https://docs.aws.amazon.com/dms/latest/userguide/CHAP_Endpoints.html

The right tool for the right job	
What are my requirements?	
Enterprise class relational database	Amazon Relational Database Service (Amazon RDS)
Fast and flexible NoSQL database service for any scale	Amazon DynamoDB
Operating system access or application features not supported by AWS database services	Databases on EC2
Specific case-driven requirements (Machine learning, data warehouse, graphs)	AWS purpose-built database services

As the cloud continues to drive down the cost of storage and compute, a new generation of applications has emerged, creating a new set of requirements for databases. These applications need databases to store terabytes to petabytes of new types of data, provide access to the data with millisecond latency, process millions of requests per second, and scale to support millions of users anywhere in the world. To support these requirements, you need both relational and non-relational databases that are purpose-built to handle the specific needs of your applications. AWS offers the broadest range of databases built for your specific application use cases.

Let's review some of the essential AWS database services in more detail.

Knowledge check



Which of the following is a key-value (NoSQL) database?

- A. Amazon Aurora
- B. Amazon DynamoDB
- C. Amazon RDS
- D. Amazon NoSQL-DB

Knowledge check



Which of the following is a key-value (NoSQL) database?

- A. Amazon Aurora
- B. Amazon DynamoDB
- C. Amazon RDS
- D. ~~Amazon NoSQL DB~~

Knowledge check



Which of the following is a key-value (NoSQL) database?

- A. Amazon Aurora
- B. Amazon DynamoDB
- C. ~~Amazon RDS~~
- D. ~~Amazon NoSQL DB~~

Knowledge check



Which of the following is a key-value (NoSQL) database?

- A. Amazon Aurora
- B. Amazon DynamoDB
- C. Amazon RDS
- D. Amazon NoSQL DB

B is correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

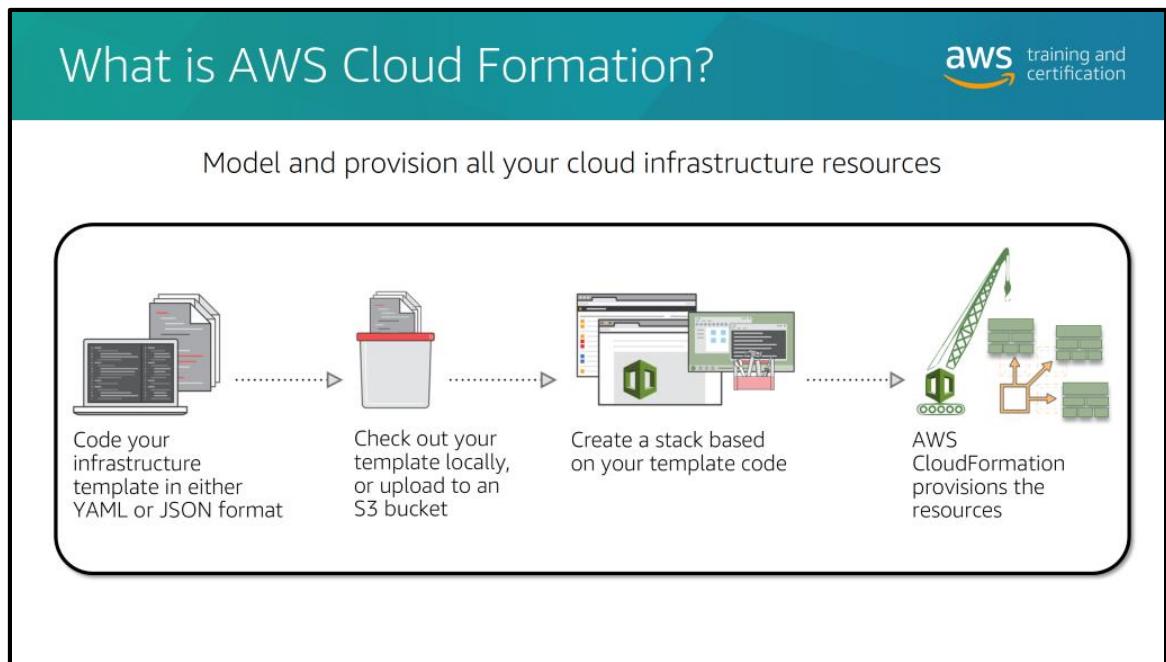
42

Automate deployment

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



4.5



Model it all

Use AWS CloudFormation to model your entire infrastructure in a text file. This template becomes the single source of truth for your infrastructure. This helps you to standardize infrastructure components used across your organization, enabling configuration compliance and faster troubleshooting.

Automate and deploy

AWS CloudFormation provisions your resources in a safe, repeatable manner, allowing you to build and rebuild your infrastructure and applications, without having to perform manual actions or write custom scripts. AWS CloudFormation takes care of determining the right operations to perform when managing your stack, and rolls back changes automatically if errors are detected.

It's only code

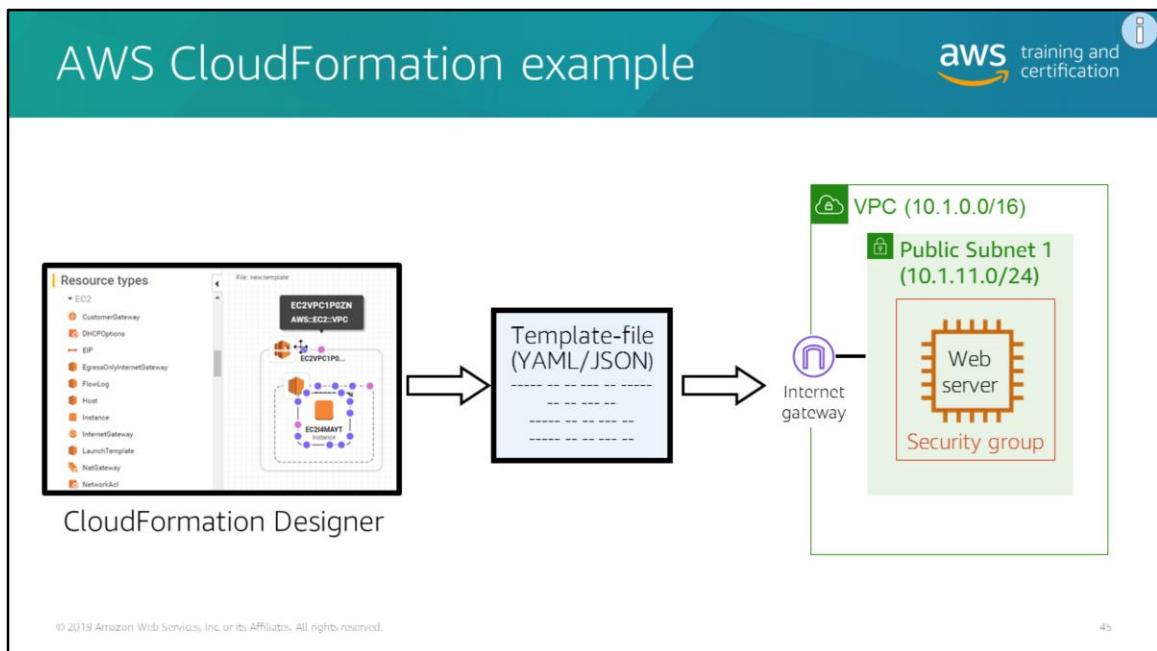
By codifying your infrastructure, you can treat your infrastructure as code. You can author it with any code editor, check it into a version control system, and review the files with team members before deploying into production. If you want to design visually, you can use AWS CloudFormation Designer to help you get started with AWS

CloudFormation templates. If you would prefer to write the templates yourself, AWS CloudFormation supports either JavaScript Object Notation (JSON) or YAML (YAML Ain't Markup Language) to describe what AWS resources you want to create and configure.

To get started with AWS CloudFormation, follow these general steps:

1. Code your infrastructure from scratch with the AWS CloudFormation template language, in either YAML or JSON format, or start from many available sample templates.
2. Check out your template code locally, or upload it into an S3 bucket.
3. To create a stack based on your template code, use AWS CloudFormation through the console, AWS CLI, or AWS APIs.

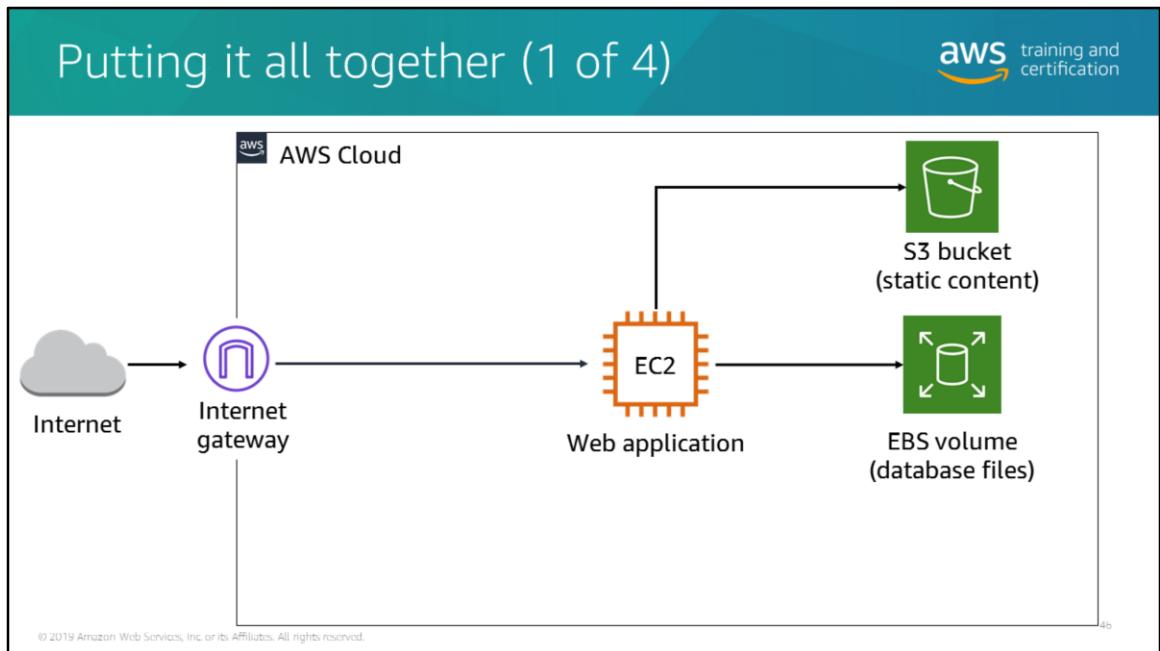
AWS CloudFormation provisions and configures the stacks and resources you specified in your template.



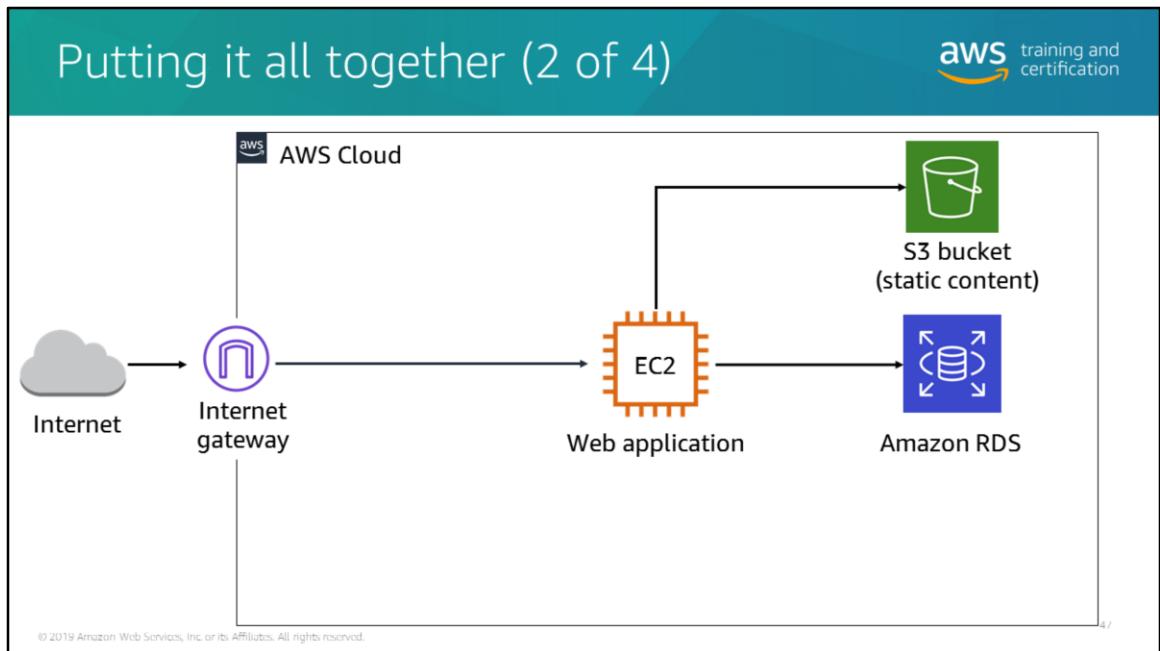
AWS CloudFormation Designer (Designer) is a visual tool for creating, viewing, and modifying AWS CloudFormation templates. With Designer, you can diagram your template resources by using a drag-and-drop interface, and then edit their details by using the integrated JSON and YAML editor. Designer helps you to quickly see the interrelationship between a template's resources and easily modify templates.

Designer is part of the AWS CloudFormation console. To use it, open Designer at <https://console.aws.amazon.com/cloudformation/designer>, and sign in with your AWS credentials.

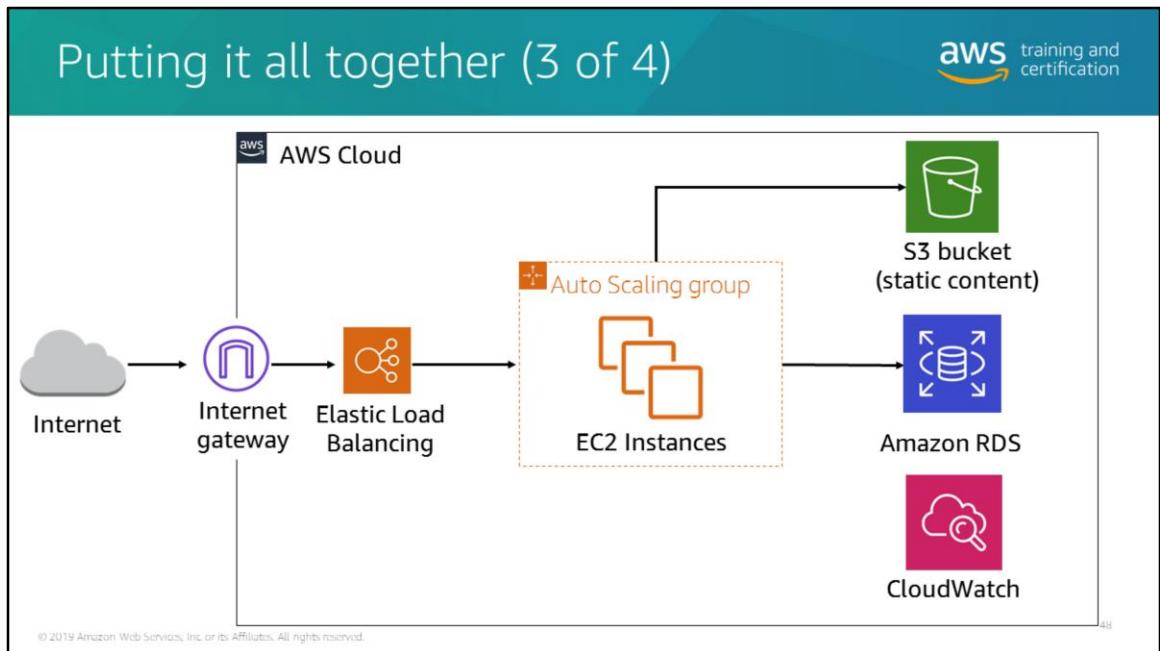
This is an example of creating a stack for a simple web server by using Designer. You can use the interface to visually build the components of your architecture and then save it as a template file that you can use to deploy your architecture in AWS.



Let's return to our project and see how we can use AWS services to improve performance, scalability, and efficiency.

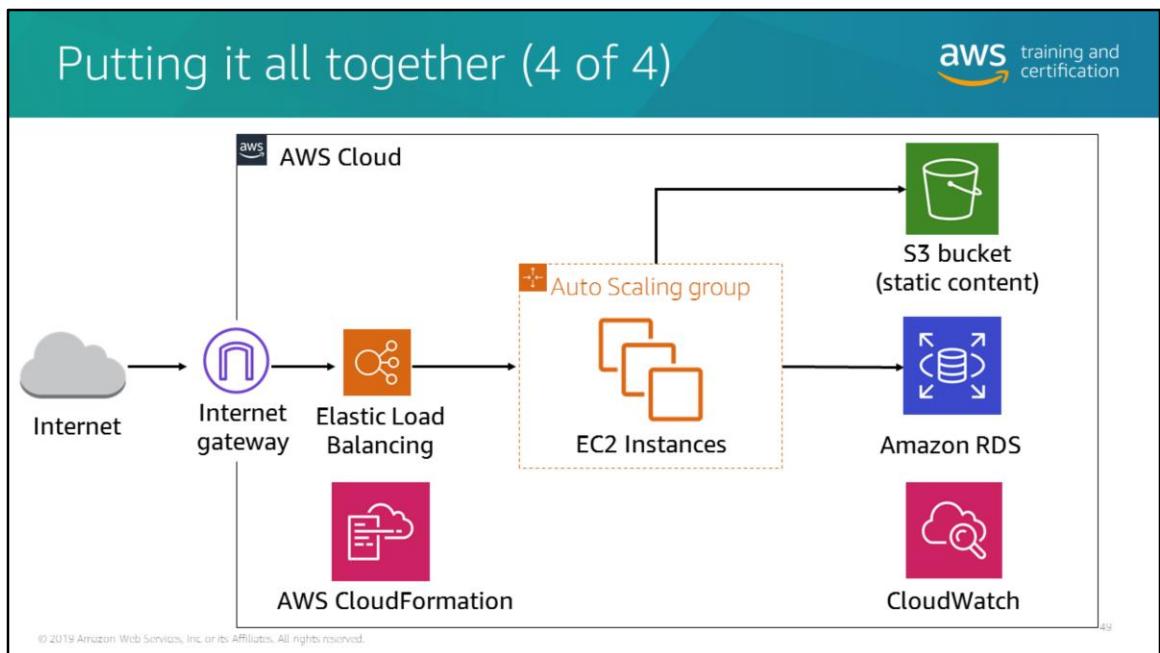


Replacing our legacy database with Amazon RDS takes much load off the instance running your application. As a managed service, Amazon RDS also decreases the amount of management and maintenance you need to do.



Replacing the single EC2 instance with an Auto Scaling group of instances greatly improves utilization. You can use CloudWatch to add and remove instances as needed, improving cost efficiency. The Elastic Load Balancing improves performance, balancing incoming requests across instances in the group.

We could make even more improvements to this architecture by using other AWS services, which we discuss later in this course.



There is deployment to consider. Now that you have a well-architected environment, it would be nice to be able to deploy all these disparate services in one common and meaningful way. AWS CloudFormation enables us to save this environment as a template that can be used quickly and easily to deploy similar environments as we grow our cloud applications.

How can I deploy without managing infrastructure?

aws training and certification

Quickly deploy and manage applications with AWS Elastic Beanstalk

- Upload your application code
- The service handles:
 - ✓ Resource provisioning
 - ✓ Load balancing
 - ✓ Automatic scaling
 - ✓ Monitoring
- Support applications that scale to serve millions of users

Application code

Sample application
Get started right away with sample code.

Upload your code
Upload a source bundle from your computer or copy one from Amazon S3.



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

50

AWS Elastic Beanstalk helps you to get web applications up and running on AWS. Developers upload their application code, and the service automatically handles all the details such as resource provisioning, load balancing, automatic scaling, and monitoring. Elastic Beanstalk is ideal if you have a PHP, Java, Python, Ruby, Node.js, .NET, Go, or Docker web application. Elastic Beanstalk uses core AWS services, such as Amazon EC2, Amazon Elastic Container Service (Amazon ECS), AWS Application Auto Scaling, and Elastic Load Balancing, to support applications that need to scale to serve millions of users. To get started, you upload your application code. The service supports the following operations:

- Resource provisioning
- Load balancing
- Automatic scaling
- Monitoring

AWS Elastic Beanstalk features



- Wide selection of application platforms
- Variety of application deployment options
- Monitoring
- Application health
- Monitoring, logging, and tracing
- Management and updates
- Scaling
- Customization
- Compliance

Wide selection of application platforms

Elastic Beanstalk supports web applications written in many popular languages and frameworks. It requires no or minimal code changes to go from development machine to the cloud. You can choose from a variety of application platforms, such as Java, .NET, Node.js, PHP, Ruby, Python, Go, and Docker, to deploy your web applications.

Variety of application deployment options

Elastic Beanstalk enables you to deploy your code through the console, Elastic Beanstalk CLI, Visual Studio, and Eclipse. You can choose from multiple deployment policies: all at once, rolling, rolling with an additional batch, immutable, and blue/green. These policy choices enable you to choose between speed and safety of deploying your applications while reducing the administrative burden.

Monitoring

Elastic Beanstalk provides a unified user interface to monitor and manage the health of your applications.

Application health

Elastic Beanstalk collects 40+ key metrics and attributes to determine the health of your application. You can use the Elastic Beanstalk Health dashboard to visualize overall application health and customize application health checks, health permissions, and health reporting in one unified interface.

Monitoring, logging, and tracing

Elastic Beanstalk integrates with CloudWatch and AWS X-Ray. You can leverage monitoring the dashboard to view key performance metrics such as latency, CPU utilization, and response codes. You can also set up CloudWatch alarms to get notified when metrics exceed your chosen thresholds.

Management and updates

You can choose to have Elastic Beanstalk automatically update to the latest version of your Elastic Beanstalk environment by using Managed Platform Updates. The immutable deployment mechanism ensures that these updates for new patches and minor platform versions are done in a safe manner to reduce the impact to users. For ongoing management, you can also customize application properties, create alarms, and enable email notifications using the Amazon Simple Notification Service (Amazon SNS).

Scaling

Elastic Beanstalk leverages Elastic Load Balancing and AWS Auto Scaling to automatically scale your application in and out based on your application's specific needs. In addition, multiple Availability Zones give you an option to improve application reliability and availability by running in more than one zone.

Customization

With Elastic Beanstalk, you have the freedom to select the AWS resources, such as Amazon EC2 instance type, that are optimal for your application. Additionally, you can use Elastic Beanstalk to retain full control over the AWS resources powering your application. If you decide that you want to take over some (or all) of the elements of your infrastructure, you can do so seamlessly by using Elastic Beanstalk management capabilities.

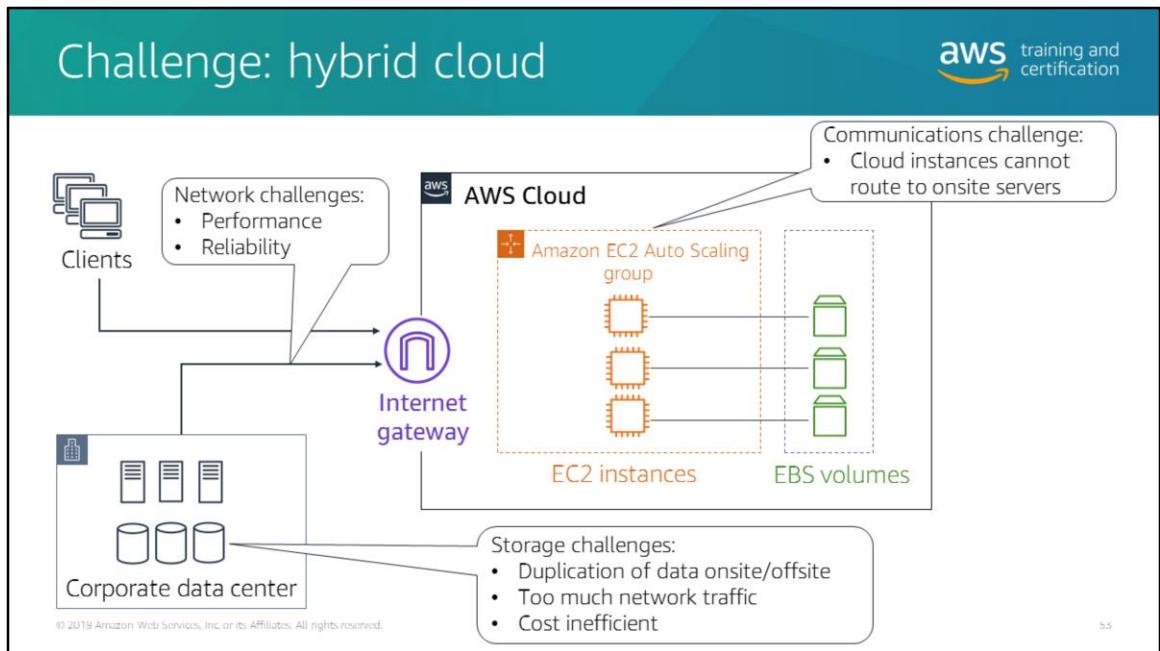
Compliance

Elastic Beanstalk meets the criteria for ISO, PCI, SOC 1, SOC 2, and SOC 3 compliance along with the criteria for HIPAA eligibility. This means that applications running on Elastic Beanstalk can process regulated financial data or protected health information (PHI).

Connect and share data

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.





Here is another example of an environment that is functional but suffers from inefficiencies. Your web application is functioning well in the cloud, but you have requirements to do some processing by using legacy applications running onsite. This results in a lot of data moving between your cloud application and your on-premises servers, which is inefficient and subject to performance issues from connecting to the cloud from the internet.

We will explore AWS services that can improve the performance, reliability, scalability, and efficiency of this solution.

What is AWS Direct Connect?



A dedicated network connection from your premises to AWS



Reduces network costs



Creates consistent network performance



Provides private connectivity to your Amazon VPC



Scales easily

AWS Direct Connect is a cloud service solution that makes it easy to establish a dedicated network connection from your premises to AWS. Using AWS Direct Connect, you can establish private connectivity between AWS and your data center, office, or colocation environment. In many cases, this capability can reduce your network costs, increase bandwidth throughput, and provide a more consistent network experience than internet-based connections.

AWS Direct Connect lets you establish a dedicated network connection between your network and one of the AWS Direct Connect locations. Using open standard 802.1q VLANs, you can partition this dedicated connection into multiple virtual interfaces. Doing so enables you to use the same connection to access public resources. These can include objects stored in Amazon S3 using a public IP address space and private resources, such as Amazon EC2 instances running within an Amazon Virtual Private Cloud (VPC) using private IP space, while maintaining network separation between the public and private environments. You can configure virtual interfaces at any time to meet your changing needs.

Reduces your bandwidth costs

If you have bandwidth-heavy workloads that you want to run in AWS, AWS Direct

Connect reduces your network costs into and out of AWS in two ways. First, by transferring data to and from AWS directly, you can reduce your bandwidth commitment to your internet service provider. Second, all data transferred over your dedicated connection is charged at the reduced AWS Direct Connect data transfer rate rather than at internet data transfer rates.

Consistent network performance

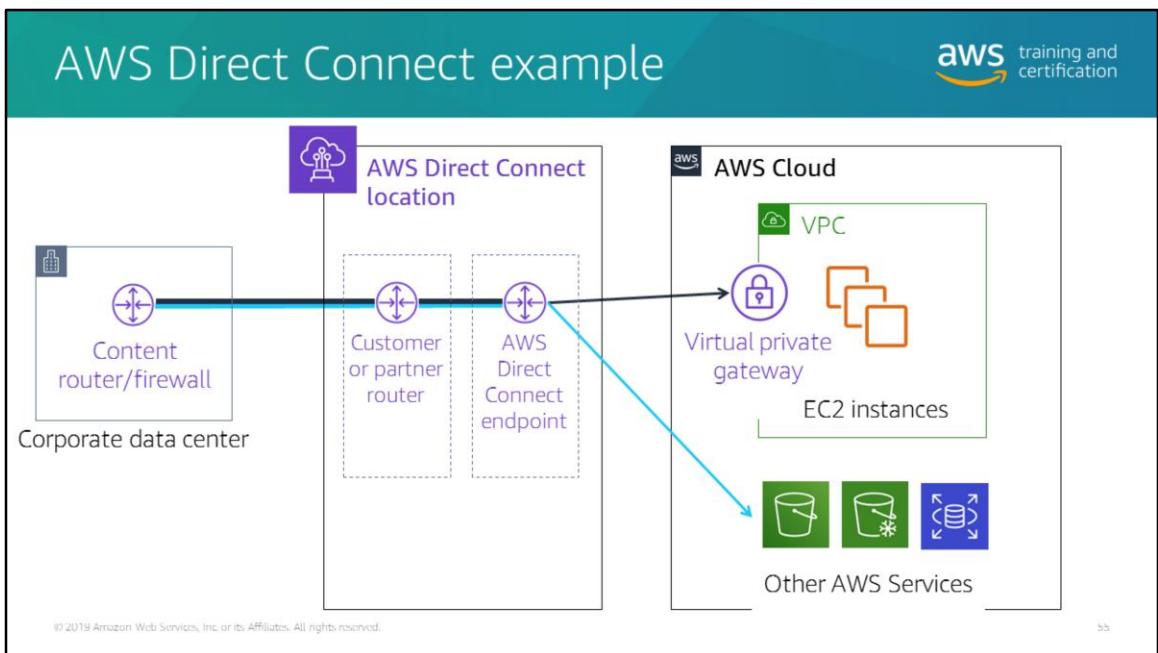
Network latency over the internet can vary given that the internet is constantly changing how data moves from point A to point B. With AWS Direct Connect, you choose the data that uses the dedicated connection and how that data is routed which can provide a more consistent network experience over Internet-based connections.

Private connectivity to your Amazon VPC

You can use AWS Direct Connect to establish a private virtual interface from your on-premises network directly to your Amazon VPC. This connectivity provides you with a private, high-bandwidth network connection between your network and your VPC. With multiple virtual interfaces, you can even establish private connectivity to multiple VPCs while maintaining network isolation.

Easy scalability

AWS Direct Connect makes it easy to scale your connection to meet your needs. AWS Direct Connect provides 1-Gbps and 10-Gbps connections. If you need more capacity, you can provision multiple connections. Instead of establishing a VPN connection over the internet to your Amazon VPC, you can also AWS Direct Connect. Using the service avoids the need to use VPN hardware that frequently can't support data transfer rates over 4 Gbps.



AWS Direct Connect links your internal network to an AWS Direct Connect location over a standard Ethernet, fiber-optic cable. One end of the cable is connected to your router, the other to an AWS Direct Connect router. With this connection, you can create *virtual interfaces* directly to public AWS services (for example, to Amazon S3) or to Amazon VPC, bypassing internet service providers in your network path. An AWS Direct Connect location provides access to AWS in the Region with which it is associated. You can use a single connection in a public Region or AWS GovCloud (US) to access public AWS services in all other public Regions.

The diagram shows how AWS Direct Connect interfaces with your network. In this example, your network is colocated with an existing AWS Direct Connect location. Other connection options include working with an AWS Direct Connect partner who is a member of the AWS Partner Network (APN) or working with an independent service provider to connect to AWS Direct Connect.

The following are the key components that you use for AWS Direct Connect:

- Connections
Create a connection in an AWS Direct Connect location to establish a network

connection from your premises to an AWS Region.

- Virtual interfaces

Create a virtual interface to enable access to AWS services. A public virtual interface enables access to public services, such as Amazon S3. A private virtual interface enables access to your VPC.

What is Amazon Route 53?



A highly available and scalable Domain Name System (DNS) web service



Register domain names



Route internet traffic to the resources for your domain



Check the health of your resources

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5b

Amazon Route 53 is a highly available and scalable Domain Name System (DNS) web service. You can use Route 53 to perform three main functions:

Register domain names

Your website needs a name, such as *example.com*. Use Route 53 to register a name, known as a *domain name*, for your website or web application.

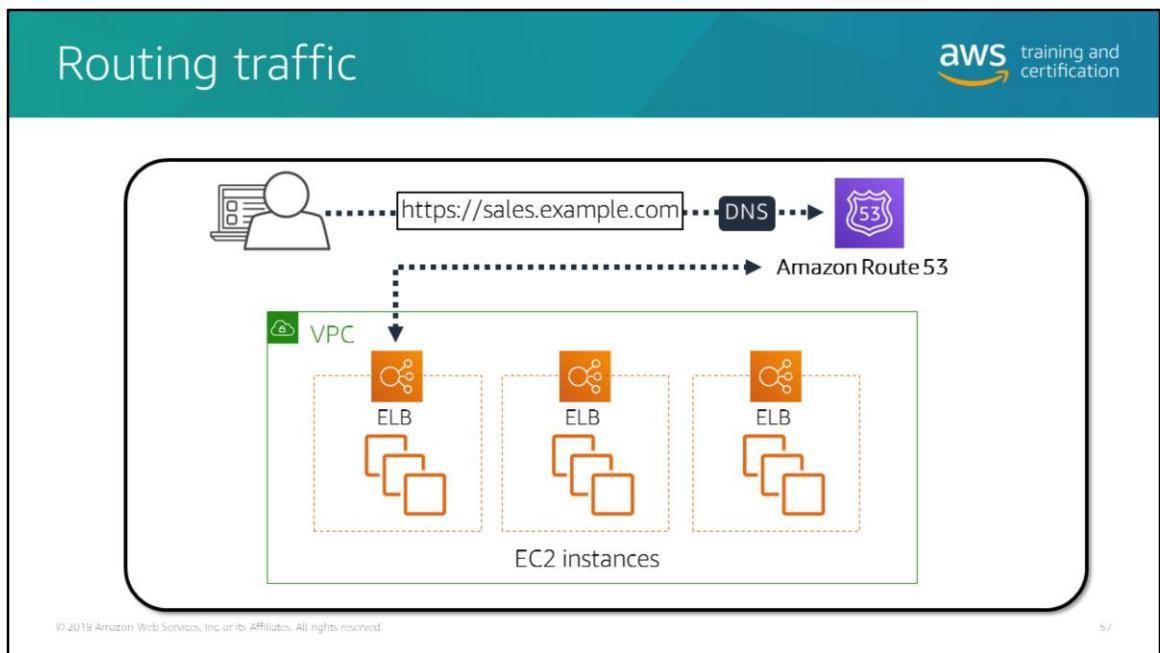
Route internet traffic to the resources for your domain

When a user opens a web browser and enters your domain name (*example.com*) or subdomain name (*apex.example.com*) in the address bar, Route 53 helps connect the browser with your website or web application.

Check the health of your resources

Route 53 sends automated requests over the internet to a resource, such as a web server, to verify that it's reachable, available, and functional. You also can choose to receive notifications when a resource becomes unavailable and choose to route internet traffic away from unhealthy resources.

You can use any combination of these functions. For example, you can use Route 53 to both register your domain name and to route internet traffic for the domain, or you can use Route 53 to route internet traffic for a domain that you registered with another domain registrar.



When you create a VPC using Amazon VPC, you automatically get DNS resolution within the VPC from the Route 53 Resolver. By default, Resolver answers DNS queries for VPC domain names such as domain names for EC2 instances or Elastic Load Balancing load balancers. Resolver performs recursive lookups against public name servers for all other domain names. You can also configure DNS resolution between your VPC and your network over a AWS Direct Connect or VPN connection.

Forward DNS queries from resolvers on your network to the Route 53 Resolver.
DNS resolvers on your network can forward DNS queries to Resolver in a specified VPC. This allows your DNS resolvers to easily resolve domain names for AWS resources, such as EC2 instances or records, in a Route 53 private hosted zone.

Conditionally, forward queries from a VPC to resolvers on your network.
You can configure Resolver to forward queries that it receives from EC2 instances in your VPCs to DNS resolvers on your network. To forward selected queries, you create Resolver rules that specify the domain names for the DNS queries that you want to forward (such as `example.com`), and the IP addresses of the DNS resolvers on your network that you want to forward the queries to. If a query matches multiple rules (`example.com`, `acme.example.com`), Resolver chooses the rule with the most specific match (`acme.example.com`) and forwards the query to the IP addresses that you specified in that rule.

What is Amazon Elastic File System (Amazon EFS)?



A scalable, elastic, cloud-native file system for Linux

-  Dynamic elasticity
-  Scalable performance
-  Shared file storage
-  Fully managed
-  Cost-effective

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

58

Dynamic elasticity

Amazon EFS automatically and instantly scales the storage capacity of your file system up or down as you add or remove files without disrupting your applications. This capability gives you the storage you need when you need it. You create your file system and start adding files with no need to provision storage in advance.

Scalable performance

Amazon EFS is designed to provide the throughput, IOPS, and low latency needed for Linux workloads. Throughput and IOPS scale as a file system grows and can burst to higher throughput levels for short periods of time to support the unpredictable performance needs of file workloads. For the most demanding workloads, Amazon EFS can support performance over 10 GB/sec and up to 500,000 IOPS.

Shared file storage

Amazon EFS provides secure access for thousands of connections. Amazon EC2 instances and on-premises servers can simultaneously access a shared Amazon EFS file system by using a traditional file permissions model, file locking capabilities, and a hierarchical directory structure through the NFSv4 protocol. Amazon EC2 instances

can access your file system across Availability Zones and Regions, whereas on-premises servers can access the file system by using AWS Direct Connect or AWS VPN.

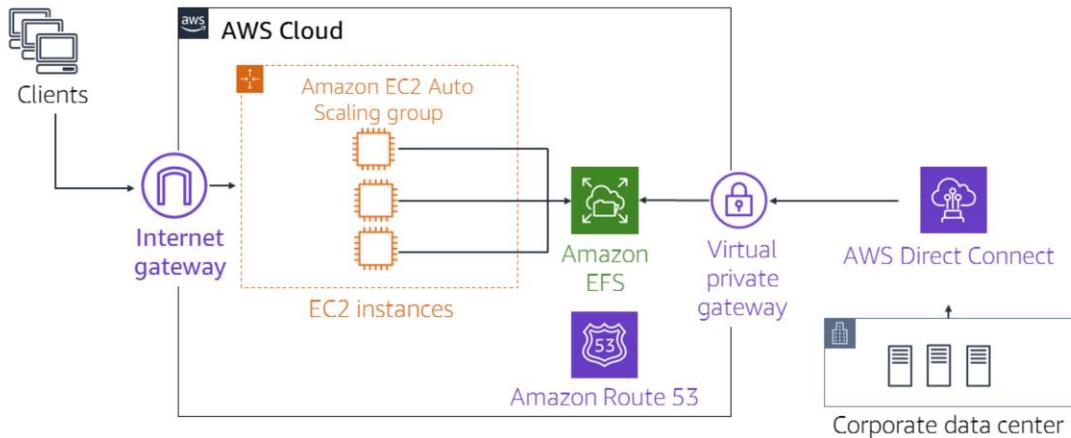
Fully managed

Amazon EFS is a fully managed service providing shared file system storage for Linux workloads. Its simple interface enables you to create and configure file systems quickly. The service also manages the file storage infrastructure for you, removing the complexity of deploying, patching, and maintaining the underpinnings of a file system.

Cost-effective

With Amazon EFS storage, you pay only for what you use. There is no need to provision storage in advance, and there are no minimum commitments or upfront fees. With Lifecycle Management, you can automatically move files that have not been accessed for 30 days to a cost-optimized storage class, reducing storage costs by up to 85%.

Putting it all together



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

59

Here is an improved solution. All application data is now stored in the cloud in a shared elastic file system. On-premises servers have fast, reliable access to AWS through AWS Direct Connect. Route 53 provides resolution as needed for cloud and local compute resources to find each other.

Deliver content faster

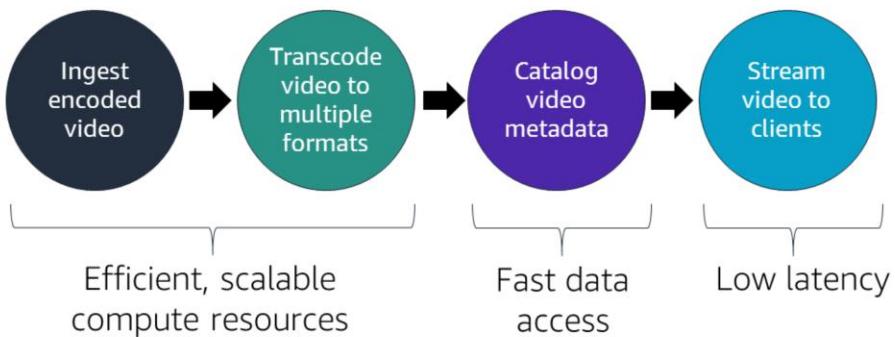
© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Challenge: Media streaming service



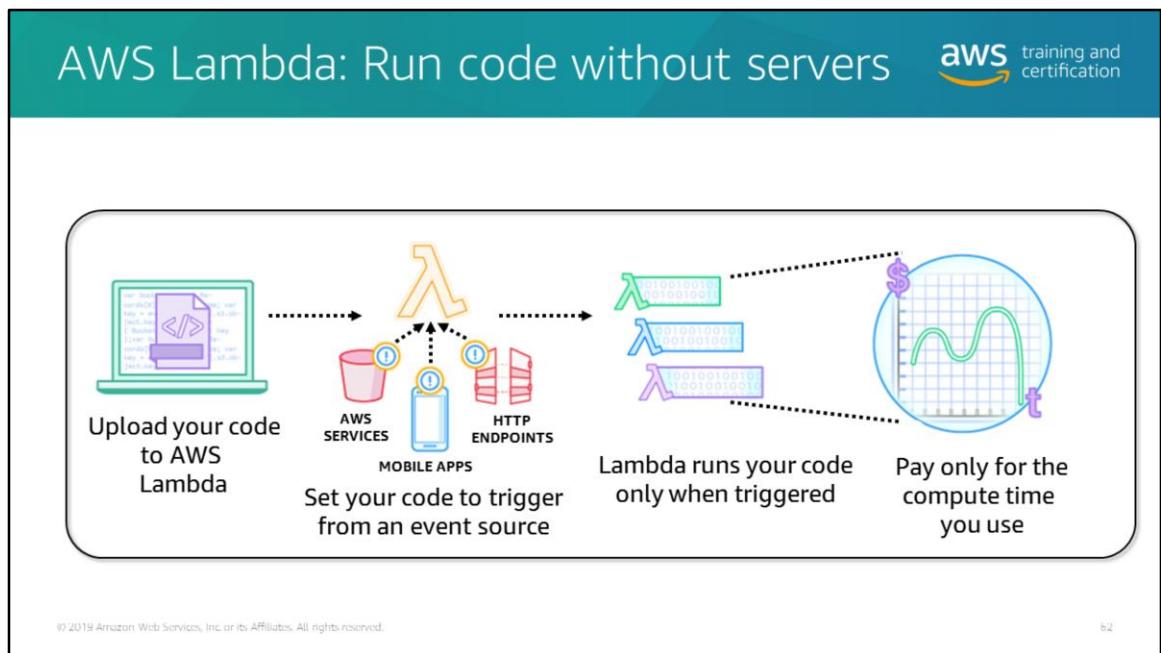
The architecture must meet the following requirements:



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

b1

In this example, rather than improving an existing environment, you will build a cloud-native architecture. The requirements for your solution are shown in this diagram. You can meet most of these requirements by using AWS services that you have already learned about, but there are additional services in this topic that will make your architecture faster, more scalable, and more cost-efficient.



You can use AWS Lambda to run code without provisioning or managing servers. You pay only for the compute time you consume—there is no charge when your code is not running.

The service enables you to run code for virtually any type of application or backend service. After you upload your code, Lambda takes care of everything required to run and scale your code with high availability.

You can set up your code to automatically trigger from other AWS services, or you can call it directly from any web or mobile app.

Benefits of Lambda



-  Supports multiple programming languages
-  Completely automated administration
-  Built-in fault tolerance
-  Supports orchestration of multiple functions
-  Pay per use pricing

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

b5

Bring your own code

With Lambda, there are no new languages, tools, or frameworks to learn. Lambda supports Java, Node.js, C#, and Python code. You can use any library, native or third-party.

Completely automated administration

Lambda manages all the infrastructure to run your code on highly available, fault-tolerant infrastructure, freeing you to focus on building differentiated backend services. With Lambda, you seldom have to update the underlying OS when a patch is released or worry about resizing or adding new servers as your usage grows. Lambda seamlessly deploys your code, does all the administration, maintenance, and security patches, and provides built-in logging and monitoring through CloudWatch.

Built-in fault tolerance

Lambda has built-in fault tolerance. Lambda maintains compute capacity across multiple Availability Zones in each Region to help protect your code against individual machine or data center facility failures. Both Lambda and the functions running on the service provide predictable and reliable operational performance. The service was designed to provide high availability for both the service itself and for the

functions it operates. There are no maintenance windows or scheduled downtimes.

Automatic scaling

Lambda invokes your code only when needed and automatically scales to support the rate of incoming requests without requiring you to configure anything. There is no limit to the number of requests your code can handle. Lambda typically starts running your code within milliseconds of an event, and because Lambda scales automatically, the performance remains consistently high as the frequency of events increases. Because your code is stateless, Lambda can start as many instances of it as needed without lengthy deployment and configuration delays.

Orchestrate multiple functions

You can coordinate multiple Lambda functions for complex or long-running tasks by building workflows with AWS Step Functions. Use Step Functions to define workflows that trigger a collection of Lambda functions by using sequential, parallel, branching, and error-handling steps. With Step Functions and Lambda, you can build stateful, long-running processes for applications and backends.

Integrated security model

Lambda allows your code to securely access other AWS services through its built-in AWS SDK and integration with IAM. Lambda runs your code within a VPC by default. You can also configure Lambda to access resources behind your own VPC, allowing you to leverage custom security groups and network access control lists to provide your Lambda functions access to your resources within a VPC.

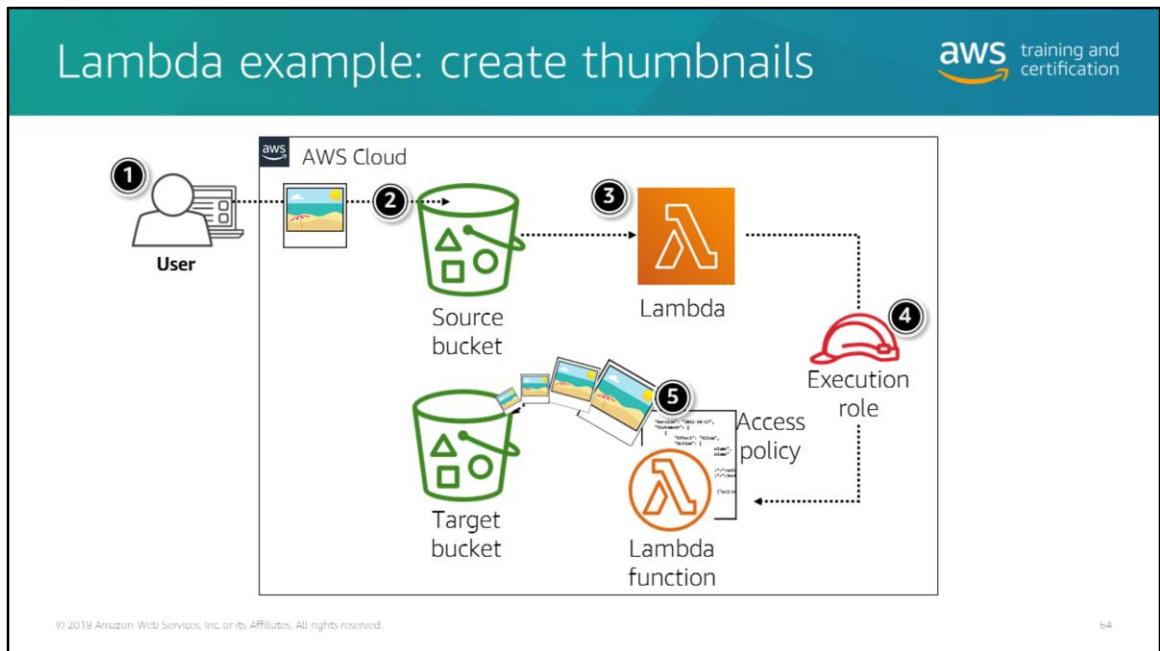
Lambda is SOC-, HIPAA-, PCI-, and ISO-compliant. For the latest in Lambda certification and compliance readiness, see the full services in scope.

Pay per use

With Lambda, you pay only for the requests served and the compute time required to run your code. Billing is metered in increments of 100 milliseconds, making it cost-effective and easy to scale automatically from a few requests per day to thousands per second.

Flexible resource model

You choose the amount of memory you want to allocate to your functions, and Lambda allocates proportional CPU power, network bandwidth, and disk I/O.



Suppose that you want to create a thumbnail for each image (.jpg and .png objects) that is uploaded to an S3 bucket. You can create a Lambda function that Amazon S3 can invoke when objects are created. Then the Lambda function can read the image object from the source bucket and create a thumbnail image in a target bucket. Here's how it works:

1. A user uploads an object to the source bucket in Amazon S3 (object-created event).
2. Amazon S3 detects the object-created event.
3. Amazon S3 publishes the `s3:ObjectCreated:*` event to Lambda by invoking the Lambda function and passing event data as a function parameter.
4. Lambda executes the Lambda function by assuming the execution role that you specified at the time you created the Lambda function.
5. From the event data it receives, the Lambda function knows the source bucket name and object key name. The Lambda function reads the object and creates a thumbnail using graphics libraries, and saves it to the target bucket.

Knowledge check



What is the first step in getting started with AWS Lambda?

- A. Provision EC2 instances.
- B. Deploy an OS image.
- C. Pay for estimated compute time.
- D. Upload your code.

Knowledge check



What is the first step in getting started with AWS Lambda?

- A. Provision EC2 instances.
- B. Deploy an OS image.
- C. Pay for estimated compute time.
- D. Upload your code.

Knowledge check



What is the first step in getting started with AWS Lambda?

- A. Provision EC2 instances.
- ~~B. Deploy an OS image.~~
- ~~C. Pay for estimated compute time.~~
- D. Upload your code.

Knowledge check



What is the first step in getting started with AWS Lambda?

- A. Provision EC2 instances.
- B. Deploy an OS image.
- C. Pay for estimated compute time.
- D. Upload your code.

D is correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

68

What is Amazon Simple Notification Service (Amazon SNS)?



Fully managed pub/sub messaging for distributed or serverless applications



Reliably deliver messages with durability



Automatically scale your workload



Simplify your architecture



Keep messages private and secure

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

by

Reliably deliver messages with durability

Amazon SNS provides durable storage for all the messages it receives. Before Amazon SNS confirms to you that it received your request, it stores the message in multiple Availability Zones. Running within the Amazon network infrastructure and data centers, Amazon SNS topics are available whenever your applications need them. All messages published to Amazon SNS are stored redundantly across multiple geographically separated servers and data centers. Amazon SNS reliably delivers messages to all valid AWS endpoints, such as Amazon SQS queues and AWS Lambda functions.

Automatically scale your workload

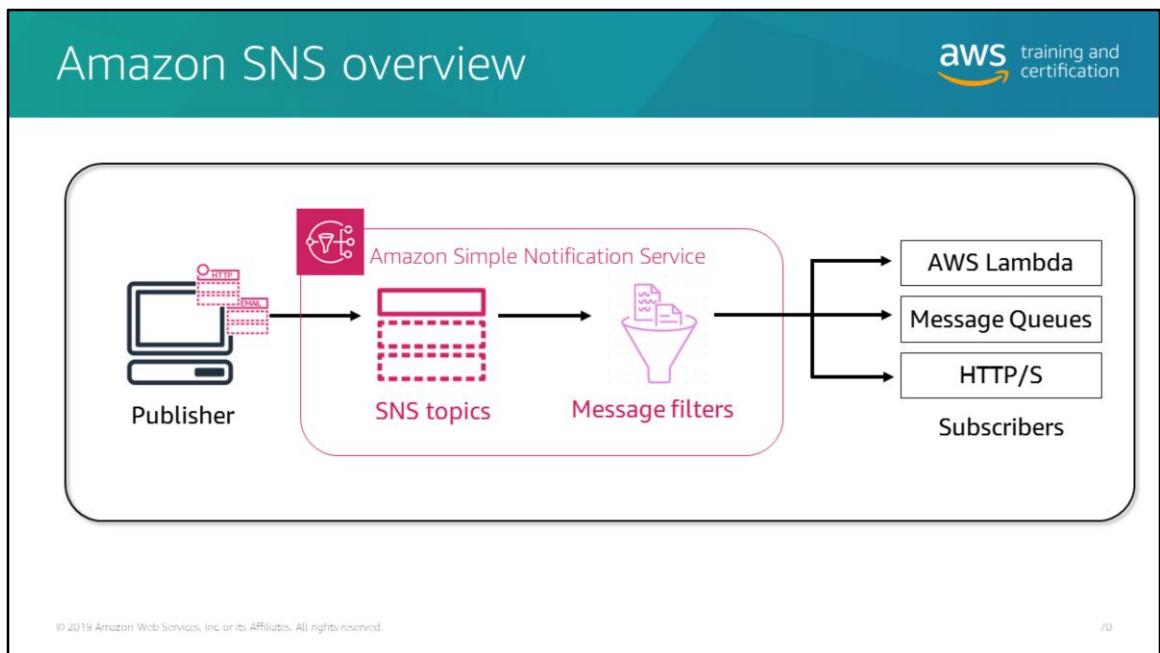
Amazon SNS leverages the AWS Cloud to dynamically scale with your application. Amazon SNS is a fully managed service, taking care of the cumbersome tasks related to capacity planning, provisioning, monitoring, and patching. The service is designed to handle high-throughput, bursty traffic patterns. Moreover, there is no upfront cost, and no need to acquire, install, configure, or upgrade messaging software.

Simplify your architecture with message filtering

Amazon SNS helps you simplify your pub/sub messaging architecture by offloading the message filtering logic from your subscriber systems and the message routing logic from your publisher systems. With Amazon SNS message filtering, subscribing endpoints receive only the messages of interest instead of all messages published to the topic. Amazon CloudWatch gives visibility into your filtering activity, and AWS CloudFormation enables you to deploy subscription filter policies in an automated and secure manner.

Keep messages private and secure

Amazon SNS topic owners can keep sensitive data secure by setting topic policies that restrict who can publish and subscribe to a topic. Amazon SNS also ensures that data is encrypted in transit by applying Amazon ATS certificates to support its HTTPS API, and it can also encrypt data at rest by using AWS KMS keys. Additionally, by using AWS PrivateLink, you can privately publish messages to Amazon SNS topics from your Amazon VPC subnets without traversing the public internet. Amazon SNS can also support use cases in regulated markets, and is in-scope with compliance programs, including HIPAA, PCI, ISO, FIPS, SOC, and FedRAMP.



Amazon SNS enables message filtering and fanout to a large number of subscribers, including serverless functions, queues, and distributed systems. Additionally, Amazon SNS fans out notifications to users through mobile push messages, SMS, and email.

A topic is like a broadcasting station. You can publish messages to a topic, and anyone interested in these messages can subscribe to the topic. Then, the interested parties are notified about the published messages. The software that broadcasts topics is called a *topic publisher* and the software that subscribes to broadcasts is called a *topic subscriber*.

Here are some use cases for messaging:

- Service-to-service communication

You have two services or systems that need to communicate with each other.

Suppose that a website (the frontend) has to update a customer's delivery address in a customer relationship management (CRM) system (the backend). Alternatively, you can set up a load balancer in front of the backend CRM service and call its API actions directly from the frontend website. You can also set up a queue: have the frontend website code send messages to the queue and have the backend CRM service to consume them.

- Asynchronous work item backlogs

You have a service that has to track a backlog of actions to be executed. Suppose a hotel booking system needs to cancel a booking, and this process takes a long time (from a few seconds to a minute). You can execute the cancellation synchronously, but then you risk annoying the customer who has to wait for the webpage to load. You can also track all pending cancellations in your database and keep polling and executing cancellations. Alternatively, you can put a message into a queue and have the same hotel booking system consume messages from that queue and perform asynchronous cancellations.

- State change notifications

You have a service that manages some resource and other services that receive updates about changes to those resources. Suppose that an inventory tracking system tracks products stocked in a warehouse. Whenever the stock is sold out, the website must stop offering that product. Whenever the stock is close to being depleted, the purchasing system must place an order for more items. Those systems can keep querying the inventory system to learn about these changes (or even directly examine the database). Alternatively, the inventory system can publish notifications about stock changes to a topic and any interested program can subscribe to learn about those changes.

What is Amazon CloudFront?

A fast, secure, and global content delivery network (CDN)

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Amazon CloudFront is a fast content delivery network (CDN) service that securely delivers data, videos, applications, and APIs to customers globally with low-latency and high-transfer speeds, all within a developer-friendly environment.

CloudFront works seamlessly with AWS services, including AWS Shield for DDoS mitigation, Amazon S3, Elastic Load Balancing, or Amazon EC2 as origins for your applications, and Lambda@Edge to run custom code closer to customers' users and to customize the user experience.

Some of the advantages of CloudFront are:

CloudFront global edge network

To deliver content with lower latency to end users, CloudFront uses a global network of 138 Points of Presence (127 edge locations and 11 regional edge caches) in 63 cities across 29 countries.

Faster performance

Provides network optimizations for better performance.

Secure

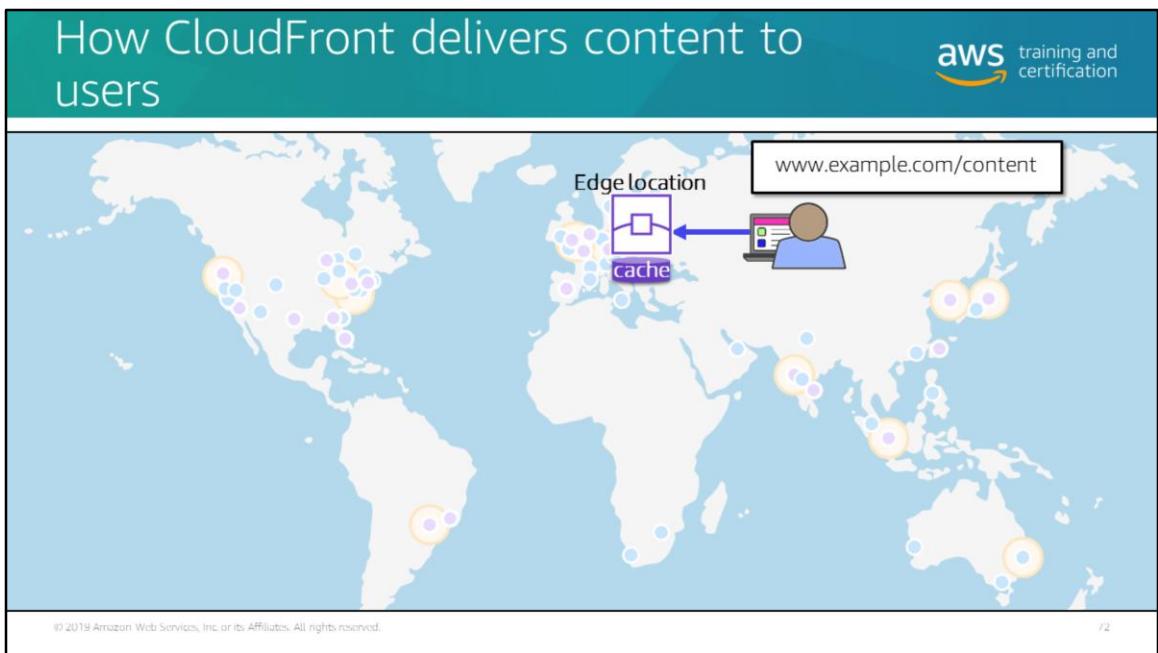
Provides protection against network and application layer attacks.

Programmable and DevOps-friendly

Includes full-featured API operations and DevOps tools

Cost Effective

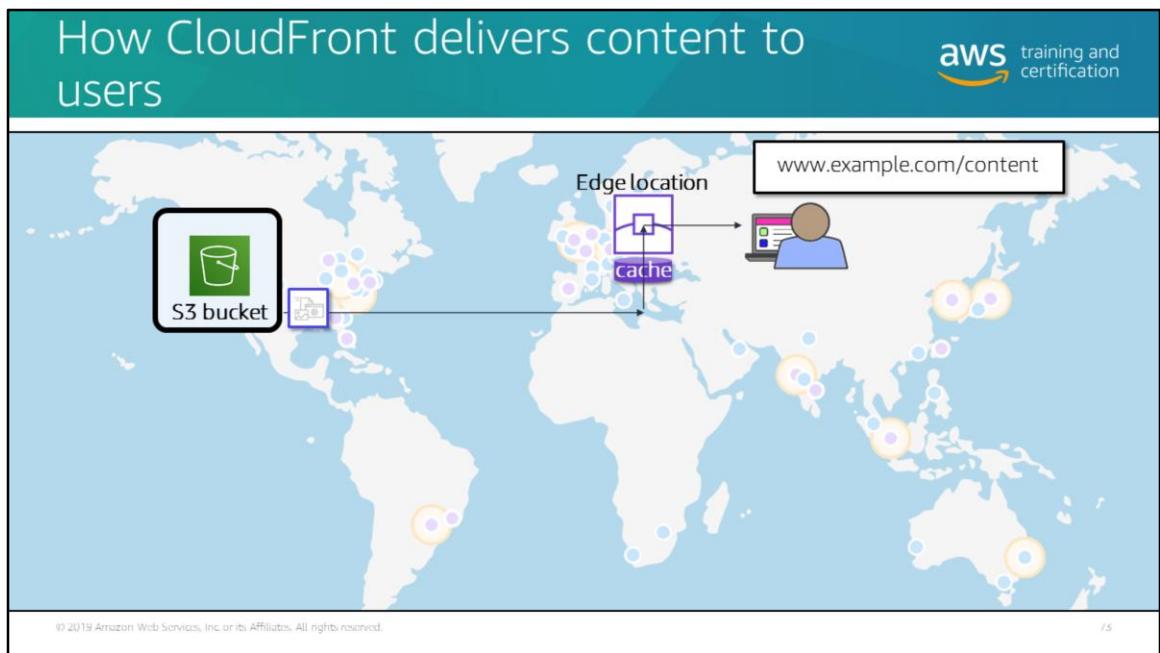
Provides flexible pay-as-you-go publicly available pricing and committed-traffic private pricing.



After you configure CloudFront to deliver your content, here's what happens when users request your objects:

1. A user accesses your website or application and requests one or more objects, such as an image file and an HTML file.
2. DNS routes the request to the CloudFront edge location that can best serve the request—typically the nearest CloudFront edge location in terms of latency—and routes the request to that edge location.
3. In the edge location, CloudFront checks its cache for the requested files. If the files are in the cache, CloudFront returns them to the user. If the files are not in the cache, it does the following:
 - a) CloudFront compares the request with the specifications in your distribution. Then it forwards the request for the files to the applicable origin server for the corresponding file type—for example, to your Amazon S3 bucket for image files and to your HTTP server for the HTML files.
 - b) The origin servers send the files back to the CloudFront edge location.
 - c) As soon as the first byte arrives from the origin, CloudFront begins to forward the files to the user. CloudFront also adds the files to the cache in the edge location for the next time someone requests those files.

You can control how long your files stay in a CloudFront cache before CloudFront forwards another request to your origin. Reducing the duration allows you to serve dynamic content. Increasing the duration means that your users get better performance because your files are more likely to be served directly from the edge cache. A longer duration also reduces the load on your origin. By default, each file automatically expires after 24 hours, but you can change the default behavior by adjusting the TTL (Time To Live) settings for CloudFront.



After you configure CloudFront to deliver your content, here's what happens when users request your objects:

1. A user accesses your website or application and requests one or more objects, such as an image file and an HTML file.
2. DNS routes the request to the CloudFront edge location that can best serve the request—typically the nearest CloudFront edge location in terms of latency—and routes the request to that edge location.
3. In the edge location, CloudFront checks its cache for the requested files. If the files are in the cache, CloudFront returns them to the user. If the files are not in the cache, it does the following:
 - a) CloudFront compares the request with the specifications in your distribution. Then it forwards the request for the files to the applicable origin server for the corresponding file type—for example, to your Amazon S3 bucket for image files and to your HTTP server for the HTML files.
 - b) The origin servers send the files back to the CloudFront edge location.
 - c) As soon as the first byte arrives from the origin, CloudFront begins to forward the files to the user. CloudFront also adds the files to the cache in the edge location for the next time someone requests those files.

You can control how long your files stay in a CloudFront cache before CloudFront forwards another request to your origin. Reducing the duration allows you to serve dynamic content. Increasing the duration means that your users get better performance because your files are more likely to be served directly from the edge cache. A longer duration also reduces the load on your origin. By default, each file automatically expires after 24 hours, but you can change the default behavior by adjusting the TTL (Time To Live) settings for CloudFront.

What is Amazon ElastiCache?



Fully managed Redis or Memcached-compatible in-memory data store



Extreme performance



Fully Managed



Scalable



Amazon ElastiCache for Redis

Versatile in-memory data store



Amazon ElastiCache for Memcached

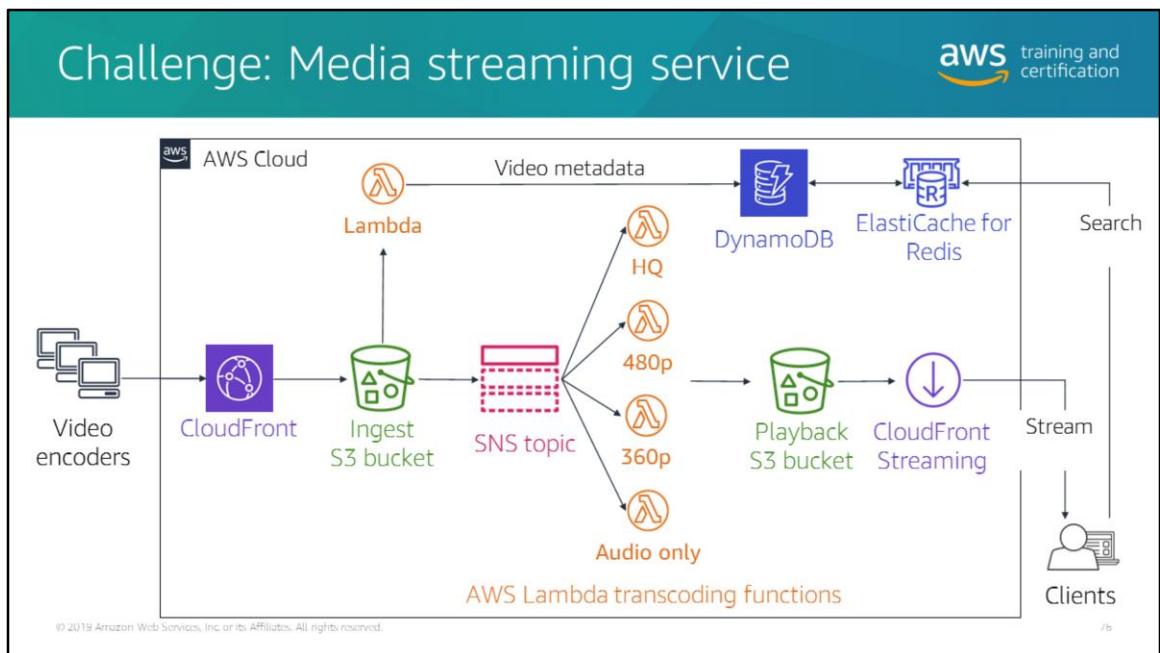
Scalable caching tier for data-intensive apps

Amazon ElastiCache is a web service that makes it easy to deploy, operate, and scale an in-memory data store or cache in the cloud. The service improves application performance by allowing developers to retrieve information from fast, managed, in-memory data stores instead of relying on slower disk-based databases. When a read request is sent, the caching layer checks to determine whether it has the answer. If it doesn't, the request is sent to the database. Meeting read requests through the caching layer in this manner is more efficient and delivers higher performance than what can be had from a traditional database alone. It is also more cost-effective. Amazon ElastiCache supports two open-source in-memory engines: *Redis* and *Memcached*.

The primary use case for Memcached is caching; it is easy to use and scale. ElastiCache is protocol-compliant with Memcached, so tools used with existing Memcached environments work seamlessly with ElastiCache. Memcached is well suited for caching relatively small and static data whereby the primary concern is fast read performance.

Redis is an in-memory NoSQL data store that supports persistence, availability, and scripting. It comes with a set of versatile in-memory data structures that make it easy

to create a variety of custom applications. Redis is often used for caching, session management, pub/sub, and leaderboards. Because of its speed and ease of use, Redis is a popular choice for web, mobile, gaming, adtech and IoT applications that require proven performance. Redis has a broader set of features than Memcached and performs well for both reads and writes.



Here is an architecture that meets the requirements stated earlier. The solution must broadcast live video with adaptive bitrates, which requires a transcoding of the video stream to multiple quality levels for consumption on desktop, mobile, and connected devices. Encoded video is published to Amazon S3, through an Amazon CloudFront distribution for accelerated uploads. When the files have been written to Amazon S3, the operation triggers a message to be published to an SNS topic. This sends the Amazon S3 event to multiple Lambda functions to be processed independently. Using multiple Lambda functions allows for efficient, serverless processing. Transcoded video that is ready for playback is stored in a playback S3 bucket and streamed to clients by using CloudFront to minimize latency.

Another requirement was to catalog video metadata. When video files are uploaded, a separate Lambda function is triggered to extract the metadata and catalog it in a DynamoDB database. This is good for long-term storage, but users will expect very high performance when they search for content to view. To satisfy this requirement, ElastiCache for Redis has been set up in front of the database to manage the content index and user authentication tokens in-memory for submilliseconds responses at scale.

Key Takeaways



Amazon CloudWatch	Have complete visibility of your cloud resources and applications
Elastic Load Balancing Application Auto Scaling	Deploy highly available applications that scale with demand
AWS Database Services	Run SQL or NoSQL databases without the management overhead
AWS CloudFormation	Programmatically deploy repeatable infrastructure
AWS Elastic Beanstalk	Deploy your application in the simplest way possible
AWS Direct Connect	Provision a dedicated network connection from your premises to AWS
Amazon Route 53	Run a highly available and scalable Domain Name System (DNS) web service
AWS Lambda	Run code without managing servers
Amazon CloudFront	Deliver your content across a massively scaled and globally available network

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

11

Lab 2: Build a Web Server on Amazon EC2



- Create a VPC
- Create subnets
- Configure a security group
- Launch an EC2 instance into a VPC

Module 4: Security



Module goals



- Secure your infrastructure
- Manage authentication and authorization
- Assess your security and compliance
- Protect your infrastructure from Distributed Denial of Service (DDoS) attacks
- Maintain compliance

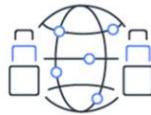
Secure your infrastructure

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Security is our top priority

aws training and certification



Designed for security Constantly monitored Highly automated Highly available Highly accredited

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

At AWS, cloud security is the highest priority. As an AWS customer, you will benefit from a data center and network architecture that is built to meet the requirements of the most security-sensitive organizations.

The AWS infrastructure has been architected to be one of the most flexible and secure cloud computing environments available today. It is designed to provide an scalable, highly reliable platform that enables customers to deploy applications and data quickly and securely.

All AWS customers benefit from a data center and network architecture built to satisfy the requirements of AWS most security-sensitive customers. This means that you get a resilient infrastructure, designed for high security, without the capital outlay and operational overhead of a traditional data center.

AWS innovates rapidly at scale, continually incorporating your feedback into AWS services. This benefits you because AWS solutions improve over time, and core security services are continually evolving. These include identity and access management (IAM), logging and monitoring, encryption and key management,

network segmentation, and standard DDoS protection. You also get advanced security services designed by engineers with deep insight into global security trends, which enables your team to proactively address emerging risks in real time. This means that you can choose the security that meets your needs as you grow, without incurring upfront expenses and with much lower operational costs than if you manage your own infrastructure.

Automating security tasks on AWS enables you to be more secure by reducing human configuration errors and giving your team more time to focus on other work that is critical to your business. Your security teams can use security automation and API integration to become more responsive and agile, making it easier to work closely with developer and operations teams to create and deploy code faster and more securely.

The slide has a teal header bar with the title 'Security of the cloud' and the AWS logo. Below the header is a white content area. Inside the content area, there is a rounded rectangle containing a bulleted list. Below this is a large orange box labeled 'AWS' on its left side. The orange box contains two rows of service names: 'Foundation services' (Compute, Storage, Database, Network) and 'AWS global infrastructure' (Availability Zones, Regions, Edge Locations). At the bottom of the slide, there is a small footer with copyright information.

- Hosts, network, software, facilities
- Protection of the AWS global infrastructure is top priority
- Availability of third-party audit reports

AWS

Foundation services

Compute Storage Database Network

AWS global infrastructure

Availability Zones Regions Edge Locations

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Because security is a priority for us, let's review the AWS shared responsibility model.

After you, the customer, starts using AWS, Amazon shares the responsibility of securing the data in AWS with you, making AWS security a shared responsibility. This concept is known as the shared responsibility model of cloud security.

Because AWS customers retain control over their data, they also retain responsibilities relating to that content as part of the AWS “shared responsibility” model.

Let's examine this closer to determine what security AWS and the customer are responsible for in this model.

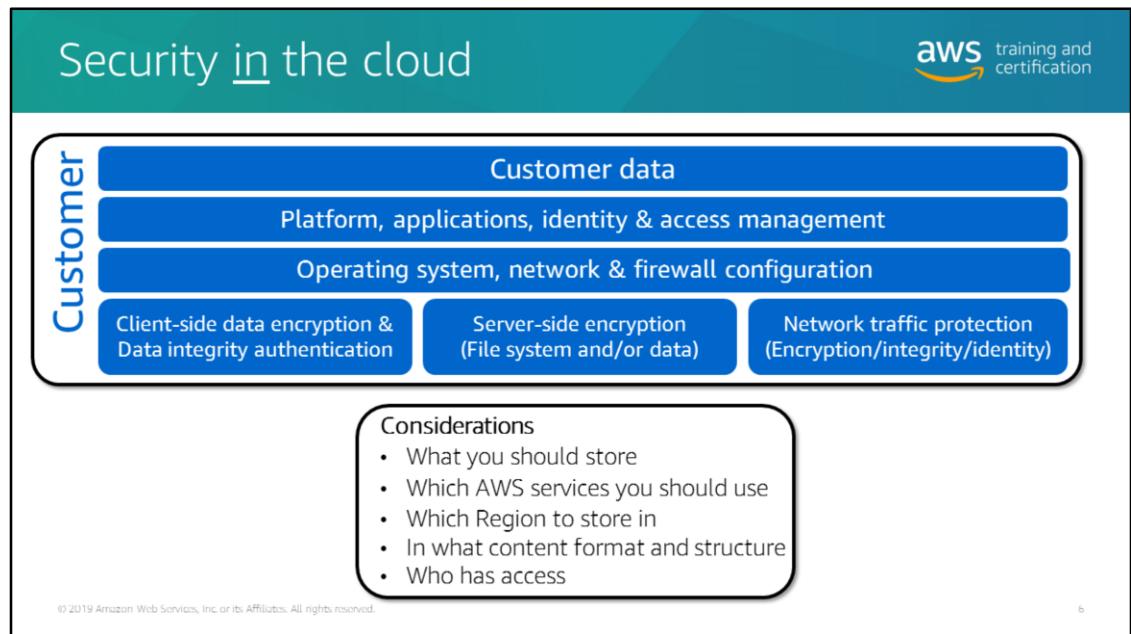
Security of the cloud

AWS is responsible for what is known as security **of** the cloud.

Under the shared responsibility model, AWS operates, manages, and controls the components from the host operating system and virtualization layer down to the physical security of the facilities in which the services operate. It means that AWS is

responsible for protecting the global infrastructure that runs all the services offered in the AWS Cloud, including AWS Regions, Availability Zones, and edge locations. The AWS global infrastructure includes the facilities, network, hardware, and operational software (e.g., host OS, virtualization software, etc.) that support the provisioning and use of these resources.

Protecting this infrastructure is the number one priority of AWS. Although you can't visit AWS data centers or offices to see this protection firsthand, Amazon provides several reports from third-party auditors who have verified its compliance with a variety of computer security standards and regulations.



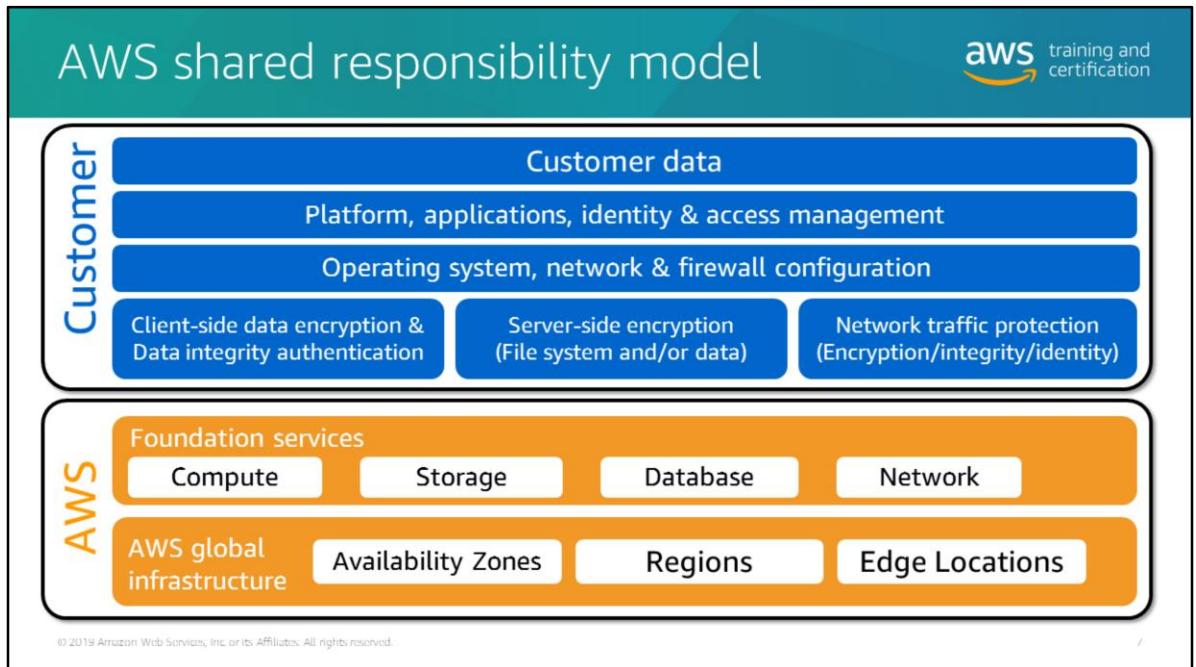
Security in the Cloud

While AWS secures and maintains the cloud infrastructure, customers are responsible for securing everything they put *in* the cloud.

When using AWS services, customers maintain complete control over their content and are responsible for managing critical-content security requirements, including:

- The content they choose to store on AWS
- The AWS services that are used with the content
- The country in which that content is stored
- The format and structure of that content and whether it is masked, anonymized, or encrypted
- Who has access to that content and how those access rights are granted, managed, and revoked

Customers retain control of what security they choose to implement to protect their own data, platform, applications, identity and access management, and operating system. This means that the shared responsibility model changes depending on the AWS services the customer uses.



You have reviewed the AWS shared responsibility model that shows how security and compliance are shared between AWS and the customer.

Discussion: Who's responsible for what?



Unmanaged services

- Amazon EC2
- Amazon EBS

Managed services

- Amazon RDS
- Amazon S3
- Amazon DynamoDB

Operations

- Guest OS patching
- Database patching
- Firewall configuration
- Disaster recovery
- User data

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

8

In addition to protecting this global infrastructure, AWS is responsible for the security configuration of its products that are considered foundational or managed services that include compute, storage, database, and networking.

Infrastructure services, such as Amazon EC2, Amazon EBS, and Amazon VPC, run on top of the AWS global infrastructure. They vary in terms of availability and durability objectives but generally operate within the specific Region where they have been launched. You can build systems that meet availability objectives exceeding those of individual AWS services by employing resilient components in multiple Availability Zones.

For abstracted services, such as Amazon S3 and Amazon DynamoDB, AWS operates the infrastructure layer, the operating system, and platforms. You access the endpoints to store and retrieve data. Amazon S3 and Amazon DynamoDB are tightly integrated with AWS Identity and Access Management (IAM). You are responsible for managing your data (including classifying your assets), and for using IAM tools to apply access control list (ACL)-type permissions to individual resources at the platform level, or permissions based on identity or user responsibility at the IAM user or IAM group level.

Here are a few examples of these types of services :

- Amazon DynamoDB
- Amazon RDS
- Amazon Redshift
- Amazon EMR
- Amazon WorkSpaces

Security, identity, and compliance products



AWS Artifact
AWS Certificate Manager
Amazon Cloud Directory
AWS CloudHSM
Amazon Cognito
AWS Directory Service
AWS Firewall Manager
Amazon GuardDuty
AWS Identity and Access Management

Amazon Inspector
AWS Key Management Service
Amazon Macie
AWS Organizations
AWS Shield
AWS Secrets Manager
AWS Single Sign-On
AWS WAF

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

9

AWS Artifact

The AWS Artifact portal provides on-demand access to AWS' security and compliance documents, also known as audit artifacts.

AWS Certificate Manager

AWS Certificate Manager lets you easily provision, manage, and deploy Secure Sockets Layer/Transport Layer Security (SSL/TLS) certificates.

Amazon Cloud Directory

Amazon Cloud Directory enables you to build flexible cloud-native directories for organizing hierarchies of data along multiple dimensions.

AWS CloudHSM

AWS CloudHSM helps you meet corporate, contractual and regulatory compliance requirements for data security by using dedicated Hardware Security Module (HSM) appliances within the AWS cloud.

Amazon Cognito

Amazon Cognito lets you add user sign-up/sign-in and access control to your web and mobile apps quickly and easily.

AWS Directory Service

AWS Directory Service for Microsoft Active Directory (Enterprise Edition), also known as AWS Microsoft AD, enables your directory-aware workloads and AWS resources to use managed Active Directory in the AWS Cloud.

[AWS Firewall Manager](#)

AWS Firewall Manager is a security management service that makes it easier to centrally configure and manage AWS WAF rules across your accounts and applications.

[Amazon GuardDuty](#)

Amazon GuardDuty is a managed threat detection service that provides you with a more accurate and easy way to continuously monitor and protect your AWS accounts and workloads.

[AWS Identity and Access Management \(IAM\)](#)

Use IAM to control user access to AWS services. Create and manage users and groups, and grant or deny access.

[Amazon Inspector](#)

Amazon Inspector is an automated security assessment service that helps improve the security and compliance of applications deployed on AWS.

[AWS Key Management Service](#)

AWS Key Management Service (AWS KMS) is a managed service that makes it easy for you to create and control the encryption keys used to encrypt your data.

[Amazon Macie](#)

Amazon Macie is a machine learning-powered security service to discover, classify, and protect sensitive data.

[AWS Organizations](#)

AWS Organizations offers policy-based management for multiple AWS accounts. With Organizations, you can create groups of accounts and then apply policies to those groups.

[AWS Shield](#)

AWS Shield is a managed Distributed Denial of Service (DDoS) protection service that safeguards web applications running on AWS.

[AWS Secrets Manager](#)

AWS Secrets Manager enables you to easily rotate, manage, and retrieve database credentials, API keys, and other secrets throughout their lifecycle.

[AWS Single Sign-On](#)

AWS Single Sign-On (SSO) is a cloud SSO service that makes it easy to centrally manage SSO access to multiple AWS accounts and business applications.

[AWS WAF](#)

AWS WAF is a web application firewall that helps protect your web applications from common web exploits that could affect application availability, compromise security, or consume excessive resources.

For the scope of this course, we will expand on AWS Identity and Access Management, Amazon Inspector and AWS Shield.

Manage authentication and authorization

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



AWS Identity and Access Management (IAM)

Securely control access to AWS resources

 IAM user	A person or application that interacts with AWS
 Group	Collection of users with identical permissions
 Role	Temporary privileges that an entity can assume

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

11

IAM enables you to manage access to AWS services and resources securely. Using IAM, you can create and manage AWS users and groups, and use permissions to allow and deny their access to AWS resources.

IAM is a feature of your AWS account that is offered at no additional charge. You are charged only for use of other AWS services by your users.

Using IAM, you can create and manage AWS users and groups, and use permissions to allow and deny their access to AWS resources.

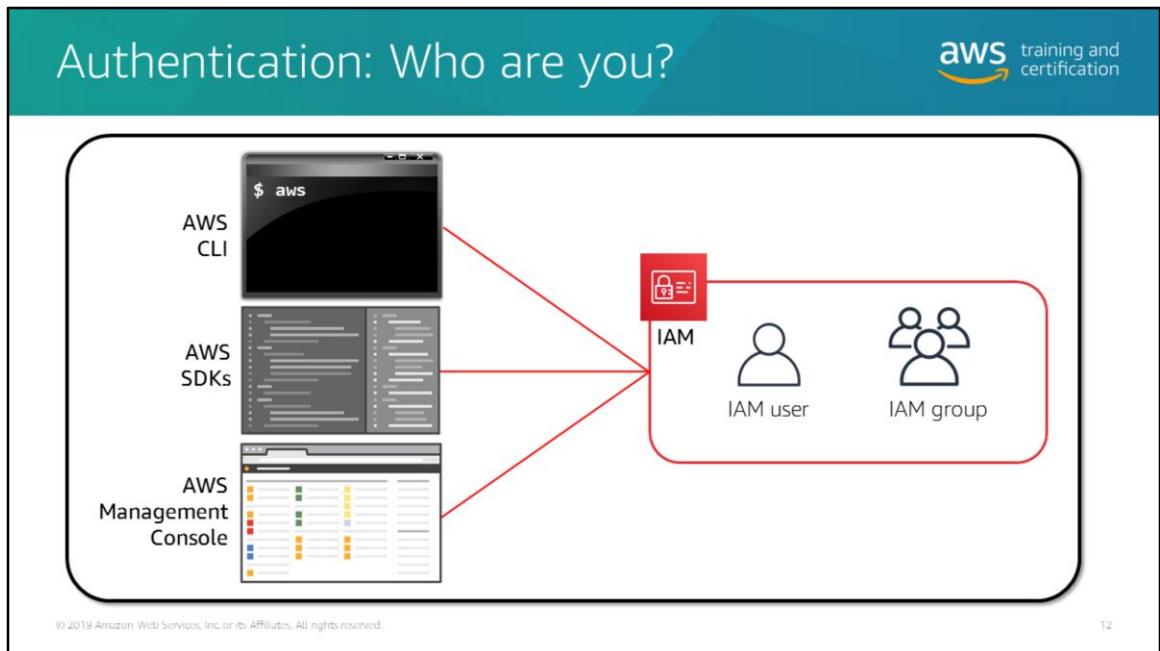
IAM allows you to:

- **Manage IAM users and their access.** You can create users in IAM, assign them individual security credentials (in other words, access keys, passwords, and multi-factor authentication devices), or request temporary security credentials to provide users access to AWS services and resources. You can manage permissions to control which operations a user can perform.

Using groups for easy administration - A group is a collection of IAM users. Use groups to assign permissions to a collection of users, which can make it easier to

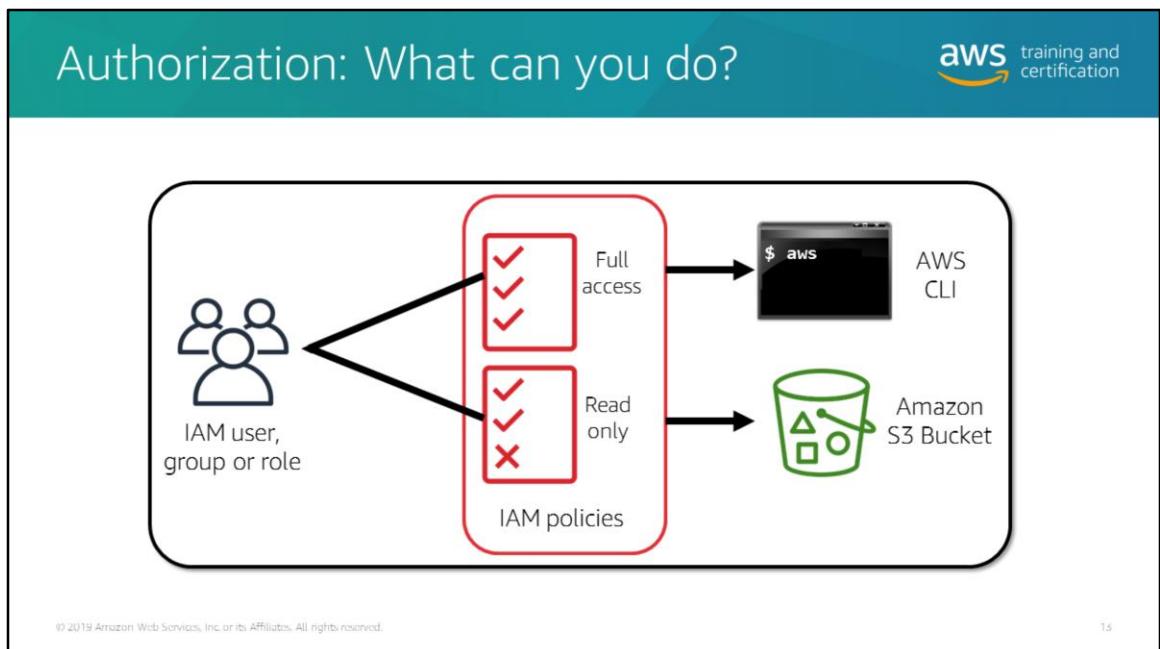
manage the permissions for those users. For example, you could have a group called *Admins* and give that group the types of permissions that administrators typically need. Any user in that group automatically has the permissions that are assigned to the group. If a new user joins your organization and requires administrator permissions, you can assign the appropriate permissions by adding the user to that group. Similarly, if a person changes jobs in your organization, instead of editing that user's permissions, you can remove him or her from the old group and add him or her to the new group.

- **Manage IAM roles and their permissions.** You can create roles in IAM and manage permissions to control which operations can be performed by the entity, or AWS service, that assumes the role.
- **Manage federated users and their permissions.** You can enable identity federation to allow existing identities (users, groups, and roles) from your corporate directory to access the console, call AWS API operations, and access resources, without the need to create an IAM user for each identity.



Authentication is the process of verifying that a user, host, or other resource is who they claim to be, and is usually accomplished with digital certificates signed by a certificate authority (CA). Users must be authenticated before they can access AWS services and resources. Users can access AWS services through the console, AWS CLI, AWS SDKs, or AWS APIs. You can create individual IAM users within your AWS account that correspond to users in your organization. Each user can have his or her own credentials for authenticating to AWS. IAM groups can also be created within your AWS account.

As the number of users managing your AWS environment increases, it is helpful to manage permissions for multiple IAM users by using IAM groups, which ensure that each member of the group will be provisioned with the same permissions policy.



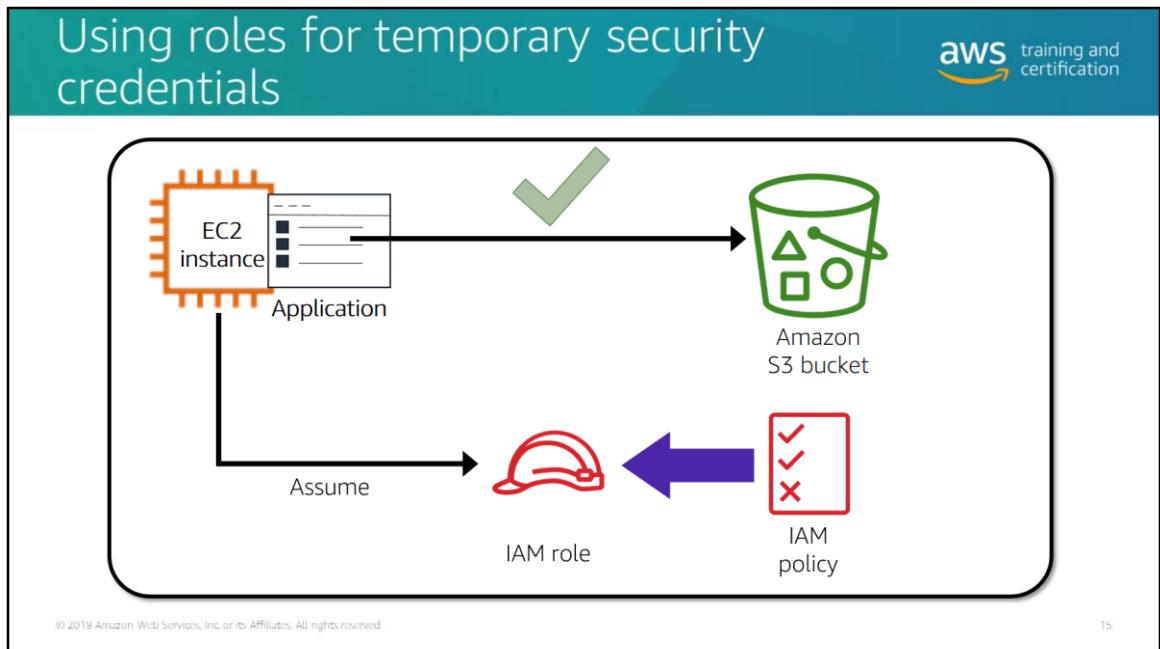
Authorization is the process of determining what permissions a user or other resource has. After a user has been authenticated, they must be authorized to access AWS services. By default, IAM users do not have permissions to access any resources or data in your account. You can grant permissions to a user by creating a *policy*, which is a document in JSON format that explicitly lists permissions to allow or deny access to resources in an AWS account.

IAM roles



- IAM users, applications, and services may assume IAM roles
- Roles uses an IAM policy for permissions

You can also assign IAM policies to an IAM role. An *IAM role* is similar to a user in that it is an AWS identity with permissions that determine what the identity can and cannot do in AWS. A role does not have any long-term defined credentials, such as password or access keys, associated with it. Instead, if a user is assigned to a role, access keys are created dynamically and provided to the user temporarily. Use IAM roles to delegate access to users, applications, or services that don't normally have access to your AWS resources. A user who assumes a role temporarily gives up his or her own permissions and instead takes on the permissions of the role.



15

So how can an IAM role be used in a real-world scenario? Well, suppose that you have a custom application that is hosted on an Amazon EC2 instance. This application needs to interact with objects stored in an Amazon S3 bucket. One way for the application to access the S3 bucket is to embed your AWS credentials in the application code, but doing so may compromise your credentials.

Also, changing or rotating the credentials would require an update in the code each time. The preferred and more secure option is to use an IAM role to pass temporary security credentials as part of an instance profile. The application would use the identity assumed by the instance to access the Amazon S3 bucket.

The screenshot shows the 'Create an AWS account' form on the left and a 'Recommendations' box on the right. The 'Email address' field is highlighted with a red border. The 'Recommendations' box contains five items:

- Delete root user access keys** (key icon)
- Create an IAM user** (person icon)
- Grant administrator access** (key icon)
- Use IAM credentials to interact with AWS** (calendar icon)
- Enable MFA** (MFA icon)

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

When you first create an AWS account, you begin with a single identity that has complete access to all AWS services and resources in the account. This identity is called the *AWS account root user* and is accessed by signing in with the email address and password that you used to create the account.

Because you can't restrict permissions for root user credentials, AWS recommends that you delete your root user access keys. If you require administrator-level permissions, create an IAM user, grant that user full administrator access, and then use those credentials to interact with AWS. If you need to modify or revoke your permissions, you can delete or modify the policies that are associated with that IAM user.

For more information on best practices, see
<https://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html#lock-away-credentials>.

Best practices



- Delete access keys for the AWS account root user
- Activate multi-factor authentication (MFA)
- Only give IAM users permissions they need
- Use roles for applications
- Rotate credentials regularly
- Remove unnecessary users and credentials
- Monitor activity in your AWS account

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

1 /

IAM best practices

- Delete your access keys for the AWS account root user, because they provide unrestricted access to your AWS resources.

Access keys are long-term credentials for an IAM user or the AWS account root user. You can use access keys to sign programmatic requests to the AWS CLI or AWS API (directly or by using the AWS SDK). Access keys consist of two parts: an access key ID (for example, AKIAIOSFODNN7EXAMPLE) and a secret access key (for example, wJalrXUtnFEMI/K7MDENG/bPxRfCYEXAMPLEKEY). Similar to a user name and password, you must use both the access key ID and secret access key together to authenticate your requests. Manage your access keys as securely as you do your user name and password. Instead, use IAM user access keys or temporary security credentials.

- Activate multi-factor authentication (MFA) on your AWS credentials to add another layer of protection to help keep your account secure.
- Create IAM users and give them only the permissions they require. Do not use

your AWS account credentials for day-to-day interaction with AWS, because the AWS account root user provides unrestricted access to your AWS resources.

- Use IAM groups to assign permissions to your IAM users to simplify managing and auditing permissions in your account.
- Apply an IAM password policy to require your IAM users to create strong passwords and to rotate their passwords regularly.
- Use roles for applications that run on Amazon EC2 instances.
- Delegate by using roles instead of by sharing credentials.
- Rotate credentials regularly.
- Remove unnecessary users and credentials.
- Use policy conditions for extra security.
- Monitor activity in your AWS account.

For more information, see <http://docs.aws.amazon.com/IAM/latest/UserGuide/best-practices.html>.

Knowledge check



Your web application requires AWS credentials and authorization to use AWS services. Which IAM entity should be used?

- A. User
- B. Group
- C. Role
- D. MFA

Knowledge check



Your web application requires AWS credentials and authorization to use AWS services. Which IAM entity should be used?

- A. User
- B. Group
- C. Role
- D. MFA

C is correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

19

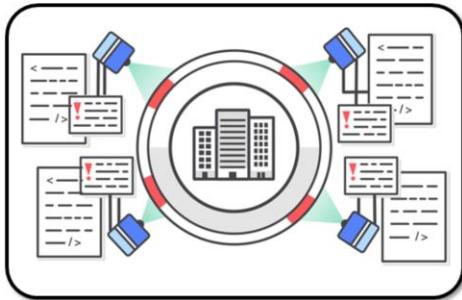
Assess your security and compliance

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



20

Challenges of threat assessment



- Expensive
- Complex
- Time-consuming
- Difficult to track IT changes

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

21

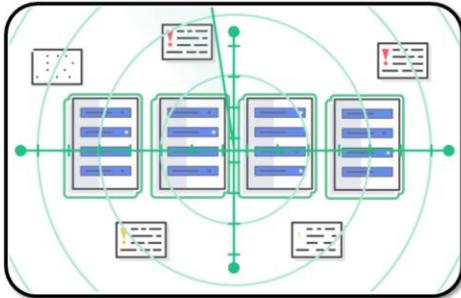
No matter the size of your organization, security matters. The needs of the business and IT may fluctuate, but threat assessment is crucial. But that's not a simple task. Most of the available tools are costly and time-consuming to configure, maintain, and making it difficult to use as regular part of the development lifecycle.

Application testing is key to moving fast, but staying safe at speed can be challenging. Security assessments are highly manual, resulting in delays or missed security checks. Valuable security subject matter experts could be spending a significant amount of their time on routine security assessments.

What is Amazon Inspector?

aws training and certification

Automated security assessment as a service



- Assesses applications for vulnerabilities
- Produces a detailed list of security findings
- Leverages security best practices

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

22

Amazon Inspector is an automated security assessment service that helps improve the security and compliance of applications deployed on AWS. Amazon Inspector automatically assesses applications for vulnerabilities or deviations from best practices. After performing an assessment, Amazon Inspector produces a detailed list of security findings prioritized by level of severity. You can view these findings directly or as part of detailed assessment reports, which are available from the Amazon Inspector console or AWS API.

To help you get started quickly, Amazon Inspector includes a knowledge base of hundreds of rules mapped to common security best practices and vulnerability definitions. Examples of built-in rules include checking whether the remote root login is enabled or whether vulnerable software versions are installed. These rules are regularly updated by AWS security researchers.

Amazon Inspector improves the security posture by allowing you to:

- **Identify application security issues**

Amazon Inspector helps you to identify security vulnerabilities and deviations from security best practices in applications, both before they are deployed and while

they are running in a production environment. This helps improve the overall security posture of your applications deployed on AWS.

- **Integrate security into DevOps**

Amazon Inspector is agent-based, API-driven service. This makes it easy for you to build right into your existing DevOps process, decentralizing and automating vulnerability assessments. This capability also empowers your development and operations teams to make security assessments an integral part of the deployment process.

- **Increase development agility**

Amazon Inspector helps you to reduce the risk of introducing security issues during development and deployment by automating the security assessment of your applications and proactively identifying vulnerabilities. This enables you to develop and iterate on new applications quickly and assess compliance with best practices and policies.

- **Leverage AWS security expertise**

The AWS security organization is continuously assessing the AWS environment and updating a knowledge base of security best practices and rules. Amazon Inspector makes this expertise available to you as a service, which simplifies the process of establishing and enforcing best practices within your AWS environment.

- **Streamline security compliance**

Amazon Inspector gives security teams and auditors visibility into the security testing that is being performed during the development of applications on AWS. This streamlines the process of validating and demonstrating that security, compliance standards, and best practices are being followed throughout the development process.

- **Enforce security standards**

Using Amazon Inspector, you can define standards and best practices for your applications and validate adherence to these standards. This simplifies enforcement of your organization's security standards and best practices. The ability to enforce custom standards helps to proactively manage security issues before they impact your production application.

The screenshot shows the 'Amazon Inspector findings' interface. At the top right is the AWS training and certification logo. Below the header is a section titled 'Amazon Inspector - Findings'. A sub-section header 'Inspector findings are potential security issues discovered during Inspector's assessment of the specified application. Learn more.' is present. Below this is a table with the following columns: Severity, Application, Assessment, Rule package, and Finding. The first row in the table has its 'Severity' column highlighted with a red box. The table contains 24 rows, with the last row showing 'Viewing 1-10 of 24'. At the bottom left is a copyright notice: '© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.' and at the bottom right is the number '25'.

Severity	Application	Assessment	Rule package	Finding
High ⓘ	Customer Processing	Comprehensive-Assessment	Authentication Best Practices	Instance i-aac4c46f is config...
High ⓘ	Customer Processing	Comprehensive-Assessment	Common Vulnerabilities and Ex...	Instance i-aac4c46f is vulne...
High ⓘ	Customer Processing	Comprehensive-Assessment	Authentication Best Practices	No password complexity me...
Informational ⓘ	Customer Processing	Comprehensive-Assessment	Operating System Security Best...	No potential security issues
Informational ⓘ	Customer Processing	Comprehensive-Assessment	Network Security Best Practices	No potential security issues

Findings are potential security issues discovered during the Amazon Inspector assessment of the selected assessment target. Findings contain both a detailed description of the security issues and recommendations for how to solve them. The details of the finding include the following:

- Name of the assessment target that includes the EC2 instance where this finding was registered
- Name of the assessment template that was used to produce this finding
- Assessment run start time, end time, and status
- Name of the rules package that includes the rule that triggered this finding
- Name, severity, and description of the finding
- Remediation steps

For more information, see

http://docs.aws.amazon.com/inspector/latest/userguide/inspector_findings.html.

Remediation recommendation



Finding for application - Customer Processing

Application name	Customer Processing
Assessment name	Comprehensive-Assessment
Rule package	Authentication Best Practices
Finding	Instance i-aac4c46f is configured to allow users to log in with root credentials over SSH. This increases the likelihood of a successful brute-force attack.
Severity	High ⓘ
Description	This rule helps determine whether the SSH daemon is configured to permit logging in to your EC2 instance as root.
Recommendation	It is recommended that you configure your EC2 instance to prevent root logins over SSH. Instead, log in as a non-root user and use sudo to escalate.

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

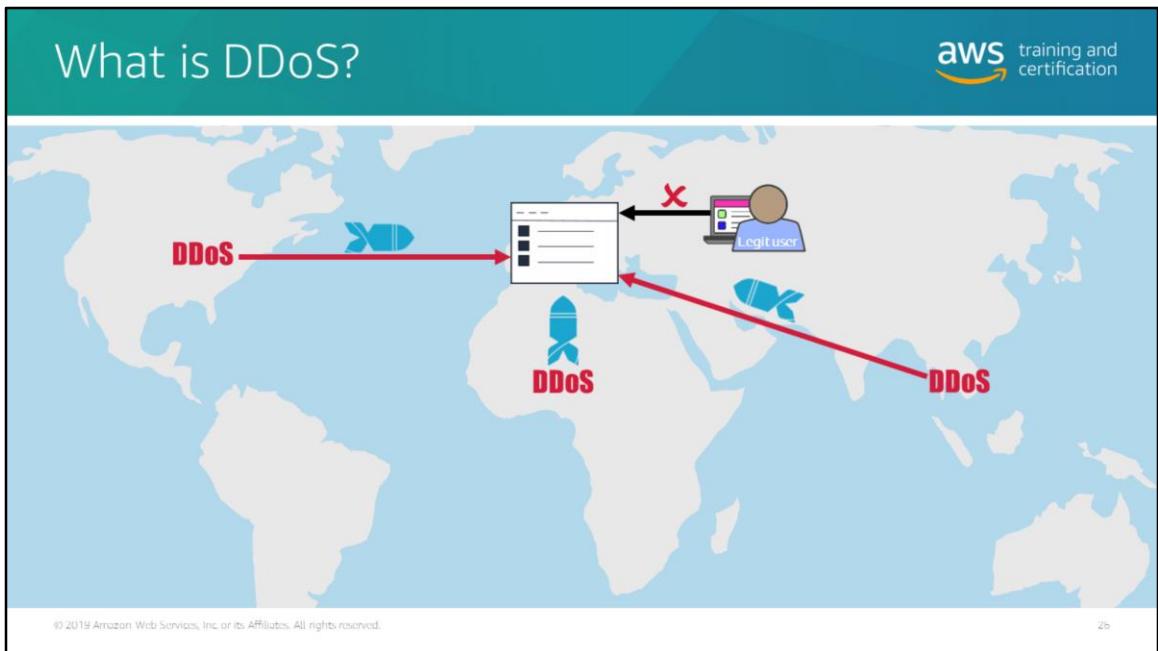
24

This is an example of a detailed recommended remediation step that you can complete to fix the potential security issue described in the finding.

Protect your infrastructure from Distributed Denial of Service (DDoS) attacks

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.





A *Denial of Service (DoS)* attack is a deliberate attempt to make your website or application unavailable to users, such as by flooding it with network traffic. To achieve this, attackers use a variety of techniques that consume large amounts of network bandwidth or tie up other system resources, disrupting access for legitimate users. In its simplest form, a lone attacker uses a single source to execute a DoS attack against a target.

But in a *Distributed Denial of Service (DDoS)* attack, an attacker uses multiple sources—which may be distributed groups of malware-infected computers, routers, IoT devices, and other endpoints—to orchestrate an attack against a target.

DDoS mitigation challenges



- Complex
- Limited bandwidth
- Involves rearchitecting
- Manual
- Degraded performance
- Time-consuming
- Expensive

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

2/

In most cases, DDoS mitigation is difficult to enable. It has the following challenges:

- Setting it up is complex.
- It usually involves re-architecting your applications.
- You can suffer bandwidth limitations if you opt to tackle it from the on-premises data center.

During the attack, operators must manually initiate the mitigation. Mitigation vendors/teams must reroute traffic via a distant scrubbing location. This, in turn, adds more time to resolution and also increases network latency.

For more information on AWS Shield Standard, see
https://aws.amazon.com/shield/features/#AWS_Shield_Standard.

What is AWS Shield?

- A managed DDoS protection service
- Always-on detection and mitigations
- Seamless integration and deployment
- Cost-efficient and customizable protection

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

28

AWS Shield is a managed DDoS protection service that safeguards applications running on AWS. AWS Shield provides always-on detection and automatic inline mitigations that reduce application downtime and latency, so there is no need to engage AWS Support to benefit from DDoS protection.

Seamless integration and deployment

With AWS Shield Standard, your AWS resources are automatically protected from the most common network and transport layer DDoS attacks. You can achieve a higher level of defense by enabling AWS Shield Advanced protection for the Elastic IP address, Elastic Load Balancing, Amazon CloudFront or Amazon Route 53 resources that you want to protect.

Customizable protection

There are two tiers of AWS Shield: Standard and Advanced. With AWS Shield Advanced, you can write customized rules to mitigate sophisticated application-layer attacks. These customizable rules can be deployed instantly, enabling you to quickly mitigate attacks. You can set up rules proactively to automatically block bad traffic or respond to incidents as they occur. You also have 24/7 access to the AWS DDoS Response Team (DRT), who can write rules on your behalf to mitigate application-

layer DDoS attacks.

Cost-efficient

As an AWS customer, you get network layer protection against the most common DDoS attacks with AWS Shield Standard automatically. This protection does not cost you more or require additional resources or time to initiate. With AWS Shield Advanced, you get DDoS Cost Protection, a feature that prevents your AWS bill from incurring Amazon EC2, Elastic Load Balancing, Amazon CloudFront, and Route 53 usage spikes as the result of a DDoS attack.

AWS Shield Standard and AWS Shield Advanced



AWS Shield Standard (included)

- Quick detection
- Inline attack mitigation

AWS Shield Advanced

- Enhanced detection
- Advanced attack mitigation
- Visibility and attack notification
- DDoS cost protection
- Specialized support

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

29

AWS Shield Standard

- **Quick detection**

AWS Shield Standard provides always-on network flow monitoring that inspects incoming traffic to AWS and uses a combination of traffic signatures, anomaly algorithms, and other analysis techniques to detect malicious traffic in real time.

- **Inline attack mitigation**

AWS Shield Standard provides built-in automated mitigation techniques, giving you protection against the most common infrastructure attacks. Automatic mitigations are applied inline to your applications so that there is no impact to latency. AWS Shield Standard uses several techniques, such as deterministic packet filtering and priority-based traffic shaping, to automatically mitigate attacks without impact to your applications. You can also mitigate application-layer DDoS attacks by writing rules using AWS WAF. When you use AWS Shield Standard with CloudFront and Route 53, you receive comprehensive availability protection against all known infrastructure (Layer 3 and 4) attacks.

AWS Shield Advanced

- **Enhanced detection**

With AWS Shield Advanced, you have 24/7 access to the AWS DDoS Response Team (DRT), whom you can engage before, during, or after a DDoS attack. The DRT helps triage the incidents, identify root causes, and apply mitigations on your behalf. You can also engage with the DRT for any post-attack analysis.

- **Advanced attack mitigation**

AWS Shield Advanced provides you with more sophisticated automatic mitigations for attacks targeting your applications running on Amazon EC2, Elastic Load Balancing, CloudFront, and Route 53 resources. Using advanced routing techniques, AWS Shield Advanced provides additional mitigation capacity to protect against larger DDoS attacks. The DRT also applies manual mitigations for more complex and sophisticated DDoS attacks. For application-layer attacks, you can use AWS WAF to respond to incidents. With AWS WAF, you can set up proactive rules like Rate Based Blacklisting to automatically block bad traffic or respond immediately to incidents as they happen. There is no additional charge for using AWS WAF for application-layer protection. You can also engage directly with the DRT to place AWS WAF rules on your behalf, in response to an application layer DDoS attack. The DRT will diagnose the attack and, with your permission, apply mitigations on your behalf.

- **Visibility and attack notification**

You gain complete visibility into DDoS attacks with near-real-time notification through CloudWatch and detailed diagnostics on the AWS WAF and AWS Shield console. Working with the DRT, you can access post-event analysis and investigation. You can also view a summary of prior attacks from the console.

- **DDoS cost protection**

AWS Shield Advanced comes with DDoS cost protection, a safeguard from scaling charges because of a DDoS attack that causes usage spikes on Amazon EC2, Elastic Load Balancing, CloudFront, or Route 53. If any of these services scale up in response to a DDoS attack, AWS will provide AWS Shield service credits for charges as the result of usage spikes. For more details on how to request service credits, see [AWS WAF and AWS Shield Advanced Documentation](#).

- **Specialized support**

AWS Shield Advanced provides enhanced detection, inspecting network flows, and also monitoring application-layer traffic to your Elastic IP address, Elastic Load Balancing, CloudFront, or Route 53 resources. Using additional techniques, such as resource-specific monitoring, AWS Shield Advanced provides granular detection of DDoS attacks and also detects application-layer DDoS attacks, such as HTTP floods or DNS query floods, by baselining traffic on your resource and identifying anomalies.

- **Global availability**

AWS Shield Advanced is available globally on all CloudFront and Route 53 edge locations. You can protect your web applications hosted anywhere in the world by deploying CloudFront in front of your application. Your origin servers can be Amazon S3, Amazon EC2, Elastic Load Balancing, or a custom server outside of AWS. You can also enable AWS Shield Advanced directly on an Elastic IP address or Elastic Load Balancing load balancer in these AWS Regions: Northern Virginia, Oregon, Ireland, Tokyo, and Northern California.

AWS security compliance

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Assurance programs

The grid displays the following logos:

- Global:** CSA cloud security alliance, ISO 9001, ISO 27001, ISO 27017, ISO 27018, PCI Security Standard Council (Participating Organization), AICPA SOC, AICPA SOC, AICPA SOC.
- USA:** DHS, Department of Defense, FedRAMP, Department of Education, FFIEC, FIPS 140-2 Validated, FISMA Moderate, HIPAA Business Associate Agreements.
- Europe:** ITAR, NIST, SECURITÉ ET EXCHANGE COMMISSION, FIEC, C5, CYBER ESSENTIALS PLUS, CERTIFICACIÓN DE CONFIABILIDAD eNS.
- Asia Pacific:** irap, ISMS SINGAPORE, iDA SINGAPORE, Rabbit logo.

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

51

AWS has successfully completed multiple audits, attestations, and certifications. AWS publishes a Service Organization Controls SOC 1 report, and publishes under both the SSAE 16 and the ISAE 3402 professional standards as SOC 2-Security and SOC 3 report.

In addition, AWS has achieved ISO 9001, ISO 27001, ISO 27017, and ISO 27018 certifications. It has also been successfully validated as a Level 1 service provider under the Payment Card Industry (PCI) Data Security Standard (DSS), and currently offers HIPAA Business Associate Agreements to covered entities and their business associates subject to HIPAA.

In the realm of public sector certifications, AWS has achieved FedRAMP compliance, has received authorization from the US General Services Administration to operate at the FISMA Moderate level. AWS is also the platform for applications with Authorities to Operate (ATO) under the Defense Information Assurance Certification and Accreditation Program (DIACAP).

NIST, FIPS 140-2, CJIS, and DoD SRG Levels 2 and 4 are some of the other certifications AWS has received.

For more information on AWS compliance, see <http://aws.amazon.com/compliance/>.

How AWS helps customers achieve compliance



Sharing information

- Industry certifications
- Security and control practices
- Compliance reports directly under NDA

Assurance program

- Certifications/attestations
- Laws, regulations, and privacy
- Alignments/frameworks

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

52

Though customers don't communicate their use and configurations to AWS, AWS does communicate its security and control environment relevant to customers. AWS achieves this by doing the following:

- Obtaining industry certifications and independent third-party attestations
- Publishing information about the AWS security and control practices in whitepapers and website content
- Providing certificates, reports, and other documentation directly to AWS customers under NDA (as required)

Many of the compliance documents are available to the customer from the console through the AWS Artifact service.

Assurance programs: AWS engages with external certifying bodies and independent auditors to provide customers with considerable information regarding the policies, processes, and controls that AWS has established and operated.

Certifications/attestations: Compliance certifications and attestations are assessed by an independent third-party auditor and result in a certification, audit report, or attestation of compliance.

Laws, regulation, and privacy: AWS customers remain responsible for complying with applicable compliance laws and regulations. In some cases, AWS offers functionality (such as security features), enablers, and legal agreements (such as the AWS Data Processing Agreement and Business Associate Addendum) to support customer compliance.

Alignments/frameworks: Compliance alignments and frameworks include published security or compliance requirements for a specific purpose, such as a specific industry or function. AWS provides functionality (such as security features) and enablers (including compliance playbooks, mapping documents, and whitepapers) for these types of programs.

Customer responsibility



You own your certification.

Review – Design – Identify – Verify

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

55

AWS customers are required to maintain adequate governance over the entire IT control environment, regardless of how IT is deployed. Leading practices include an understanding of required compliance objectives and requirements (from relevant sources), establishment of a control environment that meets those objectives and requirements, an understanding of the validation required based on the organization's risk tolerance, and verification of the operating effectiveness of their control environment. Deployment in the AWS Cloud gives enterprises different options to apply various types of controls and various verification methods.

Strong customer compliance and governance might include the following basic approaches:

- **Review** information available from AWS together with other information to understand as much of the entire IT environment as possible, and then document all compliance requirements.
- **Design** and implement control objectives to meet the enterprise compliance requirements.
- **Identify** and document controls owned by outside parties.
- **Verify** that all control objectives are met and all key controls are designed and operating effectively.

By staying engaged in the compliance and governance process with AWS, customers can design to compliance requirements.

Knowledge check



Which of the following are best practices for security? (Select all that apply)

- A. Delete root user access keys
- ~~B. Use the same password for all users~~
- C. Use roles for applications
- D. Embed secrets in your code
- E. Activate multi-factor authentication (MFA)

Knowledge check



Which of the following are best practices for security? (Select all that apply)

- A. Delete root user access keys
- ~~B. Use the same password for all users~~
- C. Use roles for applications
- ~~D. Embed secrets in your code~~
- E. Activate multi-factor authentication (MFA)

A, C, E are correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

55

Key takeaways



- Security is the highest priority at AWS.
- The Shared Responsibility Model defines security responsibilities between AWS and the customer.
- IAM controls access to AWS services and resources securely.
- Amazon Inspector assesses the security of your AWS resources.
- AWS Shield protects applications running on AWS against DDoS attacks.
- AWS security assurance programs help customers maintain security and data compliance.

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

5b

For more information about security, see <https://aws.amazon.com/security>.

Module 5: Pricing models and cloud application support



Module goals



- Fundamentals of pricing
- Getting help with AWS
 - Plans
 - Technology
 - Programs

Fundamentals of pricing

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



How do you pay for AWS?



Pay as you go



Save when you reserve



Pay less by using more



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

4

AWS offers you a pay-as-you-go approach for pricing on more than 140 cloud services. With AWS, you pay only for the individual services you need, for as long as you use them, and without requiring long-term contracts or complex licensing. AWS pricing is similar to how you pay for utilities like water and electricity. You only pay for the services you consume, and once you stop using them, there are no additional costs or termination fees.

Pay as you go

Only pay for what you use

On premises/colocation

A diagram showing a red dollar sign icon connected by a red line to three stacks of coins, representing a fixed cost model.

AWS

A diagram showing an orange dollar sign icon connected by an orange line to three stacks of coins, representing a variable cost model.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

With AWS, you only pay for what use, helping your organization remain agile, responsive and always able to meet scale demands.

Pay-as-you-go pricing allows you to adapt to changing business needs without overcommitting budgets, which helps you improve your responsiveness to change. With a pay-as-you-go model, you can adapt your business depending on need rather than forecasts, reducing the risk of overprovisioning or missing capacity.

By paying for services on an as-needed basis, you can redirect your focus to innovation and invention, reducing procurement complexity and enabling your business to be fully elastic.

Save when you reserve: Reserved Instances



- Save up to 75 percent over equivalent on-demand capacity
- Choose
 - No upfront payments (NURI)
 - partial up-front (PURI)
 - all up-front (AURI)



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

b

For certain services like Amazon EC2 and Amazon RDS, you can invest in reserved capacity. With Reserved Instances, you can save up to 75 percent over equivalent on-demand capacity. Reserved Instances are available in three options:

- All up-front (AURI)
- Partial up-front (PURI)
- No upfront payments (NURI)

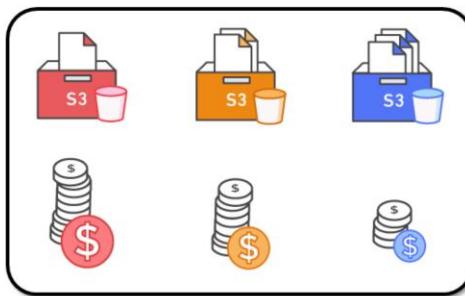
When you buy Reserved Instances, the larger the upfront payment, the greater the discount. To maximize your savings, you can pay all up-front and receive the largest discount. Partial up-front RIs offer lower discounts but give you the option to spend less up front. Lastly, you can choose to spend nothing up front and receive a smaller discount, but allowing you to free up capital to spend in other projects.

By using reserved capacity, your organization can minimize risks, more predictably manage budgets, and comply with policies that require longer-term commitments.

Use more, pay less



Automatic volume-based discounts



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

7

With AWS, you can get volume-based discounts and realize important savings as your usage increases. For services such as Amazon S3 and data transfer OUT from Amazon EC2, pricing is *tiered*, meaning the more you use, the less you pay per GB. In addition, data transfer IN is usually free of charge (there can be some exceptions). As a result, as your AWS usage needs increase, you benefit from the economies of scale that allow you to increase adoption and keep costs under control.

As your organization evolves, AWS also gives you options to acquire services that help you address your business needs. For example, the AWS storage services portfolio offers options to help you lower pricing based on how frequently you access data, and the performance you need to retrieve it. To optimize your savings, choose the right combinations of storage solutions that help you reduce costs while preserving performance, security and durability.

Pricing concepts



Compute

- Charged per hour/second*
- Varies by instance type

*Linux only

Storage

- Charged typically per GB

Data transfer

- Outbound is aggregated and charged
- Inbound has no charge (with some exceptions)
- Charged typically per GB

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

8

There are three fundamental drivers of cost with AWS: compute, storage, and outbound data transfer. These characteristics vary somewhat, depending on the AWS product and pricing model you choose.

In most cases, there is no charge for inbound data transfer or for data transfer between other AWS services within the same region. There are some exceptions, so be sure to verify data transfer rates before beginning. Outbound data transfer is aggregated across services and then charged at the outbound data transfer rate. This charge appears on the monthly statement as *AWS Data Transfer Out*.

The more data you transfer, the less you pay per GB. For compute resources, you pay hourly from the time you launch a resource until the time you terminate it, unless you have made a reservation for which the cost is agreed upon beforehand. For data storage and transfer, you typically pay per GB.

Different services are priced differently



© 2018, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

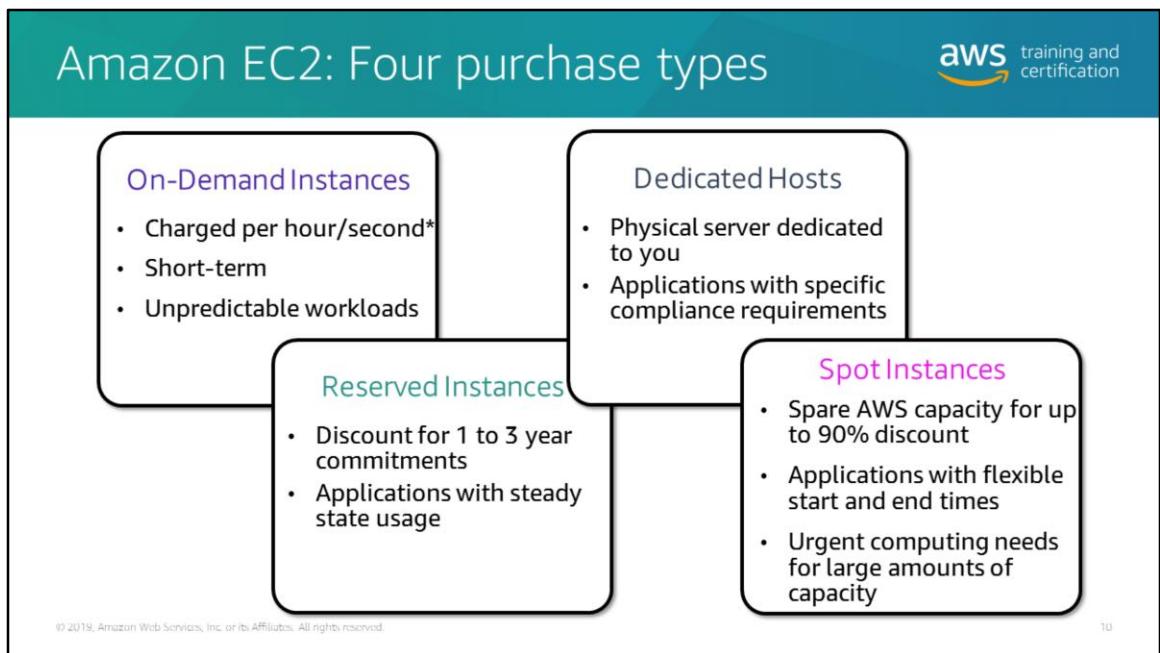
9

-  Amazon EC2
-  Amazon EBS
-  Amazon S3
-  AWS CloudFormation

Different types of services lend themselves to different pricing models. For example, Amazon EC2 pricing varies by instance type, while the Amazon Aurora database service includes charges for data input/output (I/O) and storage. This section provides an overview of pricing concepts and examples for a range of commonly used services. You can always find current price information for each AWS service at <http://aws.amazon.com/pricing>.

Let's break down the pricing characteristics for some commonly used AWS products.

For details on other services, see <https://aws.amazon.com/pricing/services/>



On-Demand

With On-Demand instances, you pay for compute capacity by per hour or per second depending on which instances you run. No longer-term commitments or upfront payments are needed. You can increase or decrease your compute capacity depending on the demands of your application and only pay the specified per hourly rates for the instance you use.

Paying by second pertains to Linux only.

Recommended for:

- Users that prefer the low cost and flexibility of Amazon EC2 without any up-front payment or long-term commitment
- Applications with short-term, spiky, or unpredictable workloads that cannot be interrupted
- Applications being developed or tested on Amazon EC2 for the first time

Reserved Instances

Reserved Instances provide you with a significant discount (up to 75 percent) compared to On-Demand instance pricing. In addition, when Reserved Instances are

assigned to a specific Availability Zone, they provide a capacity reservation, giving you additional confidence in your ability to launch instances when you need them.

For applications that have steady state or predictable usage, Reserved Instances can provide significant savings compared to using On-Demand instances.

Recommended for:

- Applications with steady state usage
- Applications that may require reserved capacity
- Customers that can commit to using Amazon EC2 over a 1- or 3-year term to reduce their total computing costs

Spot Instances

Amazon EC2 Spot instances allow you to request spare Amazon EC2 computing capacity for up to 90 percent off the On-Demand price.

Spot instances are also available to run for a predefined duration – in hourly increments up to six hours in length – at a discount of up to 30-50% compared to On-Demand pricing. Spot instances have low and predictable prices, offers the advantages of the massive operating scale of AWS and is easy to launch and manage.

Recommended for:

- Applications that have flexible start and end times
- Applications that are only feasible at very low compute prices
- Users with urgent computing needs for large amounts of additional capacity

Dedicated Hosts

A Dedicated Host is a physical Amazon EC2 server dedicated for your use. Dedicated Hosts can help you reduce costs by allowing you to use your existing server-bound software licenses, including Windows Server, SQL Server, and SUSE Linux Enterprise Server (subject to your license terms), and can also help you meet compliance requirements.

- Can be purchased On-Demand (hourly).
- Can be purchased as a Reservation for up to 70 percent off the On-Demand price.

Amazon EC2 pricing factors

When you begin to estimate the cost of using Amazon EC2, consider the following:

- **Clock hours of server time:** Resources incur charges when they are running—for example, from the time Amazon EC2 instances are launched until they are terminated, or from the time Elastic IPs are allocated until the time they are de-allocated.
- **Instance type:** Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instance types comprise varying combinations of CPU, memory, storage, and networking capacity and give you the flexibility to choose the appropriate mix of resources for your applications. Each instance type includes at least one instance size, allowing you to scale your resources to the requirements

of your target workload.

- **Pricing model:** With On-Demand Instances, you pay for compute capacity by the hour with no required minimum commitments. Reserved Instances give you the option to make a low one-time payment—or no payment at all—for each instance you want to reserve and in turn receive a significant discount on the hourly usage charge for that instance. With Spot Instances, you can bid for unused Amazon EC2 capacity.
- **Number of instances:** You can provision multiple instances of your Amazon EC2 and Amazon EBS resources to handle peak loads.
- **Load balancing:** An Elastic Load Balancer can be used to distribute traffic among Amazon EC2 Instances. The number of hours the Elastic Load Balancer runs and the amount of data it processes contribute to the monthly cost.
- **Detailed monitoring:** You can use CloudWatch to monitor your EC2 instances. By default, basic monitoring is enabled. For a fixed monthly rate, you can opt for detailed monitoring, which includes seven preselected metrics recorded once a minute. Partial months are charged on an hourly pro rata basis, at a per instance-hour rate.
- **Auto Scaling:** Auto Scaling automatically adjusts the number of Amazon EC2 instances in your deployment according to conditions you define. This service is available at no additional charge beyond CloudWatch fees.
- **Elastic IP addresses:** You can have one Elastic IP (EIP) address associated with a running instance at no charge.
- **Operating systems and software packages:** Operating system prices are included in instance prices, unless you choose to bring your own licenses. There are no additional licensing costs to run the following commercial operating systems: Red Hat Enterprise Linux, SUSE Enterprise Linux, Windows Server, and Oracle Enterprise Linux. Additionally, AWS has made it easy for you by partnering with Microsoft, IBM, and several other vendors so you can run commercial software packages, such as Microsoft SQL Server, on your Amazon EC2 instances. For commercial software packages AWS doesn't provide—such as nonstandard operating systems, Oracle Applications, Windows Server applications such as Microsoft SharePoint and Microsoft Exchange—you need to obtain a license from the vendors. You can also bring your existing license to the cloud through specific vendor programs such as Microsoft License Mobility Through Software Assurance Program.

Amazon EBS pricing model



Volumes

- Charged by GB provisioned/month
- Varies by volume type

Snapshots

- Charged by space consumed in Amazon S3
- Charged for volume copied across regions

Data transfer

- Inbound data transfer is free
- Outbound data transfer charges are tiered

Amazon EBS pricing includes three factors:

- **Volumes:** Volume storage for all EBS volume types is charged by the amount of GB you provision per month, until you release the storage.
- **Snapshots:** Snapshot storage is based on the amount of space your data consumes in Amazon S3. Because Amazon EBS does not save empty blocks, it is likely that the snapshot size will be considerably less than your volume size. Copying EBS snapshots is charged based on the volume of data transferred across regions. For the first snapshot of a volume, Amazon EBS saves a full copy of your data to Amazon S3. For each incremental snapshot, only the changed part of your Amazon EBS volume is saved. After the snapshot is copied, standard EBS snapshot charges apply for storage in the destination region.
- **Data transfer:** Consider the amount of data transferred out of your application. Inbound data transfer is free, and outbound data transfer charges are tiered.

Amazon S3 pricing model



- Amount of storage used
- Region
- Storage class
- Number and type of requests (GET, PUT, COPY)
- Amount of data transferred out of the region

Estimating Amazon S3 storage costs

With Amazon S3, you pay only for the storage you use, with no minimum fee. Prices are based on the location of your Amazon S3 bucket.

When you begin to estimate the cost of Amazon S3, consider the following:

- **Storage class:**
 - S3 Standard Storage is designed to provide 99.99999999 percent durability and 99.99 percent availability.
 - S3 Standard – Infrequent Access (S-IA) is a storage option within Amazon S3 that you can use to reduce your costs by storing less frequently accessed data at slightly lower levels of redundancy than the standard Amazon S3 storage.
 - S3 Intelligent-Tiering
 - S3 One Zone-Infrequent Access
 - You can also use Amazon S3 Glacier storage for archiving data at very low costs.
 - Amazon S3 Glacier Deep Archive
- **Storage:** Costs vary with number and size of objects stored in your Amazon S3 buckets as well as type of storage.
- **Requests:** The number and type of requests. GET requests incur charges at

different rates than other requests, such as PUT and COPY requests.

- **Data transfer:** The amount of data transferred out of the Amazon S3 region.

For more information on pricing, see

<https://aws.amazon.com/s3/pricing/?nc=sn&loc=4>.

AWS services with no additional charge



The slide features a teal header bar with the title "AWS services with no additional charge". Below the header is a white content area enclosed in a rounded rectangle. Inside this area, five AWS services are listed with their corresponding icons:

- Amazon VPC (purple cloud icon)
- Elastic Beanstalk (orange cloud icon)
- Auto Scaling (orange double-headed arrow icon)
- AWS CloudFormation (pink cloud icon)
- AWS Identity and Access Management (IAM) (red lock icon)

At the bottom left of the content area, there is a small copyright notice: "© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved." At the bottom right, there is a small number "15".

Although the service itself has no additional cost, the resources provisioned in or by them incur costs.

Cost estimating tools

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



14

AWS Free Tier

Enables you to gain free hands-on experience with the AWS platform, products, and services.

-  Sign up for an AWS account
-  Learn with 10-minute tutorials
-  Start building with AWS

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

15

The AWS Free Tier enables you to gain free, hands-on experience with the AWS platform, products, and services.

12-Months Free: These free tier offers are only available to new AWS customers, and are available for 12 months following your AWS sign-up date. When your 12 month free usage term expires or if your application use exceeds the tiers, you simply pay standard, pay-as-you-go service rates (see each service page for full pricing details). Restrictions apply; see offer terms for more details.

Always Free: These free tier offers do not automatically expire at the end of your 12 month AWS Free Tier term, but are available to both existing and new AWS customers indefinitely.

Trials: These free tier offers are short term trial offers that start from the time of first usage begins. Once the trial period expires you simply pay standard, pay-as-you-go service rates (see each service page for full pricing details).

The Amazon AWS Free Tier applies to participating services across our global regions. Your free usage under the AWS Free Tier is calculated each month across all regions and automatically applied to your bill – free usage does not accumulate. The AWS Free Tier is not available in the China (Beijing or Ningxia) region at this time.

To get started:

1. **Sign up for an AWS account.** Creating an AWS account is free and gives you immediate access to the AWS Free Tier.
2. **Take some tutorials.** Explore and learn with easy-to-follow tutorials for multiple use cases.
3. **Start building.** Build your production solution quickly and easily once you're ready.

For more information, see <https://aws.amazon.com/free>.

AWS Simple Monthly Calculator



- Estimate your monthly bill
- Per-service cost breakdown
- Aggregate monthly estimate
- Provides common customer examples

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

1b

You can use the AWS Simple Monthly Calculator to estimate your monthly bill. The calculator provides a per-service cost breakdown, as well as an aggregate monthly estimate. You can also use the calculator to see an estimation and breakdown of costs for common solutions.

Analyzing with AWS Cost Explorer



The slide features six icons with corresponding text descriptions:

-  Get started quickly
-  Set custom intervals
-  Filter/group data
-  Forecast cost and usage
-  Save progress
-  Access data programmatically

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

AWS Cost Explorer is a tool that enables you to view and analyze your costs and usage. You can explore your usage and costs using the main graph, the Cost Explorer cost and usage reports, or the Cost Explorer RI reports. You can view data for up to the last 13 months, forecast how much you're likely to spend for the next three months, and get recommendations for what Reserved Instances to purchase. You can use Cost Explorer to identify areas that need further inquiry and see trends that you can use to understand your costs.

Get started quickly: A set of default reports are included to help you quickly gain insight into your cost drivers and usage trends.

Set custom intervals: Set a custom time period, and determine whether you would like to view your data at a monthly or daily level of granularity.

Filter/group data: Dig deeper into your data by taking advantage of filtering and grouping functionality, using a variety of available dimensions.

Forecast cost and usage: Use forecasting to get a better idea of what your costs and usage may look like in the future, so that you can plan ahead.

Save progress: Once you arrive at a helpful view, save your progress as a new report that you can refer back to in the future.

Access data programmatically: Directly access the interactive, ad-hoc analytics engine that powers AWS Cost Explorer.

What Is Trusted Advisor?

A service providing guidance to help you reduce cost, increase performance, and improve security

Cost Optimization	Performance	Security	Fault Tolerance	Service Limits
				
0 ✓ 9 ▲ 0 !	3 ✓ 7 ▲ 0 !	2 ✓ 4 ▲ 11 !	0 ✓ 15 ▲ 5 !	37 ✓ 0 ▲ 1 !
\$7,516.87 Potential monthly savings				

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

18

AWS Trusted Advisor provides best practices (or checks) in five categories: cost optimization, security, fault tolerance, performance, and service limits. The status of the check is shown by using color coding on the dashboard page:

- Red: action recommended
- Yellow: investigation recommended
- Green: no problem detected

For each check, you can review a detailed description of the recommended best practice, a set of alert criteria, guidelines for action, and a list of useful resources on the topic.

Cost Optimization

See how you can save money on AWS by eliminating unused and idle resources or making commitments to reserved capacity.

Performance

Improve the performance of your service by checking your service limits, ensuring you take advantage of provisioned throughput, and monitoring for over-utilized instances.

Security

Improve the security of your application by closing gaps, enabling various AWS security features, and examining your permissions.

Fault Tolerance

Increase the availability and redundancy of your AWS application by take advantage of automatic scaling, health checks, multiple Availability Zones, and backup capabilities.

Service Limits

Checks for service usage that is more than 80 percent of the service limit. Values are based on a snapshot, so your current usage might differ. Limit and usage data can take up to 24 hours to reflect any changes.

Knowledge check



Now that you have started your migration to the cloud, you want to find out which service you use the most and where the majority of your traffic is coming from. Which tool should you use?

- A. AWS Free Tier
- B. AWS Cost Explorer
- C. AWS Simple Monthly Calculator
- D. AWS Annual Calculator

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

19

AWS Cost Explorer is the correct answer. It can be used to find patterns in how much is spent on certain AWS resources over time, identify areas that need further inquiry and see trends to assist in understanding your costs. It can also reveal which service is being used the most and which Availability Zone gets the most network traffic. Here are some of the various filters in Cost Explorer:

- API operation
- Availability Zone
- AWS Cloud service
- Custom cost allocation tags
- Amazon EC2 instance type
- Linked account(s)
- Platform
- Purchase option
- Region
- Tenancy
- Usage type
- Usage type group

Knowledge check



Now that you have started your migration to the cloud, you want to find out which service you use the most and where the majority of your traffic is coming from. Which tool should you use?

- A. AWS Free Tier
- B. AWS Cost Explorer
- C. AWS Simple Monthly Calculator
- D. AWS Annual Calculator

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

20

AWS Cost Explorer is the correct answer. It can be used to find patterns in how much is spent on certain AWS resources over time, identify areas that need further inquiry and see trends to assist in understanding your costs. It can also reveal which service is being used the most and which Availability Zone gets the most network traffic. Here are some of the various filters in Cost Explorer:

- API operation
- Availability Zone
- AWS Cloud service
- Custom cost allocation tags
- Amazon EC2 instance type
- Linked account(s)
- Platform
- Purchase option
- Region
- Tenancy
- Usage type
- Usage type group

Knowledge check



Now that you have started your migration to the cloud, you want to find out which service you use the most and where the majority of your traffic is coming from. Which tool should you use?

- A. ~~AWS Free Tier~~
- B. AWS Cost Explorer
- C. AWS Simple Monthly Calculator
- D. ~~AWS Annual Calculator~~

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

21

AWS Cost Explorer is the correct answer. It can be used to find patterns in how much is spent on certain AWS resources over time, identify areas that need further inquiry and see trends to assist in understanding your costs. It can also reveal which service is being used the most and which Availability Zone gets the most network traffic. Here are some of the various filters in Cost Explorer:

- API operation
- Availability Zone
- AWS Cloud service
- Custom cost allocation tags
- Amazon EC2 instance type
- Linked account(s)
- Platform
- Purchase option
- Region
- Tenancy
- Usage type
- Usage type group

Knowledge check



Now that you have started your migration to the cloud, you want to find out which service you use the most and where the majority of your traffic is coming from. Which tool should you use?

- A. AWS Free Tier
- B. AWS Cost Explorer
- C. AWS Simple Monthly Calculator
- D. AWS Annual Calculator

B is correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

22

AWS Cost Explorer is the correct answer. It can be used to find patterns in how much is spent on certain AWS resources over time, identify areas that need further inquiry and see trends to assist in understanding your costs. It can also reveal which service is being used the most and which Availability Zone gets the most network traffic. Here are some of the various filters in Cost Explorer:

- API operation
- Availability Zone
- AWS Cloud service
- Custom cost allocation tags
- Amazon EC2 instance type
- Linked account(s)
- Platform
- Purchase option
- Region
- Tenancy
- Usage type
- Usage type group

AWS Support

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



25

AWS Support brings Amazon's tradition of customer-obsession to the B2B technology world. We focus on helping you achieve the outcomes you need to make your business successful.

At AWS, Support goes beyond break-fix and issue resolution. AWS Support provides a mix of tools and technology, people, and programs designed to proactively help you optimize performance, lower costs, and innovate faster. We save time for your team by helping you to move faster in the cloud and focus on your core business.

Support plan overview



© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Support plan	Features
Basic	<ul style="list-style-type: none"> Customer service Support forums Service health checks Documentation, whitepapers, and best-practice guides
Developer	<ul style="list-style-type: none"> Best-practice guidance Client-side diagnostic tools Building-block architecture support
Business	<ul style="list-style-type: none"> Use-case guidance IAM for controlling individuals' access to AWS Support Full AWS Trusted Advisor An API for interacting with Support Center and Trusted Advisor Third-party software support
Enterprise	<ul style="list-style-type: none"> Application architecture guidance Infrastructure event management Technical Account Manager (TAM) White-glove case routing Management business reviews

24

Features of AWS Support Plans

AWS Support offers four support plans: Basic, Developer, Business, and Enterprise. The Basic plan is free of charge and offers support for account and billing questions and service limit increases. The other plans offer an unlimited number of technical support cases with pay-by-the-month pricing and no long-term contracts, providing the level of support that meets your needs.

All AWS customers automatically have around-the-clock access to these features of the Basic support plan:

- Customer Service: one-on-one responses to account and billing questions
- Support forums
- Service health checks
- Documentation, whitepapers, and best-practice guides

Customers with a Developer support plan have access to these additional features:

- Best-practice guidance
- Client-side diagnostic tools
- Building-block architecture support: guidance on how to use AWS products, features, and services together

In addition, customers with a Business or Enterprise support plan have access to these features:

- Use-case guidance: what AWS products, features, and services to use to best support your specific needs
- AWS Identity and Access Management (IAM) for controlling individuals' access to AWS Support
- AWS Trusted Advisor, which inspects customer environments and identifies opportunities to save money, close security gaps, and improve system reliability and performance
- An API for interacting with Support Center and Trusted Advisor, allowing for automated support case management and Trusted Advisor operations
- Third-party software support: help with Amazon EC2 instance operating systems and configuration and performance of the most popular third-party software components on AWS

In addition, customers with an Enterprise support plan have access to these features:

- Application architecture guidance: consultative partnership supporting specific use cases and applications
- Infrastructure event management: short-term engagement with AWS Support to get a deep understanding of your use case and provide architectural and scaling guidance for an event
- Technical account manager
- White-glove case routing
- Management business reviews

For more information, see <https://aws.amazon.com/premiumsupport/compare-plans/>.

Support Plan Overview



Support Plan	Features
Basic	<ul style="list-style-type: none">Customer serviceSupport forumsService health checksDocumentation, whitepapers, and best-practice guides
Developer	<ul style="list-style-type: none">Best-practice guidanceClient-side diagnostic toolsBuilding-block architecture support
Business	<ul style="list-style-type: none">Use-case guidanceIAM for controlling individuals' access to AWS SupportFull AWS Trusted AdvisorAn API for interacting with Support Center and Trusted AdvisorThird-party software support
Enterprise	<ul style="list-style-type: none">Application architecture guidanceInfrastructure event managementTechnical Account Manager (TAM)White-glove case routingManagement business reviews

**Business hours are generally defined as 8:00 AM to 6:00 PM in the customer country as set in [My Account console](#), excluding holidays and weekends. These times may vary in countries with multiple time zones.

Support documentation



- Knowledge Center (FAQs and common requests)
- AWS Documentation
- AWS Discussion Forums
- AWS Support Center

In addition to the support plans provided with tailored customer support, we provide other areas of guidance and assistance for you. The Knowledge Center is a central location where some of the most frequent questions and requests are housed. AWS Documentation is where you can find user guides, developer guides, API references, and tutorials regarding certain services. AWS Discussion Forums are forums that allow you and other users to talk amongst each other regarding your thoughts, ideas, knowledge and opinions in a central environment. AWS Support Center is where you can locate your support cases, your current support plan (which you can change), recommended videos, health events, and other helpful resources such as the ones listed above.

Key takeaways



- AWS offers a pay-as-you-go approach for pricing
- Some services have specific pricing factors and some have no additional charge*
- AWS Simple Monthly Calculator helps you estimate your monthly bill
- AWS Support offers plans to fit customers' unique needs
- Additional support such as Knowledge Center, AWS Documentation, and AWS Discussion Forums

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

2 /

Note although the service itself has no additional cost, the resources provisioned in or by them incur costs.

Lab 3: Auditing Security with AWS Trusted Advisor



- Use AWS Trusted Advisor to perform a basic audit of your AWS resources
- Modify Amazon EC2 security groups to meet best practices
- Configure Multi-Factor Authentication (MFA) (Optional; requires installation of software on a mobile device)
- Set up AWS Trusted Advisor email notifications

Module 6: Architecture

© 2019 Amazon Web Services, Inc. or its affiliates. All rights reserved.



Module goals



- The AWS Well-Architected Framework
- Reference architectures
- The future of the cloud

The AWS Well-Architected Framework

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



What is the Well-Architected Framework?



- A guide for designing infrastructures that are:
 - ✓ Secure
 - ✓ High-performing
 - ✓ Resilient
 - ✓ Efficient
- A systematic approach to evaluating and implementing architectures
- Established best practices developed through lessons learned by working with customers

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

4

AWS developed the Well-Architected Framework after reviewing thousands of customers' architectures on AWS. It is designed to help you build the most secure, high-performing, resilient, and efficient infrastructure possible in a cloud-native way. It provides a consistent approach for evaluating architectures. You can implement designs through a set of questions across five pillars with design principles that scale with your needs over time.

Five pillars of the framework

The diagram illustrates the five pillars of the Well-Architected Framework as the columns of a classical building. The building has a triangular pediment at the top. Below each pillar is a color-coded icon representing a pillar: Operational excellence (blue gears), Security (yellow shield), Reliability (pink chain), Performance efficiency (orange speedometer), and Cost optimization (green dollar bill). The entire diagram is enclosed in a rounded rectangular frame.

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The framework is divided into five areas of focus, referred to as the five pillars of the Well-Architected Framework. They include:

- Operational excellence
- Security
- Reliability
- Performance efficiency
- Cost optimization

Operational excellence



- Perform operations as code
- Annotate documentation
- Make frequent, small, reversible changes
- Refine operations procedures frequently
- Anticipate failure
- Learn from all operational failures

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

b

The operational excellence pillar includes the ability to run and monitor systems to deliver business value and to continually improve supporting processes and procedures.

Design principles

- **Perform operations as code.** Treat your entire workload, including infrastructure, the same as you would application code. Use scripting and automation to trigger actions in response to events. This helps you limit human error and enable consistent responses to events.
- **Annotate documentation.** Most often, documentation is created by hand and doesn't automatically keep up with the pace of change. Seek to automate the documentation update process to align with each change in environment. Also, you can use annotations as an input to your operations code.
- **Make frequent, small, reversible changes.** Design workloads to allow the regular update of components. Make changes in small increments that can be reversed if they fail (without affecting customers, when possible).
- **Refine operations procedures frequently.** As you evolve your workload, evolve your procedures appropriately.
- **Anticipate failure.** Research potential sources of failure so that you can remove or mitigate them. Test both failure scenarios and failure response procedures.
- **Learn from all operational failures.** Drive improvement through lessons learned from all operational events and failures.

Security



- Implement a strong identity foundation
- Enable traceability
- Apply security at all layers
- Automate security best practices
- Protect data in transit and at rest
- Prepare for security events

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

The security pillar includes the ability to protect information, systems, and assets while delivering business value through risk assessments and mitigation strategies.

Design principles

- **Implement a strong identity foundation.** Adhere to the principle of least privilege, and enforce separation of duties. Use the appropriate authorization for each interaction with your AWS resources. Centralize privilege management and reduce or even eliminate reliance on long-term credentials.
- **Enable traceability.** Monitor, alert, and audit actions and changes to your environment in real time. Integrate logs and metrics with systems to automatically respond and take action.
- **Apply security at all layers.** Instead of focusing on only protecting a single outer layer, apply a *defense in depth* approach with other security controls. Apply to all layers, for example, edge network, virtual private cloud (VPC), subnet, load balancer, every instance, operating system, and application.
- **Automate security best practices.** Automation improves your ability to securely scale more rapidly and cost effectively. Create secure architectures, including the implementation of controls that are defined and managed as code.

- **Protect data in transit and at rest.** Classify your data into sensitivity levels and use mechanisms, such as encryption and tokenization, where appropriate. Reduce or eliminate direct human access to data to reduce risk of loss or modification.
- **Prepare for security events.** Institute an incident management process. Run incident response simulations, and use tools with automation to increase your speed for detection, investigation, and recovery.

Reliability



- Test recovery procedures
- Automatically recover from failure
- Scale horizontally to increase aggregate system availability
- Stop guessing capacity
- Manage change in automation

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

8

The reliability pillar includes the ability of a system to recover from infrastructure or service disruptions, dynamically acquire computing resources to meet demand, and mitigate disruptions, such as misconfigurations or transient network issues.

Design principles

- **Test recovery procedures.** In an on-premises environment, testing is often conducted to prove that the system works in a particular scenario. Testing is not typically used to validate recovery strategies. In the cloud, you can test failure scenarios, such as when different components or tiers of the application become unresponsive, and you can validate your recovery procedures. You can use automation to simulate different failures or to recreate scenarios that led to failures before. This exposes failure pathways that you can test and rectify before a real failure scenario, reducing the risk of components failing that have not been tested before.
- **Automatically recover from failure.** By monitoring a system for key performance indicators (KPIs), you can trigger automation when a threshold is breached. This allows for automatic notification and tracking of failures and for automated recovery processes that work around or repair the failure. With more sophisticated

automation, it's possible to anticipate and remediate failures before they occur.

- **Scale horizontally to increase aggregate system availability.** Replace one large resource with multiple small resources to reduce the impact of a single failure on the overall system. Distribute requests across multiple, smaller resources to ensure that they don't share a common point of failure.
- **Stop guessing capacity.** A common cause of failure in on-premises systems is resource saturation, when the demands placed on a system exceed the capacity of that system (including denial of service attacks). In the cloud, you can monitor demand and system utilization, and automate the addition or removal of resources to maintain the optimal level to satisfy demand.
- **Manage change in automation.** Change your infrastructure by using automation. You must also manage the changes to the automation configuration itself.

Performance efficiency



- Democratize advanced technologies
- Go global in minutes
- Use serverless architectures
- Experiment more often
- Apply mechanical sympathy

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

9

The performance efficiency pillar includes the ability to use computing resources efficiently to meet system requirements and to maintain that efficiency as demand changes and technologies evolve.

Design principles

- **Democratize advanced technologies.** Some complex technologies require expertise that is not evenly dispersed across the technical community, such as NoSQL databases, media transcoding, and machine learning. In the cloud, these technologies can become services that your team can consume while focusing on product development instead of resource provisioning and management.
- **Go global in minutes.** Easily deploy your system in multiple Regions around the world with just a few clicks. This capability provides lower latency and a better experience for your customers at minimal cost.
- **Use serverless architectures.** In the cloud, serverless architectures remove the need for you to run and maintain servers to carry out traditional compute activities. For example, storage services can act as static websites, removing the need for web servers, and event services can host your code for you. This not only

removes the operational burden of managing these servers but also can lower transactional costs, because these managed services operate at cloud scale.

- **Experiment more often.** With virtual and automatable resources, you can quickly carry out comparative testing by using different types of instances, storage, or configurations.
- **Apply mechanical sympathy.** Use the technology approach that aligns best to what you are trying to achieve. For example, consider data access patterns when selecting database or storage approaches.

Cost optimization



- Adopt a consumption model
- Measure overall efficiency
- Stop spending money on data center operations
- Analyze and attribute expenditure
- Use managed services to reduce cost of ownership

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

10

The cost optimization pillar includes the ability to avoid or reduce unneeded cost or suboptimal resources.

Design principles

- **Adopt a consumption model.** Pay only for the computing resources that you consume and increase or decrease usage depending on business requirements, not by using elaborate forecasting. For example, development and test environments are typically used for only eight hours a day during the work week. You can stop these resources when they are not in use for a potential cost savings of 75 percent (40 hours versus 168 hours).
- **Measure overall efficiency.** Measure the business output of the system and the costs associated with delivering it. Use this measure to understand the gains you make from increasing output and reducing costs.
- **Stop spending money on data center operations.** AWS does the hard work of racking, stacking, and powering servers so that you can focus on your customers and business projects instead of on the IT infrastructure.

- **Analyze and attribute expenditure.** The cloud makes it easier to accurately identify the usage and cost of systems, which then allows attribution of IT costs to the various business owners. This helps measure return on investment and gives system owners an opportunity to optimize their resources and reduce costs.
- **Use managed services to reduce cost of ownership.** In the cloud, managed services remove the operational burden of maintaining servers for tasks like sending email or managing databases. And because managed services operate at cloud scale, they can offer a lower cost per transaction or service.

Reference architectures

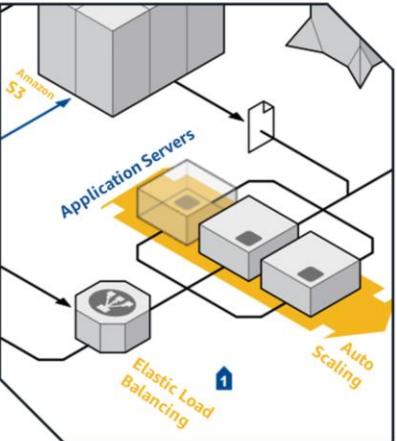
© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



12

Reference architectures are rough outlines of best practices of deployment scenarios. They're a great starting point. In this example, we will review an architecture designed for reliability.

Reference architectures



- Visually represent application architecture
- Demonstrate how services combine to form a solution
- Provide guidance on building applications
- Serve as templates to accelerate delivery

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

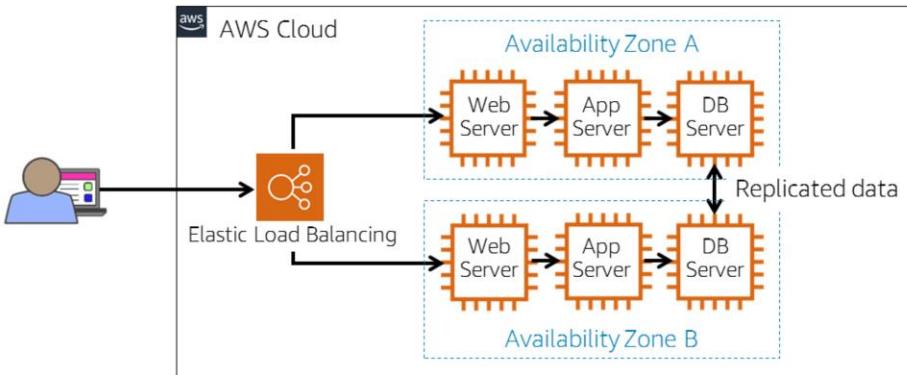
15

- Visually represent application architecture
- Demonstrate how services combine to form a solution
- Provide guidance on building applications that take full advantage of the AWS Cloud
- Serve as templates to accelerate delivery through the re-use of an effective solution

Sample reference architectures:

- Web application:
https://media.amazonwebservices.com/architecturecenter/AWS_ac_ra_web_01.pdf
- Disaster recovery architecture:
https://media.amazonwebservices.com/architecturecenter/AWS_ac_ra_disasterrecovery_07.pdf
- Serverless architecture: <https://docs.aws.amazon.com/solutions/latest/instancescheduler/architecture.html>

Example: Improving availability with Elastic Load Balancing



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

14

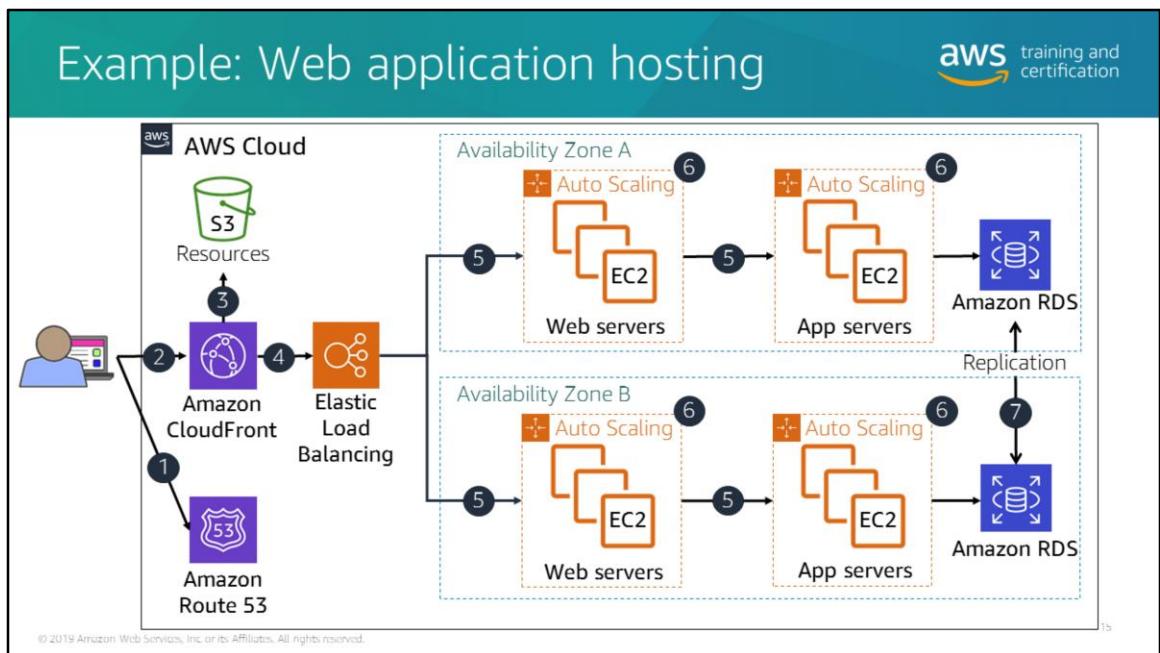
This is an example of a reference architecture that focuses on the reliability pillar of the Well-Architected Framework. Load balancing is an effective way to increase the availability of a system. Instances that fail can be replaced seamlessly behind the load balancer while other instances continue to operate. Use Elastic Load Balancing to balance across instances in multiple Availability Zones of an AWS Region.

Availability Zones are distinct geographical locations that are engineered to be insulated from failures in other Availability Zones. By placing Amazon EC2 instances in multiple Availability Zones, an application can be protected from failure at a single location. It is important to run independent application stacks in more than one Availability Zone, either in the same Region or in another Region so that if one zone fails, the application in the other zone can continue to run. When you design such a system, you need a good understanding of zone dependencies.

AWS provides services and infrastructure to build reliable, fault-tolerant, and highly available systems in the cloud.

Most of the higher-level services, such as Amazon S3, Amazon Simple Queue Service (Amazon SQS), and Elastic Load Balancing, have been built with native fault tolerance

and high availability features. Services that provide basic infrastructure, such as Amazon EC2 and Amazon EBS, provide infrastructure building blocks that, by themselves, may not be fault-tolerant. Hard drives may fail, power supplies may fail, and racks may fail. AWS provides specific features, such as Availability Zones, Elastic IP addresses, and snapshots, that a fault-tolerant and highly available system must take advantage of and use correctly. Only moving a system into the cloud doesn't make it fault-tolerant or highly available.



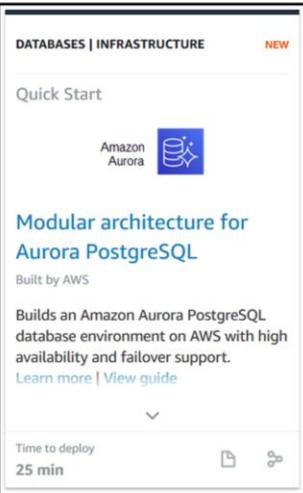
Here is another example that uses load balancing. This architecture is similar to the three-tier web application shown in an earlier lesson. In this example, the solution is spread across two Availability Zones, increasing the availability and performance of the solution. This architecture opts for using web servers as the first tier instead of relying on the load balancer, which is more limited than a fully featured web server. Also note that a second load balancer could be added between the first (web) and second (application) tiers, which might improve performance.

Here is an overview of the architecture:

1. The user's DNS requests are served by **Amazon Route 53**, a highly available DNS service. Network traffic is routed to infrastructure running in AWS.
2. **CloudFront**, a global network of edge locations, delivers static streaming and dynamic content. Requests are automatically routed to the nearest edge location so that content is delivered with the best possible performance.
3. The web application uses resources and static content that are stored on **Amazon S3**, a highly durable storage infrastructure designed for mission-critical and primary data storage.
4. HTTP requests are first handled by **Elastic Load Balancing**, which automatically

distributes incoming application traffic among multiple **Amazon EC2** instances across **Availability Zones**. It enables fault tolerance in your applications, seamlessly providing the amount of load-balancing capacity needed in response to incoming application traffic.

5. Web servers and application servers are deployed on **Amazon EC2** instances. Most organizations select an **Amazon Machine Image** (AMI), and then customize it to their needs. This custom AMI will then become the starting point for future web development.
6. Web servers and application servers are deployed in an **EC2 Auto Scaling** group. Auto Scaling automatically adjusts your capacity up or down according to conditions you define. With EC2 Auto Scaling, the number of Amazon EC2 instances that you're using increases seamlessly when demand spikes, which maintains performance. Your number of instances decreases automatically when demand lessens, which lowers costs.
7. To provide high availability, the relational database that contains an application's data is hosted redundantly on a Multi-AZ deployment of **Amazon RDS**.



The screenshot shows the AWS Quick Starts interface. At the top, it says "AWS Quick Starts" and "aws training and certification". On the left, there's a card for "Amazon Aurora". The card has a "NEW" badge at the top right. It says "Quick Start" and "Amazon Aurora" with a database icon. Below that, it says "Modular architecture for Aurora PostgreSQL" and "Built by AWS". It describes the service as building an Amazon Aurora PostgreSQL database environment on AWS with high availability and failover support. It includes links "Learn more" and "View guide". At the bottom, it says "Time to deploy 25 min" and has download and share icons.

- AWS CloudFormation templates
- Built by AWS solutions architects and partners based on AWS best practices
- Include a guide with deployment instructions

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved. 1b

Quick Starts are built by AWS solutions architects and partners to help you deploy popular technologies on AWS, based on AWS best practices for security and high availability. These accelerators reduce hundreds of manual procedures into only a few steps, so you can build your production environment quickly and start using it immediately.

Each Quick Start includes AWS CloudFormation templates that automate the deployment and a guide that discusses the architecture and provides step-by-step deployment instructions.

To get started, visit the AWS Quick Starts page online at
<https://aws.amazon.com/quickstart/>

Knowledge check



Which of the following is **NOT** a pillar of the AWS Well-Architected Framework?

- A. Security
- B. Persistence
- C. Cost Optimization
- D. Operational Excellence

Knowledge check



Which of the following is **NOT** a pillar of the AWS Well-Architected Framework?

- A. Security
- B. Persistence
- C. Cost Optimization
- D. Operational Excellence

B is correct.

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

18

Key Takeaways



- The Well-Architected Framework
 - Designed to help you build secure, high-performing, resilient, and efficient infrastructure in a cloud-native way
 - Provides a consistent approach for evaluating architectures and implementing designs
 - Established five pillars with design principles that scale with your needs over time
- AWS
 - Provides reference architectures to help you design infrastructure to fit your needs

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

19

- Designed to help build the most secure, high-performing, resilient, and efficient infrastructure possible in a cloud-native way
- Provides a consistent approach for evaluating architectures and implementing designs through a set of questions across five pillars with design principles that scale with your needs over time.
- AWS Provides references architectures to help you design infrastructure to fit your needs.

For more information, see <https://aws.amazon.com/architecture/well-architected>.

Conclusion



The future of the AWS Cloud

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.



Future of cloud computing

The diagram features four hexagonal nodes connected by white lines. The top-left node is dark blue and labeled 'Serverless'. The bottom-left node is teal and labeled 'Shrinking "edge" IoT'. The bottom-right node is purple and labeled 'Purpose-built services'. The fourth node, located at the top-right, contains three blue cubes arranged in a 2x2 grid with a third cube above them, suggesting a 3D storage or data structure. Each node has a small white hexagon with a black dot at its center.

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

22

There are a lot of exciting things happening on the cloud. Some of these are:

- Serverless Architecture, which enables you to build and run applications and services without considering servers. It reduces infrastructure management tasks, such as server or cluster provisioning, patching, operating system maintenance, and capacity provisioning;
- Purpose-built services, which are tailored to your needs for the best performance, scalability, availability and at the lowest cost; and
- Internet of Things (IoT), which is closing the gap between the physical and digital world in self-reinforcing and self-improving systems.

Powering customer innovation



- Enterprise transformation



- Robotics

- Predictive analytics



- Gaming

- Machine learning



- Enterprise applications

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

25

AWS offers a broad set of global cloud-based products to help organizations move faster and lower IT costs and scale. AWS continues to follow its customer-centric process of inventing new products and services to better assist you. It's not only about being close to customers and asking them what they want but also about deeply understanding their situation and context so that AWS can invent *on their behalf*.

The AWS Cloud platform continues to expand daily.

For more information on what's new with AWS, see <https://aws.amazon.com/new/>.

Thank you!



© 2019 Amazon Web Services, Inc. or its affiliates. All rights reserved.