AWS CERTIFIED

CLOUD PRACTITIONER 2019

TRAINING NOTES

Fast-track your exam success with the ultimate cheat sheet for the CLF-C01 exam!

- Over 150 pages of detailed and regularly updated facts; everything you need to know to pass the exam in a concise, easy-to-read format.
- Master the new exam pattern with our exam-difficulty practice questions and detailed explanations.

Neal Davis

DigitalCloud TRAINING

GETTING STARTED

Welcome

Thanks for purchasing these training notes for the AWS Certified Cloud Practitioner exam. The information in this document relates to the latest version of the CLF-C01 exam. This is a living document that will be updated periodically as the exam is updated and you have free access to download all subsequent versions.

The CLF-C01 exam covers a broad set of AWS services and the aim of putting this information together is to provide a centralized, detailed list of the facts you need to know before you sit the exam. This will shortcut your study time and maximize your chance of passing the exam first time.

We hope you get great value from this resource and wish you the best of luck with your AWS Certified Cloud Practitioner exam.

How to best prepare for your exam

Please note that this document does not read like a book or instructional text, we provide a raw, point-to-point list of facts with some tables and diagrams thrown in to help with understanding.

If you're new to AWS or to cloud computing in general there are some foundational concepts in the first few sections. It's also recommended to take an online instructor led course to familiarize yourself with the AWS platform.

The scope of coverage of services, and what information is included for each service, is based on feedback from many students who've taken the exam, as well as our own experience and may differ between AWS services.

We recommend taking our practice exams (see below) to identify your strengths and weaknesses and then use the training notes to focus your study on the knowledge areas where you need to most.

Practice Exams

When you're feeling ready to test your knowledge, check out our AWS Certified Cloud Practitioner Practice Tests on the Digital Cloud Training website. Our practice tests are designed to reflect the difficulty of the real exam and are the closest to the real AWS exam experience available today.

To accommodate different learning styles, there are multiple practice types available. Combined, these are powerful tools to prepare you for your exam.

1. **Exam simulation mode** where you get the full timed, scored, exam experience.
2. **Training mode** where you can check every answer as you go through the exam.
3. **Knowledge reviews** - once you've identified your strengths and weaknesses using the practice exams, knowledge reviews allow you to view questions from a specific knowledge area so you can focus your time where you need to most.

There are currently 6 practice exams of 65 questions each (390 questions in total), and a larger pool of questions available in the Knowledge Reviews.

Enrol now to fast-track your AWS Certified Cloud Practitioner exam success!

https://digitalcloud.training/aws-certified-cloud-practitioner-practice-tests-2019/

Bonus Offer

To gain access to a free 65-question practice exam on our interactive exam simulator, please send an email to info@digitalcloud.training with "CCP-BONUS" in the subject line and include a copy of your purchase receipt. If you have already purchased our full set of practice questions, please note that these questions are already included.

Contact, Support & Sharing

We hope you get value from these resources and please feel free to ask any questions or provide any feedback you may have.

Our private Facebook group is a great place to ask questions and share knowledge and exam tips with the community. Please join using the link below:

https://www.facebook.com/groups/awscertificationqa

The AWS platform is evolving quickly and the exam tracks these changes with a typical lag of around 6 months. We're therefore reliant on student feedback to keep track of what is appearing in the exam so please post your exam feedback to our Facebook group.

For technical support you can contact us at: support@digitalcloud.training

TABLE OF CONTENTS

Getting Started 2
- Welcome 2
- How to best prepare for your exam 2
- Practice Exams 2
- Bonus Offer 3
- Contact, Support & Sharing 3

Table of Contents 4

Compute, Storage and Network Concepts 11

Compute 11
- Introduction to Compute 11
- Compute Instances on AWS 13
- Amazon EBS & Snapshots 15
- AWS Infrastructure Services 15
- Load Balancing and Auto Scaling 17

Storage 19
- Introduction to Storage 19
- Storage Concepts 19
- Measuring Data 19
- Data Accessibility SLAs 20
- Cloud Storage Types 21

Network 25
- Introduction to Network 25
- IP Addressing 25
- Routing and Gateways 26
- The OSI Model 26
- Network Virtualization 27
- Virtual Private Cloud (VPC) 28

Cloud Computing Concepts 31

General Cloud Computing Concepts 31

The Six Advantages 31
- Trade capital expense for variable expense 31
- Benefit from massive economies of scale 31
- Stop guessing about capacity 32
- Increase speed and agility 32
- Stop spending money running and maintaining data centers 32
- Go global in minutes 32

Types of Cloud Computing 32
- Infrastructure as a Service (IaaS) 32
- Platform as a Service (PaaS) 33
- Software as a Service (SaaS) 33

Types of Cloud Deployment 33
- Public Cloud 33
- Hybrid 34
- On-premises 34

Cloud Computing Concepts Practice Questions 34
AWS Global Infrastructure 37
General 37
Regions 38
Availability Zones 38
Edge Locations and Regional Edge Caches 39
AWS Global Infrastructure Practice Questions 40
Identity and Access Management 43
General 43
Authentication Methods 45
IAM Users 46
Groups 47
Roles 47
Policies 48
AWS STS 49
IAM Best Practices 51
Identity and Access Management Practice Questions 51
AWS Compute 54
Amazon EC2 54
Amazon EC2 Container Service (ECS) 56
AWS Lambda 57
Amazon LightSail 58
Instances 58
Databases 58
AWS Compute Practice Questions 59
AWS Storage 62
Amazon Simple Storage Service (S3) 62
AWS Snowball 64
Amazon Elastic Block Store (EBS) 64
Amazon Elastic File Service (EFS) 66
AWS Storage Practice Questions 67
AWS Networking 69
General Amazon Virtual Private Cloud (VPC) 69
Subnets 71

Firewalls 71
VPC Wizard 72
NAT Instances 73
NAT Gateways 73
AWS Direct Connect 74
AWS Networking Practice Questions 75
AWS Databases 78
Amazon Relational Database Services (RDS) 78
Amazon DynamoDB 80
Amazon RedShift 81
Amazon ElastiCache 82
AWS Database Practice Questions 83
Elastic Load Balancing and Auto Scaling 85
Amazon Elastic Load Balancing (ELB) 85
Application Load Balancer (ALB) 85
Network Load Balancer (NLB) 85
Classic Load Balancer (CLB) 85
AWS Auto Scaling 86
Elastic Load Balancing and Auto Scaling Practice Questions 87
Content Delivery and DNS Services 89
Amazon Route 53 89
Amazon CloudFront 89
Content Delivery and DNS Services Practice Questions 89
Monitoring and Logging Services 92
Amazon CloudWatch 92
AWS CloudTrail 93
Monitoring and Logging Services Practice Questions 94
Notification Services 97
Amazon Simple Notification Service (SNS) 97
Notification Services Practice Questions 98
Billing and Pricing 100
General Pricing Information 100
Amazon EC2 pricing 101
Amazon Simple Storage Service (S3) Pricing 103

Amazon Glacier pricing 103
AWS Snowball Pricing 103
Amazon Relational Database Service (RDS) Pricing 104
Amazon CloudFront Pricing 104
AWS Lambda Pricing 104
Amazon Elastic Block Store (EBS) Pricing 105
Amazon DynamoDB Pricing 105
AWS Support Plans 106
Resource Groups and Tagging 107
AWS Organizations and Consolidated Billing 108
AWS Quick Starts 109
AWS Cost Calculators and Tools 109
AWS Cost Explorer 109
AWS Simple Monthly Calculator 110
TCO Calculator 110
Billing and Pricing Practice Questions 110
Cloud Security 113
General 113
Benefits of AWS Security 113
Compliance 113
AWS WAF & AWS Shield 114
AWS WAF 114
AWS Shield 114
AWS Key Management Service (KMS) 114
AWS CloudHSM 115
AWS Artifact 115
AWS Inspector and AWS Trusted Advisor 115
AWS Inspector 115
AWS Trusted Advisor 116
AWS Personal Health Dashboard 116
Penetration Testing 117
Cloud Security Practice Questions 118
Shared Responsibility Model 121
Shared Responsibility Model Practice Questions 122
Architecting for the Cloud 125
The Cloud Computing Difference 125

IT assets become programmable resources ... 125
Global, available and unlimited capacity ... 125
Higher level managed services ... 125
Security built-in ... 125

Design Principles ... 126
Scalability ... 126

Disposable Resources Instead of Fixed Servers ... 127

Automation ... 128

Loose Coupling ... 129

Services, Not Servers ... 129

Databases ... 130
Relational Databases ... 130
NoSQL Databases ... 131
Search ... 132

Removing Single Points of Failure ... 132
Introducing Redundancy ... 133
Detect Failure ... 133
Durable Data Storage ... 133
Automated Multi-Data Center Resilience ... 133
Fault Isolation and Traditional Horizontal Scaling ... 134
Optimize for Cost ... 134
Reserved Capacity ... 135
Spot Instances ... 135

Caching ... 136
Application Data Caching ... 136
Edge Caching ... 136

Security ... 136
Utilize AWS Features for Defence in Depth ... 137
Offload Security Responsibility to AWS ... 137
Reduce Privileged Access ... 137
Security as Code ... 137
Real-Time Auditing ... 138

Architecting for the Cloud Practice Questions ... 138

Additional Tools and Services ... 141

Compute ... 141
Amazon Elastic Container Service for Kubernetes (EKS) ... 141
AWS Batch ... 141
AWS Elastic Beanstalk ... 142

Storage ... 142
AWS Storage Gateway ... 142

Database ... 143
Amazon Elasticache ... 143
Amazon Neptune ... 143

Migration ... 143

AWS Migration Hub 143
AWS Database Migration Service 144
AWS Server Migration Service 144

Networking & Content Delivery 144
Amazon API Gateway 144
AWS Direct Connect 145

Developer Tools 145
AWS CodeStar 145
AWS CodeCommit 145
AWS CodeBuild 146
AWS CodeDeploy 146
AWS CodePipeline 146
AWS X-Ray 147

Management Tools 147
AWS CloudFormation 147
AWS Config 147
AWS OpsWorks 148
AWS Service Catalog 148
AWS Systems Manager 148
AWS Managed Services 148

Analytics 149
Amazon Athena 149
Amazon EMR 149
Amazon CloudSearch 149
Amazon Elasticsearch 150
Amazon Kinesis 150
AWS Data Pipeline 150
AWS Glue 151

Media Services 151
Amazon Elastic Transcoder 151

Security, Identity and Compliance 151
Amazon Cognito 151
AWS Certificate Manager 151
AWS CloudHSM 152
AWS Directory Service 152
AWS Artifact 152

Machine Learning 153
Amazon Rekognition 153
Amazon SageMaker 153
Amazon Comprehend 153
Amazon Transcribe 153

Mobile Services 154
AWS AppSync 154
AWS Device Farm 154

Application Integration 154
AWS Step Functions 154
Amazon MQ 155
Amazon SQS 155

Amazon SWF 155
Internet of Things 155
AWS IoT Core 155
Desktop & App Streaming 156
Amazon Workspaces 156
Conclusion 157
Other Books by this author 158
AWS Certified Solutions Architect Associate Training Notes 158
AWS Certified Solutions Architect Associate Practice Questions 159
About the Author 160

COMPUTE, STORAGE AND NETWORK CONCEPTS

This section provides a basic overview of some important compute, storage and networking concepts. It is aimed at those who are not from a technical role, or who are new to cloud computing or IT in general.

Please note that the content within this section does not relate directly to the exam blueprint, it is foundational knowledge which will help you to understand some of the technical concepts that are presented later.

COMPUTE

Introduction to Compute

Along with storage and networking, compute is one of the key foundational building blocks of the cloud computing infrastructure layer. We will discuss the basic concepts you need to understand to get started with compute on AWS.

Fundamentally the term "compute" refers to physical servers comprised of the processing, memory, and storage required to run an operating system such as Microsoft Windows or Linux, and some virtualized networking capability.

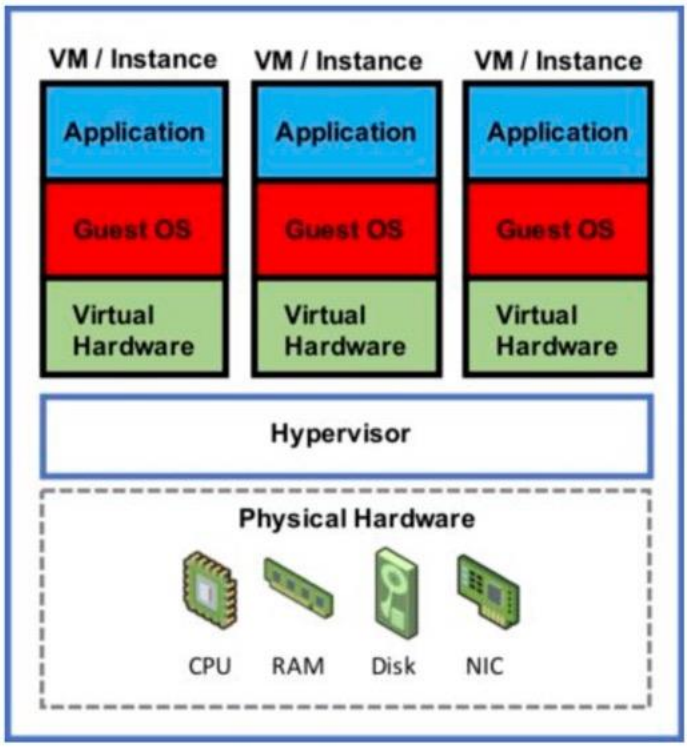
The components of a compute server include the following:

- **Processor or Central Processing Unit (CPU)** - the CPU is the brains of the computer and carries out the instructions of computer programs
- **Memory or Random Access Memory (RAM)** - within a computer memory is very high speed storage for data stored on an integrated-circuit chip
- **Storage** - the storage location for the operating system files (and optionally data). This is typically a local disk stored within the computer or a network disk attached using a block protocol such as iSCSI
- **Network** - physical network interface cards (NICs) to support connectivity with other servers

When used in cloud computing, the operating system software that is installed directly on the server is generally a hypervisor which provides a hardware abstraction layer onto which additional operating systems can be run as virtual machines (VMs) or "instances". This technique is known as **hardware virtualization**.

A VM is a container within which virtualized resources including CPU (vCPU), memory and storage are presented, and an operating system can be installed. Each VM is isolated from other VMs running on the same host hardware and many VMs can run on a single physical host, with each potentially installed with different operating system software.

The diagram below depicts hardware virtualization with guest VMs running on top of a host OS:

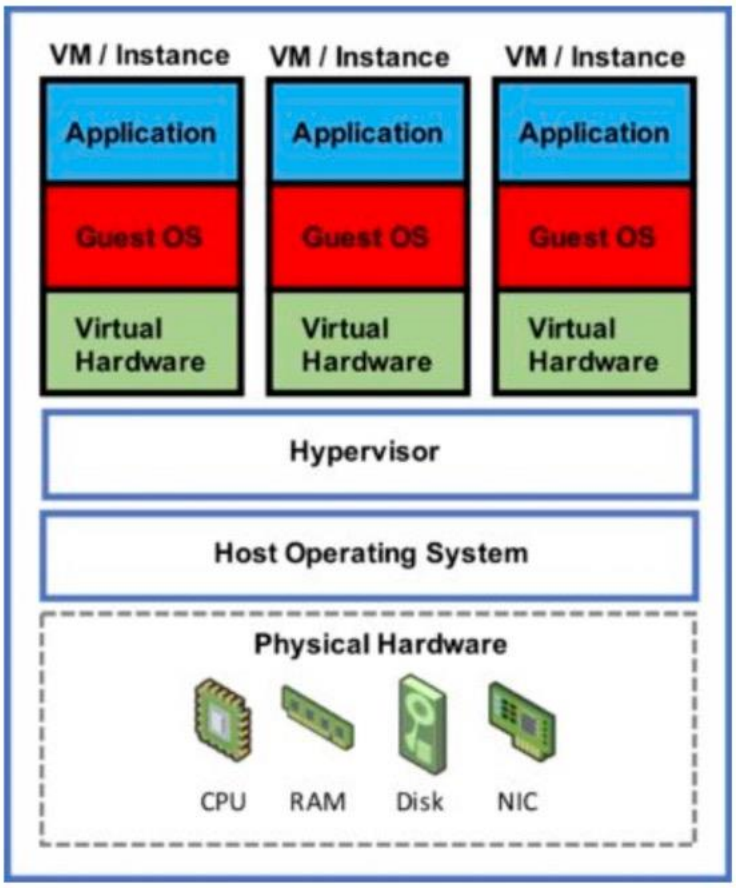


There are two main types of hypervisor:

- **Type 1** - the hypervisor is installed directly on top of the hardware and is considered a "bare-metal" hypervisor
- **Type 2** - the hypervisor software runs on top of a host operating system

Examples of Type 1 hypervisors include VMware ESXi and Microsoft Hyper-V and examples of Type 2 hypervisors include VMware Workstation and Oracle Virtual Box. Type 1 hypervisors typically provide better performance and security than Type 2 hypervisors.

The diagram above shows a hardware virtualization stack using a Type 1 hypervisor. The diagram below depicts a Type 2 hypervisor:



As you can see, the key difference is that a there is an additional host operating system layer that sits directly above the physical hardware and beneath the hypervisor layer.

Until recently Amazon Web Services (AWS) used the Xen hypervisor but has now moved to an internally developed hypervisor based on the Kernel-based Virtual Machine (KVM) technology. KVM is generally considered to be a Type 1 hypervisor

Compute Instances on AWS

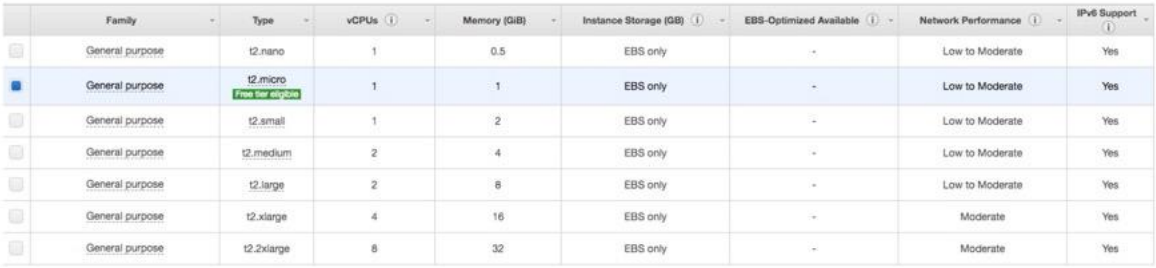
In AWS compute is consumed through the **Elastic Compute Cloud (EC2)** which is a web service from which you can launch "instances" which are essentially VMs running on the AWS KVM hypervisor.

Amazon EC2 provides secure, resizable compute capacity in the cloud on a pay-as-you-go basis with no fixed term contracts (unless you choose reserved instances to reduce cost).

There are a large selection of instance types you can choose from which come with varying specifications for vCPU, memory, and storage allocation.

Virtual networking is included with all instances and varies in performance level from low (unspecified performance) up to 25 Gigabit.

The image below shows a few "General Purpose" instance types. Note the different configurations for vCPU, Memory and Network Performance:



Family	Type	vCPUs	Memory (GiB)	Instance Storage (GB)	EBS-Optimized Available	Network Performance	IPv6 Support
General purpose	t2.nano	1	0.5	EBS only	-	Low to Moderate	Yes
General purpose	t2.micro Free tier eligible	1	1	EBS only	-	Low to Moderate	Yes
General purpose	t2.small	1	2	EBS only	-	Low to Moderate	Yes
General purpose	t2.medium	2	4	EBS only	-	Low to Moderate	Yes
General purpose	t2.large	2	8	EBS only	-	Low to Moderate	Yes
General purpose	t2.xlarge	4	16	EBS only	-	Moderate	Yes
General purpose	t2.2xlarge	8	32	EBS only	-	Moderate	Yes

Instance types are categorized into families based on how the instance specifications are optimized for different usage scenarios. Optimizations that are available include compute, memory, storage, graphics processing (GPU) or general purpose usage.

The following table shows the instance families currently available and describes the use cases they are best suited for:

Family	Hint	Purpose/Design
D	DATA	Heavy data usage (e.g. file servers, DWs)
R	RAM	Memory optimised
M	MAIN	General purpose (e.g. app servers)
C	COMPUTE	Compute optimised
G	GRAPHICS	Graphics intensive workloads
I	IOPS	Storage I/O optimised (e.g. NoSQL, DWs)
F	FAST	FPGA hardware acceleration for applications
T	CHEAP (think T2)	Lowest cost (e.g. T2-micro)
P	GPU	GPU requirements
X	EXTREME RAM	Heavy memory usage (e.g. SAP HANA, Apache Spark)

When deploying an instance on AWS the first step is to select an **Amazon Machine Image (AMI)**. An AMI is essentially a template that includes the information required to launch an instance in EC2. An AMI includes the following:

- A template for the root volume for the instance (for example, an operating system, an application server, and applications)
- Launch permissions that control which AWS accounts can use the AMI to launch instances

- A block device mapping that specifies the storage volumes to attach to the instance when it's launched

AWS provide a number of AMIs based on various operating systems and configurations. You can also select from the AWS Marketplace, AMIs that have been shared by the community, and your own AMIs that you have previously saved (registered).

Amazon EBS & Snapshots

Most EC2 instance types use the **Elastic Block Store (EBS)** for persistent storage. EBS volumes are durable, block-level storage volumes that can be attached to a single EC2 instance. There are a several different volume types available that differ in performance characteristics and price. These include:

- General Purpose SSD (gp2)
- Provisioned IOPS SSD (io1)
- Throughput Optimized HDD (st1)
- Cold HDD (sc1)
- Magnetic (standard, a previous-generation type)

Each EBS volume is replicated across multiple systems within an Availability Zone (described below) to avoid the risk of data loss if a single hardware component fails. Additionally, users can take **snapshots** of their EBS volumes which are a point-in-time copy of the data.

Snapshots are incremental backups, which means that only the blocks on the device that have changed after your most recent snapshot are saved.

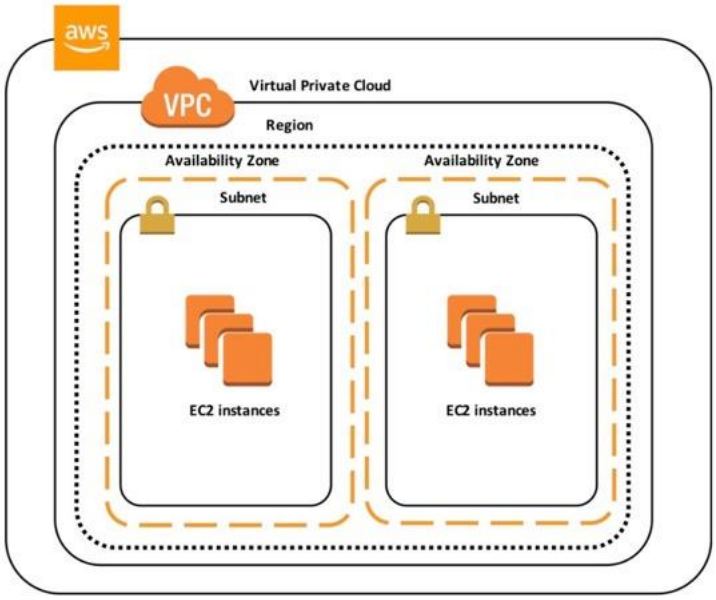
AWS Infrastructure Services

There are a number of supporting services and features on AWS that enable compute instances to be launched in a functional state. These include:

- **Virtual Private Cloud (VPC)** – A **VPC** is a virtual network that provides the networking layer of EC2. A VPC can be configured to your own requirements
- **Elastic Block Store (EBS)** – EBS provides persistent block-based storage volumes that can be attached to EC2 instances

Amazon VPCs are created within AWS Regions, which is a separate geographic area in which multiple Availability Zones (AZs, which are essentially data centers) are located. Amazon provide more information on **regions and availability zones here**.

Subnets are created within AZs and this is where Amazon EC2 instances are deployed. The following diagram depicts this AWS infrastructure:

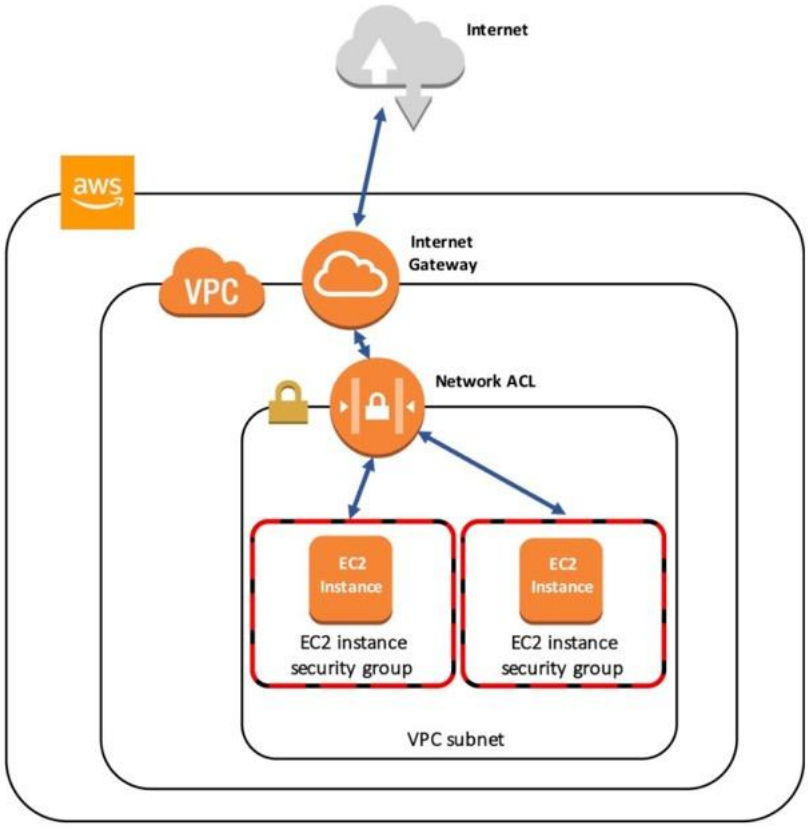


Additionally, to be able to connect to your EC2 instances on the AWS cloud it is necessary to configure **Security Groups**, which are firewalls at the instance level, and **Network Access Control Lists (NACLs)**, which are firewalls at the subnet level.

When a VPC has been properly configured, EC2 instances have been launched with public IP addresses, and Security Groups and NACLs have been configured with the correct rules, it is then possible to directly access EC2 instances from the Internet.

The following simplified diagram depicts the configuration elements required to connect to an Amazon EC2 instance from the Internet.

The diagram shows two EC2 instances with separate security groups but in the same subnet within a VPC. An Internet Gateway provides the Internet connectivity and in this configuration each EC2 instance would require a public IP address:



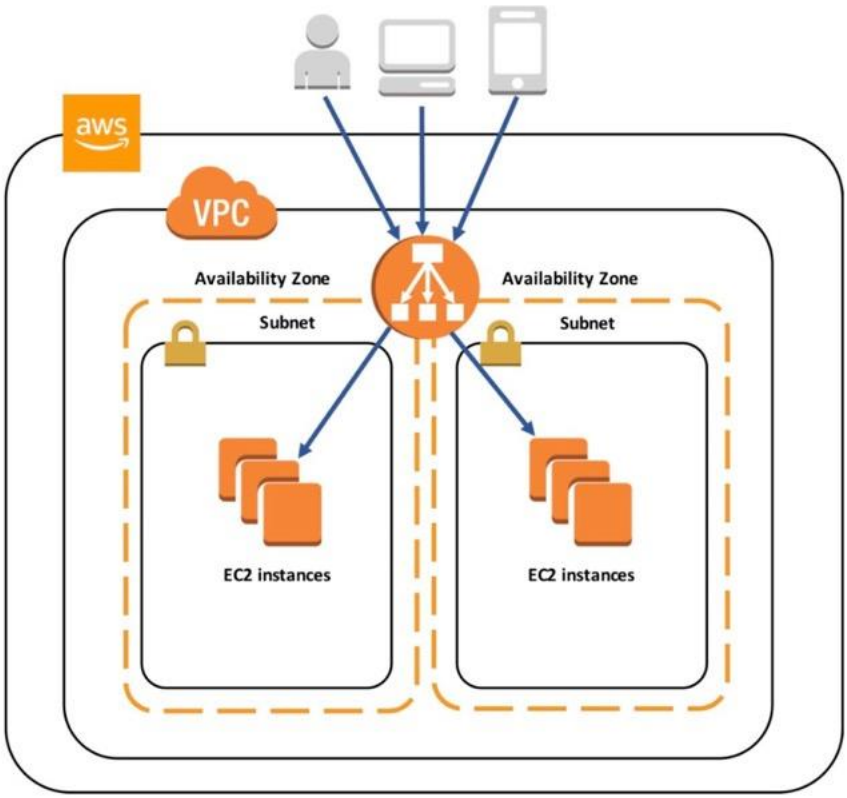
Logging on to EC2 instances involves usage of a key pair (cryptographic key) that you generate through the console and in some cases a password.

Load Balancing and Auto Scaling

Cloud applications are usually deployed in an architecture where multiple instances can share the incoming traffic load and individual instances can be easily added or removed as the load varies up or down.

AWS provide a couple of services that assist with distributing incoming connections and automatically ensuring the right number of instances are available to service the load. These are **Elastic Load Balancing** and **EC2 Auto Scaling**.

The following diagram depicts an Elastic Load Balancer (ELB) servicing a number of EC2 instances across two Availability Zones. Connections from multiple devices hit the ELB which then distributes the connections evenly across the EC2 instances.



Elastic Load Balancing provides the following benefits:

- High availability – ELB automatically distributes traffic across multiple EC2 instances in different AZs within a region
- Security – ELB provides robust security features, including integrated certificate management, user-authentication, and SSL/TLS decryption
- Elasticity – ELB is capable of handling rapid changes in network traffic patterns

EC2 Auto Scaling can complement the architecture depicted in the diagram above by dynamically scaling the number of EC2 instances based on current demand.

EC2 Auto Scaling provides the following benefits:

- Fault tolerance – Auto Scaling detects when an instance is unhealthy and replaces it
- Scalability and elasticity – Auto Scaling automatically scales the number of instances servicing your application based on demand

STORAGE

Introduction to Storage

In cloud computing, cloud storage is a service offering with which a consumer is able to read and write data to cloud-based systems that are managed by a service provider. We will discuss the basic concepts you need to understand to get started with storage on AWS.

Cloud storage is usually accessible through a web service or application programming interface (API). The underlying infrastructure is typically a virtualized infrastructure stack with disk drives that are managed by a software service layer.

The most common form of cloud storage is object storage but any method of data management that can be offered as a service across a network can be used. The other cloud storage options are block-based storage and file-based storage and these will all be discussed within this section.

Storage Concepts

Hard Disk Drives (HDD) have been around for a long while and are still in widespread use today. An HDD is a mechanical drive with spinning platters and a head that floats above the platters and moves into position to read and write data.

HDDs are also known as magnetic drives as they use magnetic polarization to record a one or zero value.

The performance of an HDD depends on a number of factors and these include the following measurements:

- **Revolutions Per Minute (RPM)** – the speed of rotation of the platters
- **Seek time** – the mean time it takes to move the head of a disk drive from one track to another
- **Input / Output Operations Per Second (IOPS)** – the number of IO transactions per second
- **Throughput** – the data transfer rate of a drive

HDDs provide good throughput, large capacity, and are extremely low cost.

Solid State Drives (SSD) store data on non-volatile microchips and have no moving parts. Non-volatile SSD chips differ from computer memory in that the data is retained when power is removed.

SSDs offer extremely high IOPS performance when compared to HDDs and also provide good throughput. SSDs are also much more expensive.

Measuring Data

Stored data is typically measured using the decimal system in kilobytes (kB), Megabytes (MB), Gigabytes (GB), Terabytes (TB) and Petabytes (PB).

In some cases the binary prefix is used such as gibibyte (GiB). A gibibyte is equal to 1024 mebibytes (MiB) while a gigabyte (GB) is equal to 1000 megabytes (MB).

To confuse matters a GB of computer memory is equal to 1024 MB (rather than 1000 MB) and some storage manufacturers have been known to use this measurement for disks too.

The following table shows how each term relates to the other in both the decimal and binary formats and the values are the number of bytes (a byte is 8 bits).

Decimal Name	Decimal Abbr.	Decimal Value	Binary Name	Binary Abbr.	Binary Value
Kilobyte	kB	1,000	**Kibibyte**	kiB	1,024
Megabyte	MB	1,000,000	**Mebibyte**	MiB	1,048,576
Gigabyte	GB	1,000,000,000	**Gibibyte**	GiB	1,073,741,824
Terabyte	TB	1,000,000,000,000	**Tebibyte**	TiB	1,099,511,627,776

The following link provides some more background on this subject:

https://en.wikipedia.org/wiki/Gibibyte

Data Accessibility SLAs

Cloud service providers will often provide service level agreements (SLAs) for the availability and durability of their storage systems.

Availability relates to system uptime, i.e. the amount of time per month or year that the storage system is operational and can deliver data upon request. Service providers aim to increase availability by designing highly available and fault tolerant storage systems.

Availability is usually expressed as a percentage of uptime in a given year. The following table shows some common availability SLAs and how much downtime each corresponds with:

Availability %	Downtime Per Month	Downtime Per Year
99% (“two nines”)	7.3 hours	3.65 days
99.5% (“two and a half nines”)	3.65 hours	1.83 days
99.9% (“three nines”)	44 mins	8.77 hours
99.95% (“three and a half nines”)	22 mins	4.38 hours
99.99% (“four nines”)	4.38 mins	52 mins

Durability relates to measuring the amount of data that may be lost due to errors occurring when writing data. In other words, durability measures the likelihood of losing some of your data.

Durability is usually expressed as a percentage of reliability and can also be interpreted as the number of files that are likely to be lost in a given year.

The following table shows the four Amazon Simple Storage Service (S3) storage classes with their respective durability SLAs and how many files could be lost per year:

AWS Storage Class	Durability %	Files lost per year per PB
Amazon S3 RRS	99.99% ("four nines")	12 million
S3 Standard	99.999999999% ("eleven nines")	.12 (one every 8 years)
S3 Standard - IA	99.999999999% ("eleven nines")	.12 (one every 8 years)
Standard	99.999999999% ("eleven nines")	.12 (one every 8 years)

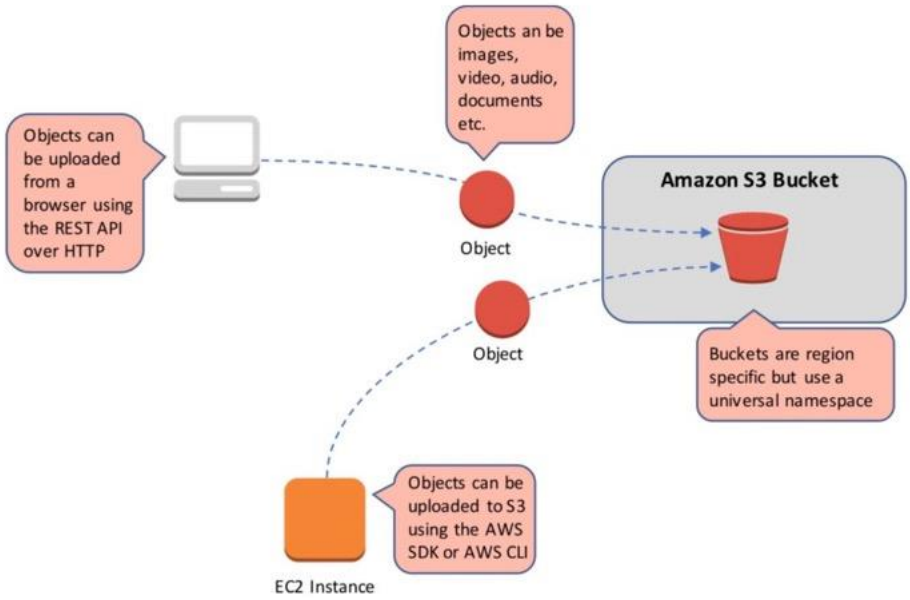
Cloud Storage Types

As mentioned earlier cloud storage is generally object-based, block-based or file-based storage. These terms relate to the type of data stored, the protocols used to access it, and the method of data storage.

Object Storage

With object storage data is managed as individual objects rather than a file hierarchy (as with a traditional file system). Each object includes the data itself, metadata (data about the data), and a globally unique identifier.

Due to its flat file structure, object storage has virtually unlimited scalability and allows the retention of massive amounts of unstructured data. The data is often replicated across multiple physical systems and facilities providing high availability and durability.



Object storage is usually accessed over Representational State Transfer (REST) and Simple Object Access Protocol (SOAP) over Hypertext Transfer Protocol (HTTP).

The **Amazon Simple Storage Service (S3)** is a key, value object-based storage system built to store and retrieve huge amounts of data from any source.

Objects in S3 are stored in a flat structure with no hierarchy. The top level containers within which objects are stored are known as buckets. Though there is no hierarchy S3 does support the concept of folders for organization (grouping of objects).

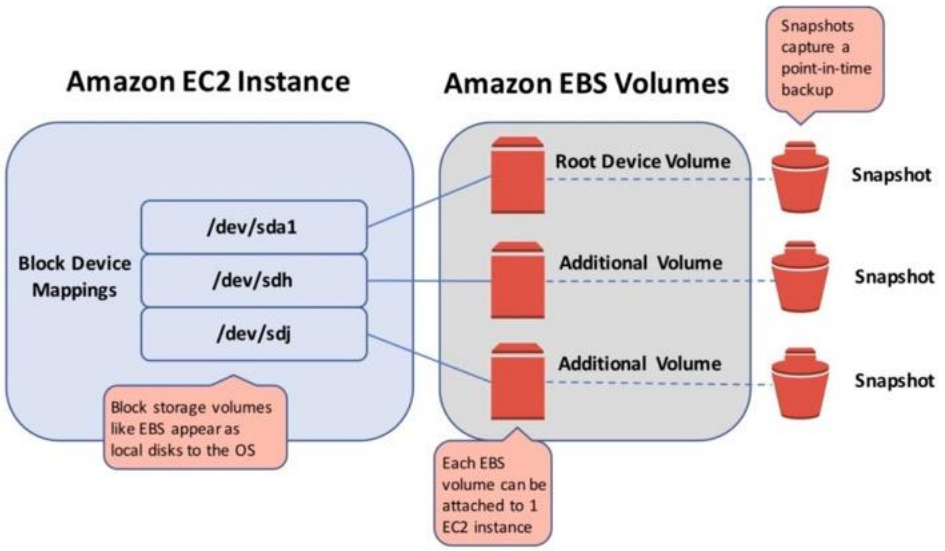
There are several S3 storage classes with varying levels of availability, durability and features. The standard class offers the following features:

- Low latency and high throughput performance
- Designed for durability of 99.999999999% of objects across multiple Availability Zones
- Data is resilient in the event of one entire Availability Zone destruction
- Designed for 99.99% availability over a given year
- Backed with the Amazon S3 Service Level Agreement for availability
- Supports SSL for data in transit and encryption of data at rest
- Lifecycle management for automatic migration of objects

Common use cases for object storage include backup, application hosting, media hosting and software delivery.

Block Storage

Data is stored and managed in blocks within sectors and tracks and is controlled by a server-based operating system. Block storage volumes appear as local disks to the operating system and can be partitioned and formatted.



Block storage is typically used in Storage Area Network (SAN) environments that use the Fibre Channel (FC) protocol as well as Ethernet networks using protocols such as iSCSI or Fibre Channel over Ethernet (FCoE).

Block storage is typically more expensive than object or file storage but provides low latency, and high and consistent performance. The costs are often highest in SAN implementations due to the specialized equipment required.

Amazon Elastic Block Store (EBS) is the AWS service for block storage. EBS provides persistent block storage volumes for use with EC2 instances in the AWS cloud.

There are several EBS volume types to choose from with varying characteristics as can be seen in the table below:

	Solid State Drives (SSD)		Hard Disk Drives (HDD)	
Volume Type	General Purpose SSD (gp2)	Provisioned IOPS SSD (io1)	Throughput Optimized (st1)	Cold HDD (sc1)
Description	Balance of price to performance	High performance SSD	Low cost HDD	Lowest cost HDD
Use Cases	• Most workloads • System boot volumes • Virtual desktops	• Critical business apps that require sustained IOPS performance • Apps that require more than 10,000 IOPS or 160 MiB/s • Large database workloads	• Streaming workloads with fast throughput • Low price • Big data • Data warehouses	• Throughput oriented storage for large volumes of infrequently accessed data • Lowest cost • Cannot be a boot volume
Volume Size	1 GiB – 16 TiB	4 GiB – 16 TiB	500 GiB – 16 TiB	500 GiB – 16 TiB
Max IOPS Per Volume	10,000	32,000	500	250
Max Throughput Per Volume	160 MiB/s	500 MiB/s	500 MiB/s	250 MiB/s

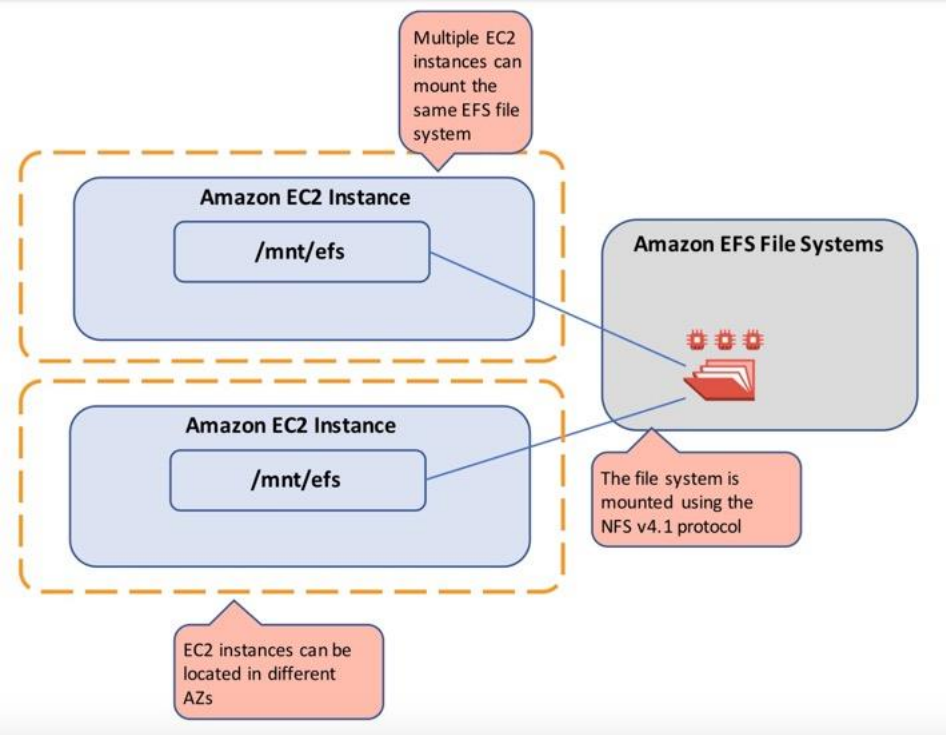
Common use cases for block storage are structured information such as file systems, databases, transactional logs, SQL databases and virtual machines (VMs).

Though the cloud service provider takes care of many aspects of performance and availability, it is also possible to implement a Redundant Array of Inexpensive Disks (RAID) array on Amazon EBS.

File Storage

File storage servers store data in a hierarchical structure using files and folders. Data is accessed as file IDs across a network using either the Server Message Block (SMB) for Windows, or Network File System (NFS) for Unix/Linux.

A file system is mounted via the network to a client computer where it then becomes accessible for reading and writing data. Files and folders can be created, updated, and deleted.



Only file-level operations can occur on a mounted file system, it is not possible to issue block level commands or format or partition the underlying storage volumes.

File storage is easy to implement and use and is generally quite inexpensive. Use cases include web serving and content management, shared corporate directories, home drives, database backups and big data analytics workloads.

The **Amazon Elastic File System (EFS)** is a simple, scalable, elastic file storage in the AWS cloud that is based on NFS. EFS provides the ability to mount a file system to many EC2 instances simultaneously and can achieve high levels of aggregate throughput and IOPS.

EFS is a regional AWS service and provides high availability and durability by storing data redundantly across Availability Zones (AZs).

NETWORK

Introduction to Network

In modern networking, network functions are increasingly becoming abstracted from the underlying switching and routing hardware layer. These virtualized resources are usually API driven, allowing developers to create, update and delete software-based network interfaces, firewalls, load balancers and routing functions through code.

Public cloud providers such as **AWS** offer many network services to customers that can be configured through graphical interfaces, command line and API endpoints. We will discuss the basic concepts you need to understand to get started with network services on AWS.

There are a few supporting concepts that are important to understand if you're working with networking in the cloud that I'll cover off first. These include IP subnetting, routing and gateways, the OSI model and network virtualization.

IP Addressing

Whatever your role in IT, you'll likely need to understand IP addressing to some level. In the cloud you need to understand how to define the IP subnet address ranges your cloud resources will use and the difference between private and public addresses (at a minimum).

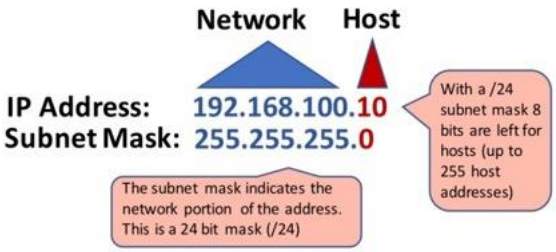
An Internet Protocol (IP) address is a label used to identify a computer on a shared network. There are two versions of IP in common use today: version 4 and version 6.

IPv4 has been around for much longer and is the most well used address range but IPv6, which has a much larger address space, is becoming increasingly common, and is supported by many **AWS services** today.

As IPv4 is the default protocol used on AWS I'll exclude IPv6 from the rest of the discussion. However, it is worth understanding IPv6 and how and why it is used. You can get more information on **IPv6 on Wikipedia**.

An IPv4 address is a 32-bit number which provides up to 4,294,967,296 possible addresses. Each address consists of a network identifier (which represents the network or subnet) and a host identifier (which represents the individual network attached device).

A subnet mask is a prefix that determines which portions of the address represent the network and which represent the hosts (devices). The following diagram depicts this:



A *classful* network design was created back in the 80s that used three classes of network (A, B & C), based on the first octet of the address and using strict octet boundaries for the entire address. This proved to lack the scalability required for the expansion of the Internet and so a classless network design was created. This is known as Classless Inter-Domain Routing (CIDR).

With CIDR variable length subnet masks can be used to allow more granular and efficient use of addresses. An example of CIDR usage is the private address space, which is a reserved address space meant for computers not directly connected to the Internet.

The table below shows how a more granular approach can be taken to allocating addresses:

CIDR Block	Subnet Mask	Max Subnets	Hosts Per Subnet
192.168.0.0/26	255.255.255.192	1024	62
192.168.0.0/27	255.255.255.224	2048	30
192.168.0.0/28	255.255.255.240	4096	14

The CIDR blocks in the table above would allow the creation of subnets with just the right numbers of hosts, this is an efficient way of assigning address blocks.

Routing and Gateways

IP addresses are the means of identifying a unique device on a network. To get to a device across a network a method of determining the best path to get there is required.

This is where routers come in. A router uses a routing protocol (or it may be directly configured) to learn the best path to reach a destination network. This data is held in a routing table.

A router is also considered a gateway when devices on a network are pointed towards it by way of a default gateway address. This address is configured in the IP settings of the networked device and specifies the target for all traffic that is destined to networks other than the local network.

The OSI Model

The Open Systems Interconnection model (OSI model) is a conceptual model that characterizes and standardizes the communication functions of communication and computing systems.

The OSI model divides data communication into 7 abstraction layers and standardizes protocols into groups of networking functionality that ensure interoperability between diverse systems irrespective of the underlying technology.

It's important to understand the 7 layers of the OSI model and where common protocols are located.

The following diagram depicts the 7 layers of the OSI model:

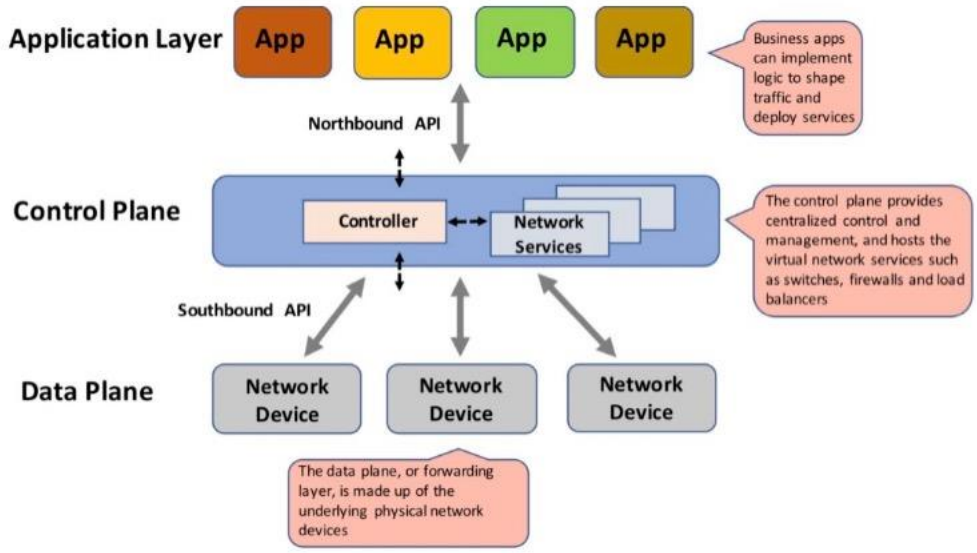
OSI Layer	Data Unit	Function / Protocols
7. Application	Data	Network process to application HTTP, SMTP, IMAP, SNMP, POP3, FTP
6. Presentation		Data representation and encryption ASCII, JPEG, MPEG, SSL, TLS, compression
5. Session		Inter-host communication NetBIOS, RPC, NFS
4. Transport	Segments	End-to-end communications and reliability TCP, UDP
3. Network	Packet/Datagram	Path determination & logical addressing (IP) IPv4, IPv6, ICMP, IPSec, ARP
2. Data Link	Bit/Frame	Physical addressing (MAC & LLC) Ethernet, 802.1x, PPP, ATM, Fiber Channel, MAC, MPLS, PPP
1. Physical	Bit	Media, signal and binary transmission Cables, connectors, hubs

A brief description of the seven layers of the OSI model can be found on **Webopedia**.

Network Virtualization

Two commonly used terms related to network virtualization are Software Defined Networking (SDN) and Network Functions Virtualization (NFV).

SDN refers to the ability to control the behavior of network devices programmatically. Usually SDN implementations offer centralized control, separation of control and forwarding functions, and the ability to programmatically control behavior using well-defined interfaces.



NFV is an approach whereby standard compute virtualization technologies are used to host network services that would traditionally run on dedicated proprietary hardware. With NFV, Virtual Machines (VMs) can run network functions such as routing, load balancing and firewalls.

SDN and NFV are considered to be complementary technologies that can be implemented together resulting in virtualized network functions that can be centrally controlled through software.

Virtual Private Cloud (VPC)

An Amazon Virtual Private Cloud (VPC) is an isolated network environment on AWS that is analogous to having a private data center in the cloud.

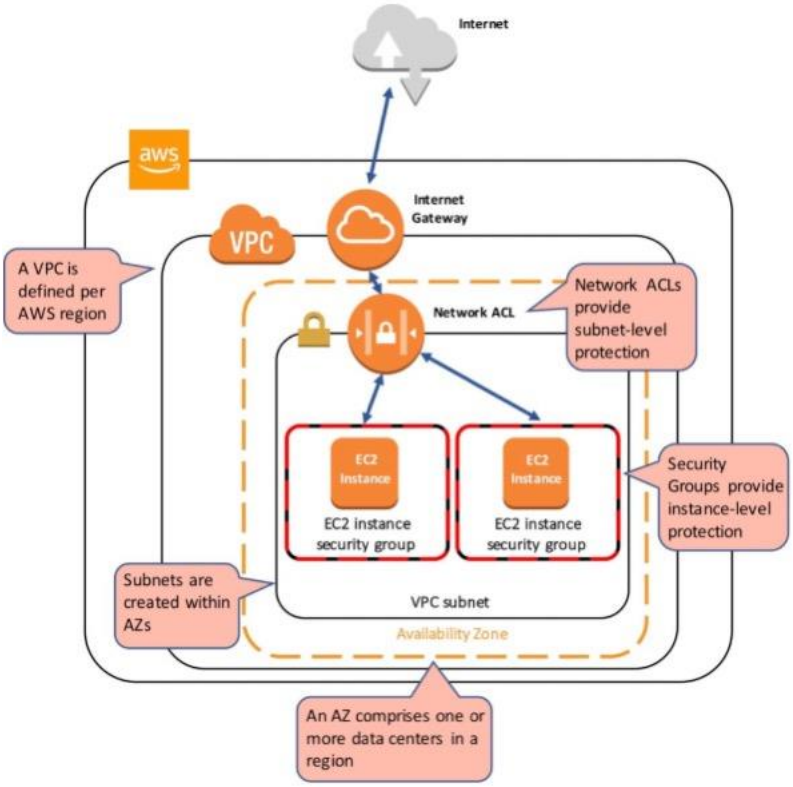
With a **VPC** you can specify your own CIDR address block, create subnets, and configure route tables and gateways. VPC allows the creation of both IPv4 and IPv6 addresses. VPCs are created within AWS regions.

Subnets

A VPC subnet is created within an Availability zone (AZ) which is comprised of one or more data centers within an AWS region. There are two or more AZs in each region and you can create many subnets in each AZ.

Each subnet can be configured as either private or public. With a private subnet instances are only assigned a private IP address (not routable on the Internet) and can only communicate with the outside world by way of a network address translation (NAT) device such as an AWS NAT Gateway.

A public subnet is a subnet in which instances are assigned a public IP address (in addition to a private IP address) and to which an Internet Gateway (IGW) is connected (this is essentially the default gateway for the instances in the subnet).



Route Tables

A route table contains a set of rules, called routes, that are used to determine where network traffic is directed. Each subnet in a VPC must be associated to a route table. A subnet can only be assigned to one route table but a route table can be assigned to multiple subnets.

An "implicit" router is associated with all VPCs and ensures that routing works between all the subnets you create. Each route in a route table specifies a destination CIDR and a target, and the router will use the most specific route that matches the traffic when making forwarding decisions.

Load Balancing

Load balancing is a method of efficiently distributing incoming network traffic across a series of backend servers or targets. With a load balancer you can evenly distribute connections to multiple servers ensuring high availability and reliability as well as providing scalability as the number of requests increases or decreases.

The AWS Elastic Load Balancing service is provided within the Elastic Compute Cloud (EC2) console and there are three different types of ELB available for use with your EC2 instances. These are:

- **Application Load Balancer (ALB)** – layer 7 load balancer that routes connections based on the content of the request
- **Network Load Balancer (NLB)** – layer 4 load balancer that routes connections based on IP protocol data

- **Classic Load Balancer (CLB**) – this is the oldest of the three and provides basic load balancing at both layer 4 and layer 7

VPN and Direct Connect

A virtual private network (VPN) is used to extend a private network across a public or untrusted network. On AWS you can create an IPsec VPN connection between your VPC and your remote network, which could be your company's on-premise data center.

Another option is AWS Direct Connect which is a network service that provides an alternative to using the Internet to connect a customer's on premise sites to AWS. With AWS Direct Connect data is transmitted through a private network connection between AWS and a customer's datacenter or corporate network.

Security

There are a number of tools and services to secure your resources in your VPC. A Security Group is an instance-level virtual firewall that controls inbound and outbound traffic. A Network ACL is a subnet-level firewall controlling traffic in and out of your subnets.

Security Group	Network ACL
Operates at the instance (interface) level	Operates at the subnet level
Supports allow rules only	Supports allow and deny rules
Stateful	Stateless
Evaluates all rules	Processes rules in order
Applies to an instance only if associated with a group	Automatically applies to all instances in the subnets its associated with

The AWS Web Application Firewall (WAF) protects web applications from common web exploits that could affect application availability, compromise security, or consume excessive resources. AWS WAF provides control over which traffic to allow or block to web applications through the definition of customizable web security rules.

AWS Shield is a managed Distributed Denial of Service (DDoS) protection service that safeguards applications running on AWS. AWS Shield offers always-on protection and provides detection and mitigation against sophisticated DDoS attacks at the network, transport and application layers.

CLOUD COMPUTING CONCEPTS

GENERAL CLOUD COMPUTING CONCEPTS

Cloud computing is the on-demand delivery of compute power, database storage, applications and other IT resources through a cloud services platform via the Internet with pay-as-you-go pricing.

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 owns and maintains the network-connected hardware required for these application services, while you provision and use what you need via a web application.

THE SIX ADVANTAGES

AWS promote the six advantages of cloud:

1. Trade capital expense for variable expense
2. Benefit from massive economies of scale
3. Stop guessing about capacity
4. Increase speed and agility
5. Stop spending money running and maintaining data centres
6. Go global in minutes

Trade capital expense for variable expense

Instead of having to invest heavily in data centers and servers before you know how you're going to use them, you can pay only when you consume computing resources, and pay only for how much you consume

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 price.

Stop guessing about capacity

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

Increase speed and agility

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

Stop spending money running and maintaining data centers

Focus on projects that differentiate your business, not the infrastructure. Cloud computing lets you focus on your own customers, rather than on the heavy lifting of racking, stacking, and powering servers.

Go global in minutes

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

TYPES OF CLOUD COMPUTING

There are three main types of cloud computing:

1. Infrastructure as a service (IaaS)
2. Platform as a service (PaaS)
3. Software as a service (SaaS)

Infrastructure as a Service (IaaS)

Infrastructure as a Service (IaaS) contains the basic building blocks for cloud IT and typically provide access to networking features, computers (virtual or on dedicated hardware), and

data storage space. IaaS provides you with the highest level of flexibility and management control over your IT resources and is most similar to existing IT resources that many IT departments and developers are familiar with today.

Platform as a Service (PaaS)

Platform as a Service (PaaS) removes the need for your organization to manage the underlying infrastructure (usually hardware and operating systems) and allows you to focus on the deployment and management of your applications. This helps you be more efficient as you don't need to worry about resource procurement, capacity planning, software maintenance, patching, or any of the other undifferentiated heavy lifting involved in running your application.

Software as a Service (SaaS)

Software as a Service (SaaS) provides you with a completed product that is run and managed by the service provider. In most cases, people referring to Software as a Service are referring to end-user applications. With a SaaS offering you do not have to think about how the service is maintained or how the underlying infrastructure is managed; you only need to think about how you will use that particular piece of software. A common example of a SaaS application is web-based email which you can use to send and receive email without having to manage feature additions to the email product or maintain the servers and operating systems that the email program is running on.

TYPES OF CLOUD DEPLOYMENT

There are three main types of cloud deployment:

1. Public Cloud or simple "Cloud" - e.g. AWS, Azure, GCP
2. Hybrid Cloud - mixture of public and private clouds
3. Private Cloud (on-premise) - managed in your own data centre, e.g. Hyper-V, OpenStack, VMware

Public 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 or can

use higher level services that provide abstraction from the management, architecting, and scaling requirements of core infrastructure.

Hybrid

A hybrid deployment is a way to connect infrastructure and applications between cloud-based resources and existing resources that are not located in the cloud. The most common method of hybrid deployment is between the cloud and existing on-premises infrastructure to extend, and grow, an organization's infrastructure into the cloud while connecting cloud resources to the internal system.

On-premises

The deployment of resources on-premises, using virtualization and resource management tools, is sometimes called the "private cloud." On-premises deployment doesn't provide many of the benefits of cloud computing but is sometimes sought for its ability to provide dedicated resources. In most cases this deployment model is the same as legacy IT infrastructure while using application management and virtualization technologies to try and increase resource utilization.

CLOUD COMPUTING CONCEPTS PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

What advantages do you get from using the AWS cloud? (choose 2)

A. Trade capital expense for variable expense
B. Stop guessing about capacity
C. Increased capital expenditure
D. Gain greater control of the infrastructure layer
E. Comply with all local security compliance programs

Question 2:

What architectural best practice aims to reduce the interdependencies between services?

A. Services, Not Servers
B. Removing Single Points of Failure
C. Automation
D. Loose Coupling

Question 3:

When instantiating compute resources, what are two techniques for using automated, repeatable processes that are fast and avoid human error? (choose 2)

A. Snapshotting
B. Bootstrapping
C. Fault tolerance
D. Infrastructure as code
E. Performance monitoring

Question 4:

What are two ways that moving to an AWS cloud can benefit an organization? (choose 2)

A. Switch to a CAPEX model
B. Increase speed and agility
C. Stop guessing about capacity
D. Depreciate assets over a longer timeframe
E. Gain greater control of data center security

Question 1 answer: A,B

Explanation:

The 6 advantages of cloud are:

1. Trade capital expense for variable expense
2. Benefit from massive economies of scale
3. Stop guessing about capacity
4. Increase speed and agility
5. Stop spending money running and maintaining data centres
6. Go global in minutes

You do not gain greater control of the infrastructure layer as AWS largely control this, and though AWS is compliant with lots of security compliance programs, not all programs in all local countries will be included

Question 2 answer: *D*

Explanation:

- As application complexity increases, a desirable attribute of an IT system is that it can be broken into smaller, loosely coupled components. This means that IT systems should be designed in a way that reduces interdependencies—a change or a failure in one component should not cascade to other components

- The concept of loos coupling includes "well-defined interfaces" which reduce interdependencies in a system by enabling interaction only through specific, technology-agnostic interfaces (e.g. RESTful APIs)

Question 3 answer: *B, D*

Explanation:

- With infrastructure as code AWS assets are programmable, so you can apply techniques, practices, and tools from software development to make your whole infrastructure reusable, maintainable, extensible, and testable
- With bootstrapping you can execute automated actions to modify default configurations. This includes scripts that install software or copy data to bring that resource to a particular state
- Snapshotting is about saving data, not instantiating resources. Fault tolerance is a method of increasing the availability of your system when components fail. Performance monitoring has nothing to do with instantiating resources

Question 4 answer: B,C

Explanation:

- Eliminate guessing on your infrastructure capacity needs. When you make a capacity decision prior to deploying an application, you often end up either sitting on expensive idle resources or dealing with limited capacity. With cloud computing, these problems go away. You can access as much or as little capacity as you need, and scale up and down as required with only a few minutes' notice
- In a cloud computing environment, new IT resources are only a click away, which means that you reduce the time to make those resources available to your developers from weeks to just minutes. This results in a dramatic increase in agility for the organization, since the cost and time it takes to experiment and develop is significantly lower
- Cloud is based on an operational expenditure (OPEX) model, not a capital expenditure (CAPEX) model
- Cloud does not provide the ability to depreciate assets over a longer timeframe as you generally do not own the assets
- Though the AWS cloud does provide significant security standards for the data center, you do not get more control as this is an AWS responsibility

AWS GLOBAL INFRASTRUCTURE

GENERAL

AWS Global Infrastructure is a key technology area covered in the Cloud Practitioner exam blueprint. The AWS Global infrastructure is built around Regions and Availability Zones (AZs)

A Region is a physical location in the world where AWS have multiple AZs

AZs consist of one or more discrete data centers, each with redundant power, networking, and connectivity, housed in separate facilities

Each region is completely independent. Each Availability Zone is isolated, but the Availability Zones in a region are connected through low-latency links

AWS are constantly expanding around the world and currently there are:

- 19 regions
- 55 availability zones

The following diagram shows the AWS global infrastructure with regions (orange circles, green are new regions), and availability zones (the number of AZs is specified within each region):

REGIONS

A region is a geographical area

Each region consists of 2 or more availability zones

Each Amazon Region is designed to be completely isolated from the other Amazon Regions

Each AWS Region has multiple Availability Zones and data centers

You can replicate data within a region and between regions using private or public Internet connections

You retain complete control and ownership over the region in which your data is physically located, making it easy to meet regional compliance and data residency requirements

Note that there is a charge for data transfer between regions

When you launch an EC2 instance, you must select an AMI that's in the same region. If the AMI is in another region, you can copy the AMI to the region you're using

Regions and Endpoints:

- When you work with an instance using the command line interface or API actions, you must specify its regional endpoint
- To reduce data latency in your applications, most Amazon Web Services offer a regional endpoint to make your requests
- An endpoint is a URL that is the entry point for a web service
- For example, https://dynamodb.us-west-2.amazonaws.com is an entry point for the Amazon DynamoDB service

AVAILABILITY ZONES

Availability Zones are physically separate and isolated from each other

AZs span one or more data centers and have direct, low-latency, high throughput and redundant network connections between each other

Each AZ is designed as an independent failure zone

When you launch an instance, you can select an Availability Zone or let AWS choose one for you

If you distribute your EC2 instances across multiple Availability Zones and one instance fails, you can design your application so that an instance in another Availability Zone can handle requests

You can also use Elastic IP addresses to mask the failure of an instance in one Availability Zone by rapidly remapping the address to an instance in another Availability Zone

An Availability Zone is represented by a region code followed by a letter identifier; for example, ***us-east-1a***

To ensure that resources are distributed across the Availability Zones for a region, AWS independently map Availability Zones to names for each AWS account

For example, the Availability Zone ***us-east-1a*** for your AWS account might not be the same location as us-east-1a for another AWS account

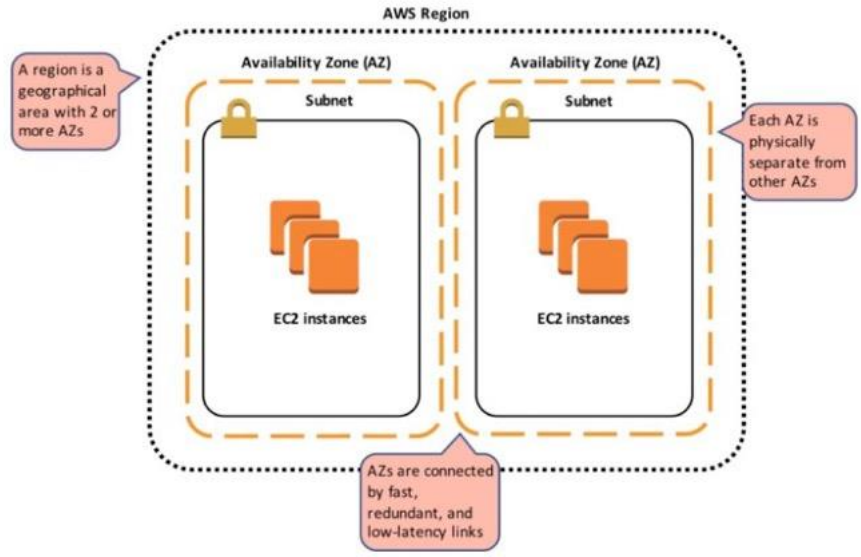
To coordinate Availability Zones across accounts, you must use the *AZ ID*, which is a unique and consistent identifier for an Availability Zone

AZs are physically separated within a typical metropolitan region and are located in lower risk flood plains

AZs use discrete UPS and onsite backup generation facilities and are fed via different grids from independent facilities

AZs are all redundantly connected to multiple tier-1 transit providers

The following diagram depicts a region with 2 availability zones:



EDGE LOCATIONS AND REGIONAL EDGE CACHES

Edge locations are Content Delivery Network (CDN) endpoints for CloudFront

There are many more edge locations than regions

Currently there are over 100 edge locations

Regional Edge Caches sit between your CloudFront Origin servers and the Edge Locations.

A Regional Edge Cache has a larger cache-width than each of the individual Edge Locations

The following diagram shows CloudFront Edge locations and Regional Edge Caches:

AWS GLOBAL INFRASTRUCTURE PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Identify the services that have a global (rather than regional) scope? (choose 2)

A. Amazon Route 53
B. Amazon S3
C. Amazon CloudFront
D. AWS Lambda
E. Amazon EC2

Question 2:

What is an availability zone composed of?

A. One or more regions
B. One or more DCs in a location
C. A collection of edge locations

D. A collection of VPCs

Question 3:

Which of the facts below are accurate in relation to AWS Regions? (choose 2)

A. Each region consists of 2 or more availability zones
B. Each region consists of a collection of VPCs
C. Each region is designed to be completely isolated from the other Amazon Regions
D. Regions have direct, low-latency, high throughput and redundant network connections between each other
E. Regions are Content Delivery Network (CDN) endpoints for CloudFront

Question 4:

Which services are managed at a regional (rather than global) level? (choose 2)

A. Amazon CloudFront
B. Amazon Route 53
C. Amazon S3
D. Amazon EC2
E. AWS IAM

Question 1 answer: A,C

Explanation:

- Amazon Route 53 and Amazon CloudFront have a global scope
- Amazon S3 uses a global namespace but buckets and objects are created within a region
- AWS Lambda is a regional service

Question 2 answer: B

Explanation:

- Availability Zones are physically separate and isolated from each other
- AZ's have direct, low-latency, high throughput and redundant network connections between each other
- A region is a geographical area
- Each region consists of 2 or more availability zones

Question 3 answer: A,C

Explanation:

- A region is not a collection of VPCs, it is composed of at least 2 AZs. VPCs exist within accounts on a per region basis

- Edge locations are (not regions Availability Zones (not regions) have direct, low-latency, high throughput and redundant network connections between each other
- Edge locations are (not regions)) are Content Delivery Network (CDN) endpoints for CloudFront

Question 4 answer: *C,D*

Explanation:

- Both Amazon EC2 and Amazon S3 are managed at a regional level. Note: Amazon S3 is a global namespace but you still create your buckets within a region
- CloudFront, Route 52 and IAM and managed at a global level

IDENTITY AND ACCESS MANAGEMENT

GENERAL

AWS Identity and Access Management (IAM) is a web service that helps you securely control access to AWS resources

You use IAM to control who is authenticated (signed in) and authorized (has permissions) to use resources

IAM makes it easy to provide multiple users secure access to AWS resources

When you first create an AWS account, you begin with a single sign-in 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

IAM can be used to manage:

- Users
- Groups
- Access policies
- Roles
- User credentials
- User password policies
- Multi-factor authentication (MFA)
- API keys for programmatic access (CLI)

Provides centralized control of your AWS account

Enables shared access to your AWS account

IAM provides the following features:

- Shared access to your AWS account
- Granular permissions
- Secure access to AWS resources for application that run on Amazon EC2
- Multi-Factor authentication
- Identity federation
- Identity information for assurance
- PCI DSS compliance
- Integrated with may AWS services
- Eventually consistent
- Free to use

You can work with AWS Identity and Access Management in any of the following ways:

- AWS Management Console
- AWS Command Line Tools
- AWS SDKs
- IAM HTTPS API

By default new users are created with NO access to any AWS services – they can only login to the AWS console

Permission must be explicitly granted to allow a user to access an AWS service

IAM users are individuals who have been granted access to an AWS account

Each IAM user has three main components:

- A user-name
- A password
- Permissions to access various resources

You can apply granular permissions with IAM

You can assign users individual security credentials such as access keys, passwords, and multi-factor authentication devices

IAM is not used for application-level authentication

Identity Federation (including AD, Facebook etc.) can be configured allowing secure access to resources in an AWS account without creating an IAM user account

Multi-factor authentication (MFA) can be enabled/enforced for the AWS account and for individual users under the account

MFA uses an authentication device that continually generates random, six-digit, single-use authentication codes

You can authenticate using an MFA device in the following two ways:

- Through the **AWS Management Console** – the user is prompted for a user name, password and authentication code
- Using the **AWS API** – restrictions are added to IAM policies and developers can request temporary security credentials and pass MFA parameters in their AWS STS API requests
- Using the **AWS CLI** by obtaining temporary security credentials from STS (aws sts get-session-token)

It is a best practice to always setup multi-factor authentication on the root account

IAM is universal (global) and does not apply to regions

IAM is eventually consistent

IAM replicates data across multiple data centres around the world

The "root account" is the account created when you setup the AWS account. It has complete Admin access and is the only account that has this access by default

It is a best practice to not use the root account for anything other than billing

Power user access allows all permissions except the management of groups and users in IAM

Temporary security credentials consist of the AWS access key ID, secret access key, and security token

IAM can assign temporary security credentials to provide users with temporary access to services/resources

To sign-in you must provide your account ID or account alias in addition to a user name and password

The sign-in URL includes the account ID or account alias, e.g.:

https://***My_AWS_Account_ID***.signin.aws.amazon.com/console/

Alternatively you can sign-in at the following URL and enter your account ID or alias manually:

https://console.aws.amazon.com/

IAM integrates with many different AWS services

IAM supports PCI DSS compliance

AWS recommend that you use the AWS SDKs to make programmatic API calls to IAM

However, you can also use the IAM Query API to make direct calls to the IAM web service

AUTHENTICATION METHODS

Console password:

- A password that the user can enter to sign in to interactive sessions such as the AWS Management Console
- You can allow users to change their own passwords
- You can allow selected IAM users to change their passwords by disabling the option for all users and using an IAM policy to grant permissions for the selected users

Access Keys:

- A combination of an **access key ID** and a **secret access key**

- You can assign two active access keys to a user at a time
- These can be used to make programmatic calls to AWS when using the **API** in program code or at a command prompt when using the **AWS CLI** or the **AWS PowerShell** tools
- You can create, modify, view or rotate access keys
- When created IAM returns the access key ID and secret access key
- The secret access is returned only at creation time and if lost a new key must be created
- Ensure access keys and secret access keys are stored securely
- Users can be given access to change their own keys through IAM policy (not from the console)
- You can disable a user's access key which prevents it from being used for API calls

Server certificates:

- SSL/TLS certificates that you can use to authenticate with some AWS services
- AWS recommends that you use the AWS Certificate Manager (ACM) to provision, manage and deploy your server certificates
- Use IAM only when you must support HTTPS connections in a region that is not supported by ACM

IAM USERS

An IAM user is an entity that represents a person or service

Can be assigned:

- An access key ID and secret access key for programmatic access to the AWS API, CLI, SDK, and other development tools
- A password for access to the management console

By default users cannot access anything in your account

The account root user credentials are the email address used to create the account and a password

The root account has full administrative permissions and these cannot be restricted

Best practice for root accounts:

- Don't use the root user credentials
- Don't share the root user credentials
- Create an IAM user and assign administrative permissions as required
- Enable MFA

IAM users can be created to represent applications and these are known as "service accounts"

You can have up to 5000 users per AWS account

Each user account has a friendly name and an ARN which uniquely identifies the user across AWS

A unique ID is also created which is returned only when you create the user using the API, Tools for Windows PowerShell or the AWS CLI

You should create individual IAM accounts for users (best practice not to share accounts)

The Access Key ID and Secret Access Key are not the same as a password and cannot be used to login to the AWS console

The Access Key ID and Secret Access Key can only be used once and must be regenerated if lost

A password policy can be defined for enforcing password length, complexity etc. (applies to all users)

You can allow or disallow the ability to change passwords using an IAM policy

Access keys and passwords should be changed regularly

GROUPS

Groups are collections of users and have policies attached to them

A group is not an identity and cannot be identified as a principal in an IAM policy

Use groups to assign permissions to users

Use the principal of least privilege when assigning permissions

You cannot nest groups (groups within groups)

ROLES

Roles are created and then "assumed" by trusted entities and define a set of permissions for making AWS service requests

With IAM Roles you can delegate permissions to resources for users and services without using permanent credentials (e.g. user name and password)

IAM users or AWS services can assume a role to obtain temporary security credentials that can be used to make AWS API calls

You can delegate using roles

There are no credentials associated with a role (password or access keys)

IAM users can temporarily assume a role to take on permissions for a specific task

A role can be assigned to a federated user who signs in using an external identity provider

Temporary credentials are primarily used with IAM roles and automatically expire

Roles can be assumed temporarily through the console or programmatically with the **AWS CLI**, **Tools for Windows PowerShell** or **API**

IAM roles with EC2 instances:

- IAM roles can be used for granting applications running on EC2 instances permissions to AWS API requests using instance profiles
- Only one role can be assigned to an EC2 instance at a time
- A role can be assigned at the **EC2 instance creation time or at any time afterwards**
- When using the AWS CLI or API instance profiles must be created manually (it's automatic and transparent through the console)
- Applications retrieve temporary security credentials from the instance metadata

Role Delegation:

- Create an IAM role with two policies:
  - Permissions policy – grants the user of the role the required permissions on a resource
  - Trust policy – specifies the trusted accounts that are allowed to assume the role
- Wildcards (*) cannot be specified as a principal
- A permissions policy must also be attached to the user in the trusted account

POLICIES

Policies are documents that define permissions and can be applied to users, groups and roles

Policy documents are written in JSON (key value pair that consists of an attribute and a value)

All permissions are implicitly denied by default

The most restrictive policy is applied

The IAM policy simulator is a tool to help you understand, test, and validate the effects of access control policies

The Condition element can be used to apply further conditional logic

AWS STS

The AWS Security Token Service (STS) is a web service that enables you to request temporary, limited-privilege credentials for IAM users or for users that you authenticate (federated users)

By default, AWS STS is available as a global service, and all AWS STS requests go to a single endpoint at https://sts.amazonaws.com

You can optionally send your AWS STS requests to endpoints in any region (can reduce latency)

All regions are enabled for STS by default but can be disabled

The region in which temporary credentials are requested must be enabled

Credentials will always work globally

STS supports AWS CloudTrail, which records AWS calls for your AWS account and delivers log files to an S3 bucket

Temporary security credentials work almost identically to long-term access key credentials that IAM users can use, with the following differences:

- Temporary security credentials are short-term
- They can be configured to last anywhere from a few minutes to several hours
- After the credentials expire, AWS no longer recognizes them or allows any kind of access to API requests made with them
- Temporary security credentials are not stored with the user but are generated dynamically and provided to the user when requested
- When (or even before) the temporary security credentials expire, the user can request new credentials, as long as the user requesting them still has permission to do so

Advantages of STS are:

- You do not have to distribute or embed long-term AWS security credentials with an application
- You can provide access to your AWS resources to users without having to define an AWS identity for them (temporary security credentials are the basis for IAM Roles and ID Federation)
- The temporary security credentials have a limited lifetime, so you do not have to rotate them or explicitly revoke them when they're no longer needed
- After temporary security credentials expire, they cannot be reused (you can specify how long the credentials are valid for, up to a maximum limit)

The AWS STS API action returns temporary security credentials that consist of:

- An access key which consists of an access key ID and a secret ID
- A session token

- Expiration or duration of validity
- Users (or an application that the user runs) can use these credentials to access your resources

With STS you can request a session token using one of the following APIs:

- AssumeRole – can only be used by IAM users (can be used for MFA)
- AssumeRoleWithSAML – can be used by any user who passes a SAML authentication response that indicates authentication from a known (trusted) identity provider
- AssumeRoleWithWebIdentity – can be used by a user who passes a web identity token that indicates authentication from a known (trusted) identity provider
- GetSessionToken – can be used by an IAM user or AWS account root user (can be used for MFA)
- GetFederationToken – can be used by an IAM user or AWS account root user

AWS recommends using Cognito for identity federation with Internet identity providers

Users can come from three sources

Federation (typically AD):

- Uses SAML 2.0
- Grants temporary access based on the users AD credentials
- Does not need to be a user in IAM
- Single sign-on allows users to login to the AWS console without assigning IAM credentials

Federation with Mobile Apps:

- Use Facebook/Amazon/Google or other OpenID providers to login

Cross Account Access:

- Lets users from one AWS account access resources in another
- To make a request in a different account the resource in that account must have an attached resource-based policy with the permissions you need
- Or you must assume a role (identity-based policy) within that account with the permissions you need

There are a couple of ways STS can be used

Scenario 1:

1. Develop an Identity Broker to communicate with LDAP and AWS STS
2. Identity Broker always authenticates with LDAP first, then with AWS STS
3. Application then gets temporary access to AWS resources

Scenario 2:

1. Develop an Identity Broker to communicate with LDAP and AWS STS
2. Identity Broker authenticates with LDAP first, then gets an IAM role associated with the user
3. Application then authenticates with STS and assumes that IAM role
4. Application uses that IAM role to interact with the service

IAM BEST PRACTICES

Lock away the AWS root user access keys

Create individual IAM users

Use AWS defined policies to assign permissions whenever possible

Use groups to assign permissions to IAM users

Grant least privilege

Use access levels to review IAM permissions

Configure a strong password policy for users

Enable MFA for privileged users

Use roles for applications that run on AWS EC2 instances

Delegate by using roles instead of sharing credentials

Rotate credentials regularly

Remove unnecessary credentials

Use policy conditions for extra security

Monitor activity in your AWS account

IDENTITY AND ACCESS MANAGEMENT PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Which file format is used to write AWS Identity and Access Management (IAM) policies?

A. DOC
B. XML
C. JBOD
D. JSON

Question 2:

To ensure the security of your AWS account, what are two AWS best practices for managing access keys? (choose 2)

A. Don't create any access keys, use IAM roles instead
B. Don't generate an access key for the root account user
C. Where possible, use IAM roles with temporary security credentials
D. Rotate access keys daily
E. Use MFA for access keys

Question 3:

When using Amazon IAM, what authentication methods are available to use? (choose 2)

A. Client certificates
B. Access keys
C. Amazon KMS
D. Server certificates
E. AES 256

Question 1 answer: D

Explanation:

- You manage access in AWS by creating policies and attaching them to IAM identities or AWS resources. A policy is an object in AWS that, when associated with an entity or resource, defines their permissions.
- AWS evaluates these policies when a principal, such as a user, makes a request. Permissions in the policies determine whether the request is allowed or denied. Most policies are stored in AWS as JSON documents

Question 2 answer: B,C

Explanation:

- Best practices include:
  - Don't generate an access key for the root account user
  - Use Temporary Security Credentials (IAM Roles) Instead of Long-Term Access Keys
  - Manage IAM User Access Keys Properly

- Rotating access keys is a recommended practice, but doing it daily would be excessive and hard to manage
- You can use MFA for securing privileged accounts, but it does not secure access keys
- You should use IAM roles where possible, but AWS do not recommend that you don't create any access keys as they also have a purpose

Question 3 answer: *B,D*

Explanation:

- Supported authentication methods include console passwords, access keys and server certificates
- Access keys are a combination of an access key ID and a secret access key and can be used to make programmatic calls to AWS
- Server certificates are SSL/TLS certificates that you can use to authenticate with some AWS services
- Client certificates are not a valid IAM authentication method
- Amazon Key Management Service (KMS) is used for managing encryption keys and is not used for authentication
- AES 256 is an encryption algorithm, not an authentication method

AWS COMPUTE

AMAZON EC2

Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud

The EC2 simple web service interface allows you to obtain and configure capacity with minimal friction

EC2 is designed to make web-scale cloud computing easier for developers

Amazon EC2's simple web service interface allows you to obtain and configure capacity with minimal friction

It provides you with complete control of your computing resources and lets you run on Amazon's proven computing environment

Amazon EC2 reduces the time required to obtain and boot new server instances to minutes, allowing you to quickly scale capacity, both up and down, as your computing requirements change

Amazon EC2 changes the economics of computing by allowing you to pay only for capacity that you actually use

Amazon EC2 provides developers the tools to build failure resilient applications and isolate them from common failure scenarios

Benefits of EC2 include:

- **Elastic Web-Scale computing** – you can increase or decrease capacity within minutes not hours and commission one to thousands of instances simultaneously
- **Completely controlled** – You have complete control include root access to each instance and can stop and start instances without losing data and using web service APIs
- **Flexible Cloud Hosting Services** – you can choose from multiple instance types, operating systems, and software packages as well as instances with varying memory, CPU and storage configurations
- **Integrated** – EC2 is integrated with most AWS services such as S3, RDS, and VPC to provide a complete, secure solution
- **Reliable** – EC2 offers a highly reliable environment where replacement instances can be rapidly and predictably commissioned with SLAs of 99.95% for each region
- **Secure** – EC2 works in conjunction with VPC to provide a secure location with an IP address range you specify and offers Security Groups, Network ACLs, and IPSec VPN features
- **Inexpensive** – Amazon passes on the financial benefits of scale by charging very low rates and on a capacity consumed basis

An Amazon Machine Image (AMI) is a special type of virtual appliance that is used to create a virtual machine within the Amazon Elastic Compute Cloud ("EC2")

AMIs come in three main categories:

- **Community AMIs** – free to use, generally you just select the operating system you want
- **AWS Marketplace AMIs** – pay to use, generally come packaged with additional, licensed software
- **My AMIs** – AMIs that you create yourself

Metadata and User Data:

- User data is data that is supplied by the user at instance launch in the form of a script
- Instance metadata is data about your instance that you can use to configure or manage the running instance
- User data is limited to 16KB
- User data and metadata are not encrypted
- Instance metadata is available at http://169.254.169.254/latest/meta-data
- The Instance Metadata Query tool allows you to query the instance metadata without having to type out the full URI or category names

Pricing

On-demand:

- Good for users that want the low cost and flexibility of 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 EC2 for the first time

Reserved:

- Applications with steady state or predictable usage
- Applications that require reserved capacity
- Users can make up-front payments to reduce their total computing costs even further
- Standard Reserved Instances (RIs) provide up to 75% off on-demand price
- Convertible RIs provide up to 54% off on-demand price – provides the capability to change the attributes of the RI as long as the exchange results in the creation of RIs of equal or greater value
- Scheduled RIs are available to launch within the time window you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month

Spot:

- Applications that have flexible start and end times
- Applications that are only feasible at very low compute prices
- Users with an urgent need for a large amount of additional compute capacity
- If Amazon terminate your instances you do not pay, if you terminate you pay for the hour

Dedicated hosts:

- Useful for regulatory requirements that may not support multi-tenant virtualization
- Great for licensing which does not support multi-tenancy or cloud deployments
- Can be purchased on-demand (hourly)
- Can be purchased as a reservation for up to 70% off the on-demand price

The following table shows the current list of EC2 instance types available:

Family	Hint	Purpose/Design
D	DATA	Heavy data usage (e.g. file servers, DWs)
R	RAM	Memory optimised
M	MAIN	General purpose (e.g. app servers)
C	COMPUTE	Compute optimised
G	GRAPHICS	Graphics intensive workloads
I	IOPS	Storage I/O optimised (e.g. NoSQL, DWs)
F	FAST	FPGA hardware acceleration for applications
T	CHEAP (think T2)	Lowest cost (e.g. T2-micro)
P	GPU	GPU requirements
X	EXTREME RAM	Heavy memory usage (e.g. SAP HANA, Apache Spark)
U	HIGH MEMORY	High memory and bare metal performance – use for in memory DBs including SAP HANA
Z	HGH COMPUTE & MEMORY	Fast CPU, high memory and NVMe-based SSDs – use when high overall performance is required
H	HIGH DISK THROUGHPUT	Up to 16 TB of HDD-based local storage

AMAZON EC2 CONTAINER SERVICE (ECS)

Amazon Elastic Container Service (ECS) is a highly scalable, high performance container management service that supports Docker containers and allows you to easily run applications on a managed cluster of Amazon EC2 instances

Amazon ECS eliminates the need for you to install, operate, and scale your own cluster management infrastructure

Using API calls you can launch and stop container-enabled applications, query the complete state of clusters, and access many familiar features like security groups, Elastic Load Balancing, EBS volumes and IAM roles

Amazon ECS can be used to schedule the placement of containers across clusters based on resource needs and availability requirements

The EC2 container registry (ECR) is a managed AWS Docker registry service for storing, managing and deploying Docker images

There is no additional charge for Amazon ECS. You pay for AWS resources (e.g. EC2 instances or EBS volumes) you create to store and run your application

Amazon ECR is integrated with Amazon EC2 Container Service (ECS)

With Amazon ECR, there are no upfront fees or commitments. You pay only for the amount of data you store in your repositories and data transferred to the Internet

AWS LAMBDA

Lambda is a serverless computing technology that allows you to run code without provisioning or managing servers

Lambda executes code only when needed and scales automatically

You pay only for the compute time you consume (you pay nothing when your code is not running)

Benefits of Lambda:

- No servers to manage
- Continuous scaling
- Subsecond metering
- Integrates with almost all other AWS services

Primary use cases for Lambda:

- Data processing
- Real-time file processing
- Real-time stream processing
- Build serverless backends for web, mobile, IOT, and 3rd party API requests

AMAZON LIGHTSAIL

Instances

LightSail provides developers compute, storage, and networking capacity and capabilities to deploy and manage websites, web applications, and databases in the cloud

LightSail includes everything you need to launch your project quickly – a virtual machine, SSD-based storage, data transfer, DNS management, and a static IP

LightSail provides preconfigured virtual private servers (instances) that include everything required to deploy and application or create a database

The underlying infrastructure and operating system is managed by LightSail

Best suited to projects that require a few dozen instances or fewer

Provides a simple management interface

Good for blogs, websites, web applications, e-commerce etc.

Can deploy load balancers and attach block storage

Public API

Limited to 20 LightSail instances, 5 static IPs, 3 DNS zones, 20 TB block storage, 40 databases, and 5 load balancers per account

Up to 20 certificates per calendar year

Can connect to each other and other AWS resources through public Internet and private (VPC peering) networking

Application templates include WordPress, WordPress Multisite, Drupal, Joomla!, Magento, Redmine, LAMP, Nginx (LEMP), MEAN, Node.js, and more

LightSail currently supports 6 Linux or Unix-like distributions: Amazon Linux, CentOS, Debian, FreeBSD, OpenSUSE, and Ubuntu, as well as 2 Windows Server versions: 2012 R2 and 2016

Databases

LightSail databases are instances that are dedicated to running databases

A LightSail database can contain multiple user-created databases, and you can access it by using the same tools and applications that you use with a stand-alone database

LightSail managed databases provide an easy, low maintenance way to store your data in the cloud

LightSail manages a range of maintenance activities and security for your database and its underlying infrastructure

LightSail automatically backs up your database and allows point in time restore from the past 7 days using the database restore tool

LightSail databases support the latest major versions of MySQL. Currently, these versions are 5.6, 5.7, and 8.0 for MySQL

LightSail databases are available in Standard and High Availability plans

High Availability plans add redundancy and durability to your database, by automatically creating standby database in a separate Availability Zone

LightSail is very affordable

LightSail plans are billed on an on-demand hourly rate, so you pay only for what you use

For every LightSail plan you use, we charge you the fixed hourly price, up to the maximum monthly plan cost

AWS COMPUTE PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Which AWS service can you use to install a third-party database?

A. Amazon RDS

B. Amazon DynamoDB

C. Amazon EC2

D. Amazon EMR

Question 2:

Which service can you use to provision a preconfigured server with little to no AWS experience?

A. Amazon Elastic Beanstalk

B. AWS Lambda

C. Amazon EC2

D. Amazon Lightsail

Question 3:

Which AWS service provides elastic web-scale cloud computing allowing you to deploy operating system instances?

A. Amazon EBS

B. AWS Lambda

C. Amazon RDS

D. Amazon EC2

Question 4:

Which of the following are valid types of Reserved Instance? (choose 2)

A. Convertible RI

B. Discounted RI

C. Scheduled RI

D. Long-Term RI

E. Special RI

Question 1 answer: C

Explanation:

All of these services are managed services except for Amazon EC2. EC2 is the only service in the list upon which you can manually install the database software of your choice

Question 2 answer: D

Explanation:

- Lightsail provides developers compute, storage, and networking capacity and capabilities to deploy and manage websites, web applications, and databases in the cloud
- Lightsail provides preconfigured virtual private servers (instances) that include everything required to deploy and application or create a database
- Deploying a server on Lightsail is extremely easy and does not require knowledge of how to configure VPCs, security groups, network ACLs etc.
- AWS Elastic Beanstalk can be used to quickly deploy and manage applications in the AWS Cloud. It is considered a PaaS service. However, you do still need to deploy within a VPC so more AWS expertise is required
- Amazon EC2 also requires AWS expertise as it deploys within a VPC
- AWS Lambda provides serverless functions not preconfigured servers

Question 3 answer: *D*

Explanation:

- Amazon EC2 provides elastic web-scale computing in the cloud allowing you to deploy Windows and Linux
- AWS Lambda lets you run code without provisioning or managing server operating systems
- Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2instances in the AWS Cloud
- Amazon Relational Database Service (Amazon RDS) makes it easy to set up, operate, and scale a relational database in the cloud

Question 4 answer: *A,C*

Explanation:

- **Standard RIs**: These provide the most significant discount (up to 75% off On-Demand) and are best suited for steady-state usage
- **Convertible RIs**: These provide a discount (up to 54% off On-Demand) and the capability to change the attributes of the RI as long as the exchange results in the creation of Reserved Instances of equal or greater value. Like Standard RIs, Convertible RIs are best suited for steady-state usage
- **Scheduled RIs**: These are available to launch within the time windows you reserve. This option allows you to match your capacity reservation to a predictable recurring schedule that only requires a fraction of a day, a week, or a month

AWS STORAGE

AMAZON SIMPLE STORAGE SERVICE (S3)

Amazon S3 is object storage built to store and retrieve any amount of data from anywhere – web sites and mobile apps, corporate applications, and data from IoT sensors or devices

You can store any type of file in S3

S3 is designed to deliver 99.999999999% durability, and stores data for millions of applications used by market leaders in every industry

S3 provides comprehensive security and compliance capabilities that meet even the most stringent regulatory requirements.

S3 gives customers flexibility in the way they manage data for cost optimization, access control, and compliance

S3 provides query-in-place functionality, allowing you to run powerful analytics directly on your data at rest in S3. And Amazon S3 is the most supported cloud storage service available, with integration from the largest community of third-party solutions, systems integrator partners, and other AWS services.

Files can be anywhere from 0 bytes to 5 TB

There is unlimited storage available

Files are stored in buckets

Buckets are root level folders

Any subfolder within a bucket is known as a “folder”

S3 is a universal namespace so bucket names must be unique globally

When you successfully upload a file to S3 you receive a HTTP 200 code

Bucket names must follow a set of rules:

- Names must be unique across all of AWS
- Names must be 3 to 63 characters in length
- Names can only contain lowercase letters, numbers and hyphens
- Names cannot be formatted as an IP address

Data consistency:

- Read after write consistency for PUTS of new objects
- Eventual consistency for overwrite PUTS and DELETES (takes time to propagate)

Objects consist of:

- Key (name of the object)
- Value (data made up of a sequence of bytes)
- Version ID (used for versioning)
- Metadata (data about the data that is stored)

Subresources:

- Access control lists
- Torrent

Built for 99.99 availability

SLA is 99.9% availability

Amazon guarantee 99.99999999% durability

Object sharing – the ability to make any object publicly available via a URL

Lifecycle management – set rules to transfer objects between storage classes at defined time intervals

Versioning – automatically keep multiple versions of an object (when enabled)

Encryption

Data secured using ACLs and bucket policies

Tiers:

- S3 standard
- S3-IA
- S3 One Zone – IA
- Glacier

Charges:

- Storage
- Requests
- Storage management pricing
- Data transfer pricing
- Transfer acceleration

When you create a bucket you need to select the region where it will be created

It is a best practice to create buckets in regions that are physically closest to your users to reduce latency

AWS SNOWBALL

With AWS Snowball (Snowball), you can transfer hundreds of terabytes or petabytes of data between your on-premises data centers and Amazon Simple Storage Service (Amazon S3)

Uses a secure storage device for physical transportation

AWS Snowball Client is software that is installed on a local computer and is used to identify, compress, encrypt, and transfer data

Uses 256-bit encryption (managed with the AWS KMS) and tamper-resistant enclosures with TPM

Snowball (80TB) (50TB model available only in the USA)

Snowball Edge (100TB) comes with onboard storage and compute capabilities

Snowmobile – exabyte scale with up to 100PB per Snowmobile

Snowball can import to S3 or export from S3

Import/export is when you send your own disks into AWS – this is being deprecated in favour of Snowball

Snowball must be ordered from and returned to the same region

To speed up data transfer it is recommended to run simultaneous instances of the AWS Snowball Client in multiple terminals and transfer small files as batches

AMAZON ELASTIC BLOCK STORE (EBS)

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

Amazon EBS volumes offer the consistent and low-latency performance needed to run your workloads. With Amazon EBS, you can scale your usage up or down within minutes – all while paying a low price for only what you provision

The following table shows a comparison of a few EBS volume types:

	Solid State Drives (SSD)		Hard Disk Drives (HDD)	
Volume Type	General Purpose SSD (gp2)	Provisioned IOPS SSD (io1)	Throughput Optimized (st1)	Cold HDD (sc1)
Description	Balance of price to performance	High performance SSD	Low cost HDD	Lowest cost HDD
Use Cases	• Most workloads • System boot volumes • Virtual desktops	• Critical business apps that require sustained IOPS performance • Apps that require more than 10,000 IOPS or 160 MiB/s • Large database workloads	• Streaming workloads with fast throughput • Low price • Big data • Data warehouses	• Throughput oriented storage for large volumes of infrequently accessed data • Lowest cost • Cannot be a boot volume
Volume Size	1 GiB - 16 TiB	4 GiB - 16 TiB	500 GiB - 16 TiB	500 GiB - 16 TiB
Max IOPS Per Volume	10,000	32,000	500	250
Max Throughput Per Volume	160 MiB/s	500 MiB/s	500 MiB/s	250 MiB/s

EBS volume data persists independently of the life of the instance

EBS volumes do not need to be attached to an instance

You can attach multiple EBS volumes to an instance

You cannot attach an EBS volume to multiple instances (use Elastic File Store instead)

EBS volumes must be in the same AZ as the instances they are attached to

Termination protection is turned off by default and must be manually enabled (keeps the volume/data when the instance is terminated)

Root EBS volumes are deleted on termination by default

Extra non-boot volumes are not deleted on termination by default

The behaviour can be changed by altering the “DeleteOnTermination” attribute

EBS Snapshots:

- Snapshots capture a point-in-time state of an instance
- Snapshots are stored on S3
- Does not provide granular backup (not a replacement for backup software)
- If you make periodic snapshots of a volume, the snapshots are incremental, which means that only the blocks on the device that have changed after your last snapshot are saved in the new snapshot
- Even though snapshots are saved incrementally, the snapshot deletion process is designed so that you need to retain only the most recent snapshot in order to restore the volume
- Snapshots can only be accessed through the EC2 APIs
- EBS volumes are AZ specific but snapshots are region specific

AMAZON ELASTIC FILE SERVICE (EFS)

EFS is a fully-managed service that makes it easy to set up and scale file storage in the Amazon Cloud

Good for big data and analytics, media processing workflows, content management, web serving, home directories etc.

EFS uses the NFSv4.1 protocol

Pay for what you use (no pre-provisioning required)

Can scale up to petabytes

EFS is elastic and grows and shrinks as you add and remove data

Can concurrently connect 1 to 1000s of EC2 instances, from multiple AZs

A file system can be accessed concurrently from all AZs in the region where it is located

By default you can create up to 10 file systems per account

On-premises access can be enabled via Direct Connect or AWS VPN

Can choose General Purpose or Max I/O (both SSD)

The VPC of the connecting instance must have DNS hostnames enabled

EFS provides a file system interface, file system access semantics (such as strong consistency and file locking)

Data is stored across multiple AZ's within a region

Read after write consistency

Need to create mount targets and choose AZ's to include (recommended to include all AZ's)

Limited region support currently

Instances can be behind an ELB

There are two performance modes:

- "General Purpose" performance mode is appropriate for most file systems
- "Max I/O" performance mode is optimized for applications where tens, hundreds, or thousands of EC2 instances are accessing the file system

Amazon EFS is designed to burst to allow high throughput levels for periods of time

AWS STORAGE PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Which Amazon S3 storage tier provides does not include a data retrieval fee and has an availability SLA of 99.99%?

A. S3 Standard

B. S3 Standard-IA

C. S3 One Zone-IA

D. Amazon Glacier

Question2:

Which AWS service can be used to host a static website?

A. Amazon S3

B. Amazon EBS

C. AWS Lambda

D. Amazon EFS

Question3:

Which feature enables fast, easy, and secure transfers of files over long distances between a client and an Amazon S3 bucket?

A. S3 Static Websites

B. S3 Copy

C. Multipart Upload

D. S3 Transfer Acceleration

Question 4:

How is data protected by default in Amazon S3?

A. Buckets are replicated across all regions

B. Objects are redundantly stored on multiple devices across multiple facilities within a region

C. Objects are redundantly stored on multiple devices across multiple facilities across all regions

D. Objects are copied across at least two Availability Zones per region

Question 1 answer: *A*

Explanation:

- All of the storage tiers listed include a data retrieval fee except for S3 Standard
- Availability SLAs are: S3 Standard = 99.99%; S3 Standard-IA = 99.9%; S3 One Zone-IA = 99%; Amazon Glacier = no SLA

Question 2 answer: *A*

Explanation:

- Amazon S3 can be used to host static websites. It is not possible to use dynamic content. You can use a custom domain name if you configure the bucket name to match
- The other services listed cannot be used to host a static website

Question 3 answer: *D*

Explanation:

- Amazon S3 Transfer Acceleration enables fast, easy, and secure transfers of files over long distances between your client and your Amazon S3 bucket. S3 Transfer Acceleration leverages Amazon CloudFront's globally distributed AWS Edge Locations
- With S3 copy you can create a copy of objects up to 5GB in size in a single atomic operation
- Multipart upload can be used to speed up uploads to S3
- S3 can also be used to host static websites

Question 4 answer: *B*

Explanation:

- Amazon S3 provides a highly durable storage infrastructure designed for mission-critical and primary data storage. Objects are redundantly stored on multiple devices across multiple facilities in an Amazon S3 region
- Amazon does not specify how data is replicated across AZs, the use the term facilities instead

AWS NETWORKING

GENERAL AMAZON VIRTUAL PRIVATE CLOUD (VPC)

A virtual private cloud (VPC) is a virtual network dedicated to your AWS account

Analogous to having your own DC inside AWS

It is logically isolated from other virtual networks in the AWS Cloud

Provides complete control over the virtual networking environment including selection of IP ranges, creation of subnets, and configuration of route tables and gateways

You can launch your AWS resources, such as Amazon EC2 instances, into your VPC

When you create a VPC, you must specify a range of IPv4 addresses for the VPC in the form of a Classless Inter-Domain Routing (CIDR) block; for example, 10.0.0.0/16

This is the primary CIDR block for your VPC

A VPC spans all the Availability Zones in the region

You have full control over who has access to the AWS resources inside your VPC

You can create your own IP address ranges, and create subnets, route tables and network gateways

When you first create your AWS account a default VPC is created for you in each AWS region

A default VPC is created in each region with a subnet in each AZ

By default you can create up to 5 VPCs per region

You can define dedicated tenancy for a VPC to ensure instances are launched on dedicated hardware (overrides the configuration specified at launch)

A default VPC is automatically created for each AWS account the first time Amazon EC2 resources are provisioned

The default VPC has all-public subnets

Public subnets are subnets that have:

- “Auto-assign public IPv4 address” set to “Yes”
- The subnet route table has an attached Internet Gateway

Instances in the default VPC always have both a public and private IP address

AZs names are mapped to different zones for different users (i.e. the AZ "ap-southeast-2a" may map to a different physical zone for a different user)

Components of a VPC:

- **A Virtual Private Cloud**: A logically isolated virtual network in the AWS cloud. You define a VPC's IP address space from ranges you select
- **Subnet**: A segment of a VPC's IP address range where you can place groups of isolated resources (maps to an AZ, 1:1)
- **Internet Gateway**: The Amazon VPC side of a connection to the public Internet
- NAT Gateway: A highly available, managed Network Address Translation (NAT) service for your resources in a private subnet to access the Internet
- **Hardware VPN Connection**: A hardware-based VPN connection between your Amazon VPC and your datacenter, home network, or co-location facility
- **Virtual Private Gateway**: The Amazon VPC side of a VPN connection
- **Customer Gateway**: Your side of a VPN connection
- **Router**: Routers interconnect subnets and direct traffic between Internet gateways, virtual private gateways, NAT gateways, and subnets
- **Peering Connection**: A peering connection enables you to route traffic via private IP addresses between two peered VPCs
- **VPC Endpoints**: Enables private connectivity to services hosted in AWS, from within your VPC without using an Internet Gateway, VPN, Network Address Translation (NAT) devices, or firewall proxies
- **Egress-only Internet Gateway**: A stateful gateway to provide egress only access for IPv6 traffic from the VPC to the Internet

Options for securely connecting to a VPC are:

- AWS managed VPN – fast to setup
- Direct Connect – high bandwidth, low-latency but takes weeks to months to setup
- VPN CloudHub – used for connecting multiple sites to AWS
- Software VPN – use 3rd party software

An Elastic Network Interface (ENI) is a logical networking component that represents a NIC

ENIs can be attached and detached from EC2 instances and the configuration of the ENI will be maintained

Flow Logs capture information about the IP traffic going to and from network interfaces in a VPC

Flow log data is stored using Amazon CloudWatch Logs

Flow logs can be created at the following levels:

- VPC
- Subnet
- Network interface

Peering connections can be created with VPCs in different regions (available in most regions now)

Data sent between VPCs in different regions is encrypted (traffic charges apply)

SUBNETS

After creating a VPC, you can add one or more subnets in each Availability Zone

When you create a subnet, you specify the CIDR block for the subnet, which is a subset of the VPC CIDR block

Each subnet must reside entirely within one Availability Zone and cannot span zones

Types of subnet:

- If a subnet's traffic is routed to an internet gateway, the subnet is known as a public subnet
- If a subnet doesn't have a route to the internet gateway, the subnet is known as a private subnet
- If a subnet doesn't have a route to the internet gateway, but has its traffic routed to a virtual private gateway for a VPN connection, the subnet is known as a VPN-only subnet

An Internet Gateway is a horizontally scaled, redundant, and highly available VPC component that allows communication between instances in your VPC and the internet

FIREWALLS

Network Access Control Lists (ACLs) provide a firewall/security layer at the subnet level

Security Groups provide a firewall/security layer at the instance level

The table below describes some differences between Security Groups and Network ACLs:

Security Group	Network ACL
Operates at the instance (interface) level	Operates at the subnet level
Supports allow rules only	Supports allow and deny rules
Stateful	Stateless
Evaluates all rules	Processes rules in order
Applies to an instance only if associated with a group	Automatically applies to all instances in the subnets its associated with

VPC WIZARD

The VPC Wizard can be used to create the following four configurations:

VPC with a Single Public Subnet:

- Your instances run in a private, isolated section of the AWS cloud with direct access to the Internet
- Network access control lists and security groups can be used to provide strict control over inbound and outbound network traffic to your instances
- Creates a /16 network with a /24 subnet. Public subnet instances use Elastic IPs or Public IPs to access the Internet

VPC with Public and Private Subnets:

- In addition to containing a public subnet, this configuration adds a private subnet whose instances are not addressable from the Internet
- Instances in the private subnet can establish outbound connections to the Internet via the public subnet using Network Address Translation (NAT)
- Creates a /16 network with two /24 subnets
- Public subnet instances use Elastic IPs to access the Internet
- Private subnet instances access the Internet via Network Address Translation (NAT)

VPC with Public and Private Subnets and Hardware VPN Access:

- This configuration adds an IPsec Virtual Private Network (VPN) connection between your Amazon VPC and your data center – effectively extending your data center to the cloud while also providing direct access to the Internet for public subnet instances in your Amazon VPC
- Creates a /16 network with two /24 subnets
- One subnet is directly connected to the Internet while the other subnet is connected to your corporate network via an IPsec VPN tunnel

VPC with a Private Subnet Only and Hardware VPN Access:

- Your instances run in a private, isolated section of the AWS cloud with a private subnet whose instances are not addressable from the Internet
- You can connect this private subnet to your corporate data center via an IPsec Virtual Private Network (VPN) tunnel
- Creates a /16 network with a /24 subnet and provisions an IPsec VPN tunnel between your Amazon VPC and your corporate network

NAT INSTANCES

NAT instances are managed **by** you

Used to enable private subnet instances to access the Internet

When creating NAT instances always disable the source/destination check on the instance

NAT instances must be in a single public subnet

NAT instances need to be assigned to security groups

NAT GATEWAYS

NAT gateways are managed **for** you by AWS

NAT gateways are highly available in each AZ into which they are deployed

They are preferred by enterprises

Can scale automatically up to 45Gbps

No need to patch

Not associated with any security groups

The table below describes some differences between NAT instances and NAT gateways:

	NAT Gateway	NAT Instance
Managed	Managed by AWS	Managed by you
Availability	Highly available within an AZ	Not highly available (would require scripting)
Bandwidth	Up to 45 Gbps	Depends on the bandwidth of the EC2 instance type selected
Maintenance	Managed by AWS	Managed by you
Performance	Optimized for NAT	Amazon Linux AMI configured to perform NAT
Public IP	Elastic IP that cannot be detached	Elastic IP that can be detached
Security Groups	Cannot associate with a Security Group	Can associate with a Security Group
Bastion Host	Not supported	Can be used as a bastion host

AWS DIRECT CONNECT

AWS Direct Connect is a network service that provides an alternative to using the Internet to connect a customer's on premise sites to AWS

Data is transmitted through a private network connection between AWS and a customer's datacenter or corporate network

Benefits:

- Reduce cost when using large volumes of traffic
- Increase reliability (predictable performance)
- Increase bandwidth (predictable bandwidth)
- Decrease latency

Each AWS Direct Connect connection can be configured with one or more virtual interfaces (VIFs)

Public VIFs allow access to public services such as S3, EC2, and DynamoDB

Private VIFs allow access to your VPC

From Direct Connect you can connect to all AZs **within the region**

You can establish IPSec connections over public VIFs to remote regions

Direct Connect is charged by port hours and data transfer

Available in 1Gbps and 10Gbps

Speeds of 50Mbps, 100Mbps, 200Mbps, 300Mbps, 400Mbps, and 500Mbps can be purchased through AWS Direct Connect Partners

Uses Ethernet trunking (802.1q)

Each connection consists of a single dedicated connection between ports on the customer router and an Amazon router

for HA you must have 2 DX connections – can be active/active or active/standby

Route tables need to be updated to point to a Direct Connect connection

VPN can be maintained as a backup with a higher BGP priority

You cannot extend your on-premise VLANs into the AWS cloud using Direct Connect

AWS NETWORKING PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Which of the following statements are correct about the benefits of AWS Direct Connect? (choose 2)

A. Quick to implement

B. Increased reliability (predictable performance)

C. Lower cost than a VPN

D. Increased bandwidth (predictable bandwidth)

E. Uses redundant paths across the Internet

Question 2:

What advantages do NAT Gateways have over NAT Instances? (choose 2)

A. Can be assigned to security groups

B. Can be used as a bastion host

C. Managed for you by AWS

D. Highly available within each AZ

E. Can be scaled up manually

Question 3:

What is the scope of an Amazon Virtual Private Cloud (VPC)?

A. It spans multiple subnets

B. It spans a single CIDR block

C. It spans all Availability Zones in all regions

D. It spans all Availability Zones within a region

Question 4:

Which of the below are valid options within the VPC Wizard? (choose 2)

A. VPC with Two Public Subnets

B. VPC with Private Subnets

C. VPC with a Single Public Subnet

D. VPC with Public and Private Subnets and Hardware VPN Access

E. VPC with a Private Subnet Only and Software VPN Access

Question 1 answer: *B,D*

Explanation:

- AWS Direct Connect is a network service that provides an alternative to using the Internet to connect customers' on premise sites to AWS
- Data is transmitted through a private network connection between AWS and a customer's datacenter or corporate network
- Benefits:
- - Reduce cost when using large volumes of traffic
- - Increase reliability (predictable performance)
- - Increase bandwidth (predictable bandwidth)
- - Decrease latency
- Direct Connect is not fast to implement as it can take weeks to months to setup (use VPN for fast deployment times)
- Direct Connect is more expensive than VPN
- Direct Connect uses private network connections, it does not use redundant paths over the Internet

Question 2 answer: *C,D*

Explanation:

- NAT gateways are managed **for** you by AWS. NAT gateways are highly available in each AZ into which they are deployed. They are not associated with any security groups and can scale automatically up to 45Gbps
- NAT instances are managed **by** They must be scaled manually and do not provide HA. NAT Instances can be used as bastion hosts and can be assigned to security groups

Question 3 answer: *D*

Explanation:

- A virtual private cloud (VPC) is a virtual network dedicated to your AWS account. A VPC spans all the Availability Zones in the region
- You can have multiple CIDR blocks in a VPC
- A VPC spans AZs, subnets are created within AZs

Question 4 answer: *C,D*

Explanation:

- The options available in the VPC Wizard are:
- - VPC with a Single Public Subnet
- - VPC with Public and Private Subnets
- - VPC with Public and Private Subnets and Hardware VPN Access
- - VPC with a Private Subnet Only and Hardware VPN Access

AWS DATABASES

AMAZON RELATIONAL DATABASE SERVICES (RDS)

Amazon Relational Database Service (Amazon RDS) is a managed service that makes it easy to set up, operate, and scale a relational database in the cloud

Relational databases are known as Structured Query Language (SQL) databases

Non-relational databases are known as NoSQL databases

RDS is an Online Transaction Processing (OLTP) type of database

RDS features and benefits:

- SQL type of database
- Can be used to perform complex queries and joins
- Easy to setup, highly available, fault tolerant, and scalable
- Used when data is clearly defined
- Common use cases include online stores and banking systems

Amazon RDS supports the following database engines:

- SQL Server
- Oracle
- MySQL Server
- PostgreSQL
- Aurora
- MariaDB

Aurora is Amazon's proprietary database

RDS is a fully managed service and you do not have access to the underlying EC2 instance (no root access)

The RDS service includes the following:

- Security and patching of the DB instances
- Automated backup for the DB instances
- Software updates for the DB engine
- Easy scaling for storage and compute
- Multi-AZ option with synchronous replication
- Automatic failover for Multi-AZ option
- Read replicas option for read heavy workloads

A DB instance is a database environment in the cloud with the compute and storage resources you specify

Encryption:

- You can encrypt your Amazon RDS instances and snapshots at rest by enabling the encryption option for your Amazon RDS DB instance
- Encryption at rest is supported for all DB types and uses AWS KMS
- You cannot encrypt an existing DB, you need to create a snapshot, copy it, encrypt the copy, then build an encrypted DB from the snapshot

DB Subnet Groups:

- A DB subnet group is a collection of subnets (typically private) that you create in a VPC and that you then designate for your DB instances
- Each DB subnet group should have subnets in at least two Availability Zones in a given region
- It is recommended to configure a subnet group with subnets in each AZ (even for standalone instances)

AWS Charge for:

- DB instance hours (partial hours are charged as full hours)
- Storage GB/month
- I/O requests/month – for magnetic storage
- Provisioned IOPS/month – for RDS provisioned IOPS SSD
- Egress data transfer
- Backup storage (DB backups and manual snapshots)

Scalability:

- You can only scale RDS up (compute and storage)
- You cannot decrease the allocated storage for an RDS instance
- You can scale storage and change the storage type for all DB engines except MS SQL

RDS provides multi-AZ for disaster recovery which provides fault tolerance across availability zones:

- Multi-AZ RDS creates a replica in another AZ and synchronously replicates to it (DR only)
- There is an option to choose multi-AZ during the launch wizard
- AWS recommends the use of provisioned IOPS storage for multi-AZ RDS DB instances
- Each AZ runs on its own physically distinct, independent infrastructure, and is engineered to be highly reliable
- You cannot choose which AZ in the region will be chosen to create the standby DB instance

Read Replicas – provide improved performance for reads:

- Read replicas are used for read heavy DBs and replication is asynchronous
- Read replicas are for workload sharing and offloading
- Read replicas provide read-only DR
- Read replicas are created from a snapshot of the master instance
- Must have automated backups enabled on the primary (retention period > 0)

AMAZON DYNAMODB

Amazon DynamoDB is a fully managed NoSQL database service that provides fast and predictable performance with seamless scalability

Dynamo DB features and benefits:

- NoSQL type of database (non-relational)
- Fast, highly available, and fully managed
- Used when data is fluid and can change
- Common use cases include social networks and web analytics

Push button scaling means that you can scale the DB at any time without incurring downtime

SSD based and uses limited indexing on attributes for performance

DynamoDB is a Web service that uses HTTP over SSL (HTTPS) as a transport and JSON as a message serialisation format

Amazon DynamoDB stores three geographically distributed replicas of each table to enable high availability and data durability

Data is synchronously replicated across 3 facilities (AZs) in a region

Cross-region replication allows you to replicate across regions:

- Amazon DynamoDB global tables provides a fully managed solution for deploying a multi-region, multi-master database
- When you create a global table, you specify the AWS regions where you want the table to be available
- DynamoDB performs all of the necessary tasks to create identical tables in these regions, and propagate ongoing data changes to all of them

Provides low read and write latency

Scale storage and throughput up or down as needed without code changes or downtime

DynamoDB is schema-less

DynamoDB can be used for storing session state

Provides two read models

Eventually consistent reads (Default):

- The eventual consistency option maximises your read throughput (best read performance)
- An eventually consistent read might not reflect the results of a recently completed write
- Consistency across all copies reached within 1 second

Strongly consistent reads:

- A strongly consistent read returns a result that reflects all writes that received a successful response prior to the read (faster consistency)

AMAZON REDSHIFT

Amazon Redshift is a fast, fully managed data warehouse that makes it simple and cost-effective to analyze all your data using standard SQL and existing Business Intelligence (BI) tools

RedShift is a SQL based data warehouse used for **analytics** applications

RedShift is an Online Analytics Processing (OLAP) type of DB

RedShift is used for running complex analytic queries against petabytes of structured data, using sophisticated query optimization, columnar storage on high-performance local disks, and massively parallel query execution

RedShift is ideal for **processing** large amounts of data for business intelligence

RedShift is 10x faster than a traditional SQL DB

RedShift uses columnar data storage:

- Data is stored sequentially in columns instead of rows
- Columnar based DB is ideal for data warehousing and analytics
- Requires fewer I/Os which greatly enhances performance

RedShift provides advanced compression:

- Data is stored sequentially in columns which allows for much better performance and less storage space
- RedShift automatically selects the compression scheme

RedShift uses replication and continuous backups to enhance availability and improve durability and can automatically recover from component and node failures

RedShift always keeps three copies of your data:

- The original
- A replica on compute nodes (within the cluster)
- A backup copy on S3

RedShift provides continuous/incremental backups:

- Multiple copies within a cluster
- Continuous and incremental backups to S3
- Continuous and incremental backups across regions
- Streaming restore

RedShift provides fault tolerance for the following failures:

- Disk failures
- Nodes failures
- Network failures
- AZ/region level disasters

AMAZON ELASTICACHE

ElastiCache is a web service that makes it easy to deploy and run Memcached or Redis protocol-compliant server nodes in the cloud

The in-memory caching provided by ElastiCache can be used to significantly improve latency and throughput for many read-heavy application workloads or compute-intensive workloads

Best for scenarios where the DB load is based on Online Analytics Processing (OLAP) transactions

ElastiCache EC2 nodes cannot be accessed from the Internet, nor can they be accessed by EC2 instances in other VPCs

Can be on-demand or reserved instances too (but not Spot instances)

ElastiCache can be used for storing session state

There are two types of ElastiCache engine:

- Memcached – simplest model, can run large nodes with multiple cores/threads, can be scaled in and out, can cache objects such as DBs
- Redis – complex model, supports encryption, master / slave replication, cross AZ (HA), automatic failover and backup/restore

AWS DATABASE PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

What pricing models are available for DynamoDB? (choose 2)

A. On-demand capacity mode

B. Spot capacity mode

C. Provisioned capacity mode

D. Dedicated capacity mode

E. Reserved capacity mode

Question 2:

Which two types of database engine can be used with Amazon ElastiCache? (choose 2)

A. Memcached

B. HANA

C. Redis

D. MongoDB

E. MemSQL

Question 3:

Which type of data storage system is typically considered to hold "structured" data?

A. Non-relational database

B. File system

C. Email system

D. Relational database

Question 4:

What is the availability model of Amazon DynamoDB?

A. Data is synchronously replicated across all regions

B. Data is asynchronously replicated across all regions

C. Data is synchronously replicated across 3 facilities in a region

D. Data is asynchronously replicated across 3 facilities in a region

Question 1 answer: *A,C*

Explanation:

- **On-demand capacity mode:** DynamoDB charges you for the data reads and writes your application performs on your tables. You do not need to specify how much read and write throughput you expect your application to perform because DynamoDB instantly accommodates your workloads as they ramp up or down
- **Provisioned capacity mode:** you specify the number of reads and writes per second that you expect your application to require. You can use auto scaling to automatically adjust your table's capacity based on the specified utilization rate to ensure application performance while reducing cost

Question 2 answer: *A,C*

Explanation:

- ElastiCache is a web service that makes it easy to deploy and run Memcached or Redis protocol-compliant server nodes in the cloud. The in-memory caching provided by ElastiCache can be used to significantly improve latency and throughput for many read-heavy application workloads or compute-intensive workloads
- Only the Memcached and Redis database engines can be used with ElastiCache, the others in the list are all in-memory databases but are not supported

Question 3 answer: *D*

Explanation:

- Relation databases such as Structured Query Language (SQL) databases hold data in a structured format. Examples are Amazon RDS and Microsoft SQL Server
- File systems, email systems and non-relational databases hold data in an "unstructured" format. This means that though there is some structure to it, the data cannot be easily searched using standard data processing algorithms or structured queries. Unstructured data is more human-friendly than machine-friendly

Question 4 answer: *C*

Explanation:

- Amazon DynamoDB stores three geographically distributed replicas of each table to enable high availability and data durability. Data is synchronously replicated across 3 facilities (AZs) in a region

ELASTIC LOAD BALANCING AND AUTO SCALING

AMAZON ELASTIC LOAD BALANCING (ELB)

ELB automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, and IP addresses

ELB can handle the varying load of your application traffic in a single Availability Zone or across multiple Availability Zones

ELB features high availability, automatic scaling, and robust security necessary to make your applications fault tolerant

There are three types of Elastic Load Balancer (ELB) on AWS:

- Application Load Balancer (ALB) – layer 7 load balancer that routes connections based on the content of the request
- Network Load Balancer (NLB) – layer 4 load balancer that routes connections based on IP protocol data
- Classic Load Balancer (CLB) – this is the oldest of the three and provides basic load balancing at both layer 4 and layer 7

Application Load Balancer (ALB)

ALB is best suited for load balancing of HTTP and HTTPS traffic and provides advanced request routing targeted at the delivery of modern application architectures, including microservices and containers

Operating at the individual request level (Layer 7), Application Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) based on the content of the request

Network Load Balancer (NLB)

NLB is best suited for load balancing of TCP traffic where extreme performance is required

Operating at the connection level (Layer 4), Network Load Balancer routes traffic to targets within Amazon Virtual Private Cloud (Amazon VPC) and is capable of handling millions of requests per second while maintaining ultra-low latencies

Network Load Balancer is also optimized to handle sudden and volatile traffic patterns

Classic Load Balancer (CLB)

CLB provides basic load balancing across multiple Amazon EC2 instances and operates at both the request level and connection level

Classic Load Balancer is intended for applications that were built within the EC2-Classic network

The CLB may be phased out over time and Amazon are promoting the ALB and NLB for most use cases within VPC

Benefits:

- ELB provides high availability and fault tolerance by allowing traffic to be directed to multiple EC2 instances

AWS AUTO SCALING

Auto Scaling automates the process of adding (scaling up) OR removing (scaling down) EC2 instances based on the traffic demand for your application

Auto Scaling helps to ensure that you have the correct number of EC2 instances available to handle the application load

You create collections of EC2 instances, called Auto Scaling Group (ASG)

You can specify the minimum number of instances in each ASG, and Auto Scaling will ensure the group never goes beneath this size

You can also specify the maximum number of instances in each ASG and the group will never go above this size

A desired capacity can be configured and Auto Scaling will ensure the group has this number of instances

You can also specify scaling policies that control when Auto Scaling launches or terminates instances

Scaling policies determine when, if, and how the ASG scales and shrinks (on-demand/dynamic scaling, cyclic/scheduled scaling)

Scaling Plans define the triggers and when instances should be provisioned/de-provisioned

A launch configuration is the template used to create new EC2 instances and includes parameters such as instance family, instance type, AMI, key pair and security groups

Benefits:

- Auto Scaling enables elasticity and scalability

ELASTIC LOAD BALANCING AND AUTO SCALING PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

What are the primary benefits of using AWS Elastic Load Balancing? (choose 2)

A. High availability

B. Elasticity

C. Automation

D. Caching

E. Regional resilience

Question 2:

Which type of Elastic Load Balancer only distributes traffic using the TCP protocol information?

A. Application Load Balancer (ALB)

B. Network Load Balancer (NLB)

C. Classic Load Balancer (CLB)

D. No load balancers operate at the TCP level

Question 3:

What do you need to create to specify how your AWS Auto Scaling Group scales and shrinks?

A. IAM Policy

B. Scaling Plan

C. Scaling Policy

D. Launch Configuration

Question 4:

Which type of scaling does AWS Auto Scaling provide?

A. Vertical

B. Linear

C. Horizontal

D. Incremental

Question 1 answer: *A,B*

Explanation:

- High availability – ELB automatically distributes traffic across multiple EC2 instances in different AZs within a region
- Elasticity – ELB is capable of handling rapid changes in network traffic patterns
- An ELB can distribute incoming traffic across your Amazon EC2 instances in a single Availability Zone or multiple Availability Zones, but not across regions (for regional resilience)
- Automation is not a primary benefit of ELB
- Caching is not a benefit of ELB

Question 2 answer: B

Explanation:

- NLBs process traffic at the TCP level (layer 4)
- ALBs process traffic at the HTTP,HTTPS level (layer 7)
- CLBs process traffic at the TCP, SSL, HTTP and HTTPS levels (layer 4 & 7)

Question 3 answer: *C*

Explanation:

- Scaling policies determine when, if, and how the ASG scales and shrinks (on-demand/dynamic scaling, cyclic/scheduled scaling)
- Scaling Plans define the triggers and when instances should be provisioned/de-provisioned
- A launch configuration is the template used to create new EC2 instances and includes parameters such as instance family, instance type, AMI, key pair and security groups
- An IAM policy is not used to control Auto Scaling

Question 4 answer: *C*

Explanation:

- AWS Auto Scaling scales horizontally by adding additional compute instances

CONTENT DELIVERY AND DNS SERVICES

AMAZON ROUTE 53

Route 53 is the AWS Domain Name Service

Route 53 performs three main functions:

- Domain registration – Route 53 allows you to register domain names
- Domain Name Service (DNS) – Route 53 translates name to IP addresses using a global network of authoritative DNS servers
- Health checking – Route 53 sends automated requests to your application to verify that it's reachable, available and functional

You can use any combination of these functions

Route 53 benefits:

- Domain registration
- DNS service
- Traffic Flow (send users to the best endpoint)
- Health checking
- DNS failover (automatically change domain endpoint if system fails)
- Integrates with ELB, S3, and CloudFront as endpoints

AMAZON CLOUDFRONT

CloudFront is a content delivery network (CDN) that allows you to store (cache) your content at "edge locations" located around the world

This allows customers to access content more quickly and provides security against DDoS attacks

CloudFront can be used for data, videos, applications, and APIs

CloudFront benefits:

- Cache content at Edge Location for fast distribution to customers
- Built-in Distributed Denial of Service (DDoS) attack protection
- Integrates with many AWS services (S3, EC2, ELB, Route 53, Lambda)

CONTENT DELIVERY AND DNS SERVICES PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

How can a Solutions Architect reduce the latency between end-users and applications or content? (choose 2)

A. Deploy applications in multiple AZs

B. Deploy applications in regions closest to the end-users

C. Use S3 Transfer Acceleration to improve application performance

D. Use Amazon CloudFront to cache content closer to end-users

E. Use larger EC2 instance types for the applications

Question 2:

Identify the services that have a global (rather than regional) scope? (choose 2)

A. Amazon Route 53

B. Amazon S3

C. Amazon CloudFront

D. AWS Lambda

E. Amazon EC2

Question 3:

Which service supports the resolution of public domain names to IP addresses or AWS resources?

A. Amazon Route 53

B. Amazon CloudFront

C. Amazon SNS

D. Hosted Zones

Question 4:

Which of the following services does Amazon Route 53 provide? (choose 2)

A. Domain registration

B. Route tables

C. Domain Name Service (DNS)

D. Auto Scaling

E. Load balancing

Question 1 answer: *B,D*

Explanation:

- To reduce latency, which corresponds with the distance over which network communications travel, you should aim to host your applications closer to your end-users. This means deploying them in the closest regions
- Deploying in multiple AZs may create resiliency but won't change latency much as AZs are geographically close to each other
- S3 Transfer Acceleration is used to improve upload speeds for S3 objects and does not affect application performance
- CloudFormation is used for deploying resources through code ("infrastructure as code")
- Using a larger instance type for your application may improve application performance but will not reduce latency

Question 2 answer: *A,C*

Explanation:

- Amazon Route 53 and Amazon CloudFront have a global scope
- Amazon S3 uses a global namespace but buckets and objects are created within a region
- AWS Lambda is a regional service

Question 3 answer: *A*

Explanation:

- Amazon Route 53 is a highly available and scalable Domain Name System (DNS) service
- A hosted zone is a collection of records for a specified domain in Route 53
- CloudFront is a content delivery network (CDN) that allows you to store (cache) your content at “edge locations” located around the world
- Simple Notification Service is used to send notifications over multiple transport protocols

Question 4 answer: *A,C*

Explanation:

- Route 53 services include domain registration, DNS, health checking (availability monitoring) and traffic management

MONITORING AND LOGGING SERVICES

AMAZON CLOUDWATCH

Amazon CloudWatch is a monitoring service for AWS cloud resources and the applications you run on AWS

CloudWatch is for performance monitoring (CloudTrail is for auditing)

Used to collect and track metrics, collect and monitor log files, and set alarms

Automatically react to changes in your AWS resources

Monitor resources such as:

- EC2 instances
- DynamoDB tables
- RDS DB instances
- Custom metrics generated by applications and services
- Any log files generated by your applications

Gain system-wide visibility into resource utilization

Monitor application performance

Monitor operational health

CloudWatch is accessed via API, command-line interface, AWS SDKs, and the AWS Management Console

CloudWatch integrates with IAM

Amazon CloudWatch Logs lets you monitor and troubleshoot your systems and applications using your existing system, application and custom log files

CloudWatch Logs can be used for real time application and system monitoring as well as long term log retention

CloudWatch Logs keeps logs indefinitely by default

CloudTrail logs can be sent to CloudWatch Logs for real-time monitoring

CloudWatch Logs metric filters can evaluate CloudTrail logs for specific terms, phrases or values

CloudWatch retains metric data as follows:

- Data points with a period of less than 60 seconds are available for 3 hours. These data points are high-resolution custom metrics.

- Data points with a period of 60 seconds (1 minute) are available for 15 days
- Data points with a period of 300 seconds (5 minute) are available for 63 days
- Data points with a period of 3600 seconds (1 hour) are available for 455 days (15 months)

Dashboards allow you to create, customize, interact with, and save graphs of AWS resources and custom metrics

Alarms can be used to monitor any Amazon CloudWatch metric in your account

Events are a stream of system events describing changes in your AWS resources

Logs help you to aggregate, monitor and store logs

Basic monitoring = 5 mins (free for EC2 Instances, EBS volumes, ELBs and RDS DBs)

Detailed monitoring = 1 min (chargeable)

Metrics are provided automatically for a number of AWS products and services

There is no standard metric for memory usage on EC2 instances

A custom metric is any metric you provide to Amazon CloudWatch (e.g. time to load a web page or application performance)

Options for storing logs:

- CloudWatch Logs
- Centralized logging system (e.g. Splunk)
- Custom script and store on S3

Do not store logs on non-persistent disks:

Best practice is to store logs in CloudWatch Logs or S3

CloudWatch Logs subscription can be used across multiple AWS accounts (using cross account access)

Amazon CloudWatch uses Amazon SNS to send email

AWS CLOUDTRAIL

AWS CloudTrail is a web service that records activity made on your account and delivers log files to an Amazon S3 bucket

CloudTrail is for auditing (CloudWatch is for performance monitoring)

CloudTrail is about logging and saves a history of API calls for your AWS account

Provides visibility into user activity by recording actions taken on your account

API history enables security analysis, resource change tracking, and compliance auditing

Logs API calls made via:

- AWS Management Console
- AWS SDKs
- Command line tools
- Higher-level AWS services (such as CloudFormation)

CloudTrail records account activity and service events from most AWS services and logs the following records:

- The identity of the API caller
- The time of the API call
- The source IP address of the API caller
- The request parameters
- The response elements returned by the AWS service

CloudTrail is not enabled by default

CloudTrail is per AWS account

You can consolidate logs from multiple accounts using an S3 bucket:

1. Turn on CloudTrail in the paying account
2. Create a bucket policy that allows cross-account access
3. Turn on CloudTrail in the other accounts and use the bucket in the paying account

You can integrate CloudTrail with CloudWatch Logs to deliver data events captured by CloudTrail to a CloudWatch Logs log stream

CloudTrail log file integrity validation feature allows you to determine whether a CloudTrail log file was unchanged, deleted, or modified since CloudTrail delivered it to the specified Amazon S3 bucket

MONITORING AND LOGGING SERVICES PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

A manager needs to keep a check on his AWS spend. How can the manager setup alarms that notify him when his bill reaches a certain amount?

A. Using CloudWatch

B. Using AWS Trusted Advisor

C. Using CloudTrail

D. By notifying AWS support

Question 2:

Which service provides visibility into user activity by recording actions taken on your account?

A. Amazon CloudWatch

B. Amazon CloudFormation

C. Amazon CloudTrail

D. Amazon CloudHSM

Question 3:

What types of monitoring can Amazon CloudWatch be used for? (choose 2)

A. Application performance

B. API access

C. Operational health

D. Infrastructure

E. Data center

Question 4:

You would like to collect custom metrics from a production application every 1 minute. What type of monitoring should you use?

A. CloudWatch with detailed monitoring

B. CloudWatch with basic monitoring

C. CloudTrail with detailed monitoring

D. CloudTrail with basic monitoring

Question 1 answer: A

Explanation:

- The best ways to do this is to use CloudWatch to configure alarms that deliver a notification when activated. The alarms can use cost metrics that trigger the alarm when a certain amount of spend has been reached

Question 2 answer: *C*

Explanation:

- CloudTrail is a web service that records activity made on your account and delivers log files to an Amazon S3 bucket
- CloudTrail is for auditing (CloudWatch is for performance monitoring)
- CloudFormation is used for deploying infrastructure through code
- CloudHSM is a hardware security module for generating, managing and storing encryption keys

Question 3 answer: *A,C*

Explanation:

- Amazon CloudWatch is a monitoring service for AWS cloud resources and the applications you run on AWS. CloudWatch performs performance monitoring and can monitor custom metrics generated by applications and the operational health of your AWS resources
- Amazon CloudTrail monitors API access
- Infrastructure and data center monitoring is not accessible to AWS customers

Question 4 answer: *A*

Explanation:

- Amazon CloudWatch is a monitoring service for AWS cloud resources and the applications you run on AWS. CloudWatch is for performance monitoring (CloudTrail is for auditing). Used to collect and track metrics, collect and monitor log files, and set alarms. Basic monitoring collects metrics every 5 minutes whereas detailed monitoring collects metrics every 1 minute
- AWS CloudTrail is a web service that records activity made on your account and delivers log files to an Amazon S3 bucket. CloudTrail is for auditing (CloudWatch is for performance monitoring). CloudTrail is about logging and saves a history of API calls for your AWS account

NOTIFICATION SERVICES

AMAZON SIMPLE NOTIFICATION SERVICE (SNS)

Amazon Simple Notification Service (Amazon SNS) is a web service that makes it easy to set up, operate, and send notifications from the cloud

Amazon SNS is used for building and integrating loosely-coupled, distributed applications

SNS provides instantaneous, push-based delivery (no polling)

SNS concepts:

- Topics – how you label and group different endpoints that you send messages to
- Subscriptions – the endpoints that a topic sends messages to
- Publishers – the person/alarm/event that gives SNS the message that needs to be sent

SNS usage:

- Send automated or manual notifications
- Send notification to email, mobile (SMS), SQS, and HTTP endpoints
- Closely integrated with other AWS services such as CloudWatch so that alarms, events, and actions in your AWS account can trigger notifications

Uses simple APIs and easy integration with applications

Flexible message delivery is provided over multiple transport protocols

Offered under an inexpensive, pay-as-you-go model with no up-front costs

The web-based AWS Management Console offers the simplicity of a point-and-click interface

Data type is JSON

SNS supports a wide variety of needs including event notification, monitoring applications, workflow systems, time-sensitive information updates, mobile applications, and any other application that generates or consumes notifications

SNS Subscribers:

- HTTP
- HTTPS
- Email
- Email-JSON
- SQS
- Application
- Lambda

SNS supports notifications over multiple transport protocols:

- HTTP/HTTPS – ssubscribers specify a URL as part of the subscription registration
- Email/Email-JSON – mmessages are sent to registered addresses as email (text-based or JSON-object)
- SQS – users can specify an SQS standard queue as the endpoint
- SMS – messages are sent to registered phone numbers as SMS text messages

Topic names are limited to 256 characters

SNS supports CloudTrail auditing for authenticated calls

SNS provides durable storage of all messages that it receives (across multiple AZs)

NOTIFICATION SERVICES PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Which service can be used for building and integrating loosely-coupled, distributed applications?

A. Amazon EBS

B. Amazon SNS

C. Amazon EFS

D. Amazon RDS

Question 2:

Which AWS service can be used to send automated notifications to HTTP endpoints?

A. Amazon SQS

B. Amazon SWF

C. Amazon SNS

D. Amazon SES

Question 1 answer: B

Explanation:

- Amazon Simple Notification Service (Amazon SNS) is a web service that makes it easy to set up, operate, and send notifications from the cloud

- Amazon SNS is used for building and integrating loosely-coupled, distributed applications
- Amazon Elastic Block Storage (EBS) provides storage volumes for EC2 instances
- Amazon Elastic File System (EFS) provides an NFS filesystem for usage by EC2 instances
- Amazon Relational Database Service (RDS) provides a managed relational database service

Question 2 answer: *C*

Explanation:

- Amazon Simple Notification Service (Amazon SNS) is a web service that makes it easy to set up, operate, and send notifications from the cloud. SNS can be used to send automated or manual notifications to email, mobile (SMS), SQS, and HTTP endpoints
- Amazon Simple Queue Service (SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications
- Amazon SWF helps developers build, run, and scale background jobs that have parallel or sequential step
- Amazon Simple Email Service (Amazon SES) is a cloud-based email sending service designed to help digital marketers and application developers send marketing, notification, and transactional emails

BILLING AND PRICING

GENERAL PRICING INFORMATION

AWS Billing and Pricing is one of the key subjects on the Cloud Practitioner exam. It is recommended to read the following whitepaper to understand how AWS pricing works: https://d1.awsstatic.com/whitepapers/aws_pricing_overview.pdf

AWS works on a pay as you go model in which you only pay for what you use, when you are using it

If you turn off resources, you don't pay for them (you may pay for consumed storage)

There are no upfront charges and you stop paying for a service when you stop using it

Aside from EC2 reserved instances you are not locked into long term contracts and can terminate whenever you choose to

Volume discounts are available so the more you use a service the cheaper it gets (per unit used)

There are no termination fees

The three fundamental drivers of cost with AWS are: compute, storage and outbound data transfer

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)

Outbound data transfer is aggregated across services and then charged at the outbound data transfer rate

Free tier allows you to run certain resources for free

Free tier includes offers that expire after 12 months and offers that never expire

Pricing policies include:

- Pay as you go
- Pay less when you reserve
- Pay even less per unit when using more
- Pay even less as AWS grows
- Custom pricing (enterprise customers only)

Free services include:

- Amazon VPC
- Elastic Beanstalk (but not the resources created)
- CloudFormation (but not the resources created)

- Identity Access Management (IAM)
- Auto Scaling (but not the resources created)
- OpsWorks
- Consolidated Billing

Fundamentally charges include:

1. Compute
2. Storage
3. Data out

AMAZON EC2 PRICING

EC2 pricing is based on:

- Clock hours of server uptime
- Machine configuration
- Machine type
- Number of instances
- Load balancing
- Detailed monitoring
- Auto Scaling (resources created)
- Elastic IP addresses
- Operating systems and software packages

There are several pricing model for AWS services, these include:

On Demand:

- Means you pay for compute or database capacity with no long-term commitments of upfront payments
- You pay for the computer capacity per hour or per second (Linux only, and applies to On-Demand, Reserved and Spot instances)
- Recommended for users who prefer low cost and flexibility without upfront payment or long-term commitments
- Good for applications with short-term, spiky, or unpredictable workloads that cannot be interrupted

Dedicated Hosts:

- A dedicated host is an EC2 servers dedicated to a single customer
- Runs in your VPC
- Good for when you want to leverage existing server-bound software licences such as Windows Server, SQL Server, and SUSE Linux Enterprise Server

- Also good for meeting compliance requirements

Dedicated Instances:

- Dedicated Instances are Amazon EC2 instances that run in a VPC on hardware that's dedicated to a single customer
- Dedicated instances are physically isolated at the host hardware level from instances that belong to other AWS accounts
- Dedicated instances may share hardware with other instances from the same AWS account that are not Dedicated instances

Spot Instances:

- Purchase spare computing capacity with no upfront commitment at discounted hourly rates
- Provides up to 90% off the On-Demand price
- Recommended for applications that have flexible start and end times, applications that are only feasible at very low compute prices, and users with urgent computing needs for a lot of additional capacity
- In the old model Spot instances were terminated because of higher competing bids, in the new model this does not happen but instances still may be terminated (with a 2 minute warning) when EC2 needs the capacity back – note: the exam may not be updated to reflect this yet

Reservations:

- Reserved instances provide significant discounts, up to 75% compared to On-Demand pricing, by paying for capacity ahead of time
- Provide a capacity reservation when applied to a specific Availability Zone
- Good for applications that have predictable usage, that need reserved capacity, and for customers who can commit to a 1 or 3-year term

Reservations apply to various services, including:

- Amazon EC2 Reserved Instances
- Amazon DynamoDB Reserved Capacity
- Amazon ElastiCache Reserved Nodes
- Amazon RDS Reserved Instances
- Amazon RedShift Reserved Instances

Reservation options include no upfront, partial upfront and all upfront

Reservation terms are 1 or 3 years

AMAZON SIMPLE STORAGE SERVICE (S3) PRICING

Storage pricing is determined by:

- **Storage class** – e.g. Standard or IA
- **Storage quantity** – data volume stored in your buckets on a per GB basis
- **Number of requests** – the number and type of requests, e.g. GET, PUT, POST, LIST, COPY
- **Lifecycle transitions requests** – moving data between storage classes
- **Data transfer** – data transferred out of an S3 region is charged

AMAZON GLACIER PRICING

Extremely low cost and you pay only for what you need with no commitments of upfront fees

Charged for requests and data transferred out of Glacier

"Amazon Glacier Select" pricing allows queries to run directly on data stored on Glacier without having to retrieve the archive. Priced on amount of data scanned, returned, and number of requests initiated

Three options for access to archives, listed in the table below:

	Expedited	Standard	Bulk
Data access time	1-5 minutes	3-5 hours	5-12 hours
Data retrievals	$0.03 per GB	$0.01 per GB	$0.0025 per GB
Retrieval requests	On-Demand: $0.01 per request Provisioned: $100 per Provisioned Capacity Unit	$0.050 per 1,000 requests	$0.025 per 1,000 requests

AWS SNOWBALL PRICING

Pay a service fee per data transfer job and the cost of shipping the appliance

Each job allows use of Snowball appliance for 10 days onsite for free

Data transfer in to AWS is free and outbound is charged (per region pricing)

AMAZON RELATIONAL DATABASE SERVICE (RDS) PRICING

RDS pricing is determined by:

- **Clock hours of server uptime** – amount of time the DB instance is running
- **Database characteristics** – e.g. databased engine, size and memory class
- **Database purchase type** – e.g. On-Demand, Reserved
- **Number of database instances**
- **Provisioned storage** – backup is included up to 100% of the size of the DB. After the DB is terminated backup storage is charged per GB per month
- **Additional storage** – the amount of storage in addition to the provisioned storage is charged per GB per month
- **Requests** – the number of input and output requests to the DB
- **Deployment type** – single AZ or multi-AZ
- **Data transfer** – inbound is free, outbound data transfer costs are tiered
- **Reserved Instances** – RDS RIs can be purchased with No Upfront, Partial Upfront, or All Upfront terms. Available for Aurora, MySQL, MariaDB, Oracle and SQL Server

AMAZON CLOUDFRONT PRICING

CloudFront pricing is determined by:

- **Traffic distribution** – data transfer and request pricing, varies across regions, and is based on the edge location from which the content is served
- **Requests** – the number and type of requests (HTTP or HTTPS) and the geographic region in which they are made
- **Data transfer out** – quantity of data transferred out of CloudFront edge locations
- There are additional chargeable items such as invalidation requests, field-level encryption requests, and custom SSL certificates

AWS LAMBDA PRICING

Pay only for what you use and charged based on the number of requests for functions and the time it takes to execute the code

Price is dependent on the amount of memory allocated to the function

AMAZON ELASTIC BLOCK STORE (EBS) PRICING

Pricing is based on three factors:

- Volumes – volume storage for all EBS volumes types is charged by the amount of GB provisioned per month
- Snapshots – based on the amount of space consumed by snapshots in S3. Copying snapshots is charged on the amount of data copied across regions
- Data transfer – inbound data transfer is free, outbound data transfer charges are tiered

AMAZON DYNAMODB PRICING

Charged based on:

- **Provisioned throughput (write)**
- **Provisioned throughput (read)**
- **Indexed data storage**
- **Data transfer** – no charge for data transfer between DynamoDB and other AWS services within the same region, across regions is charged on both sides of the transfer
- **Global tables** – charged based on the resources associated with each replica of the table (replicated write capacity units, or rWCUs)
- **Reserved Capacity** – option available for a one-time upfront fee and commitment to paying a minimum usage level at specific hourly rates for the duration of the term. Additional throughput is charged at standard rates

The table below provides more details:

Resource Type	Details	Monthly Price
Provisioned throughput (write)	One write capacity unit (WCU) provides up to one write per second, enough for 2.5 million writes per month	As low as $0.47 per WCU
Provisioned throughput (read)	One read capacity unit (RCU) provides up to two reads per second, enough for 5.2 million reads per month	As low as $0.09 per RCU
Indexed data storage	DynamoDB charges an hourly rate per GB of disk space that your table consumes	As low as $0.25 per GB

Always remember that AWS is fundamentally a service in which you pay only for what you use and can start and stop using services whenever you choose

You do not have to enter into any contracts however you may choose to do so for lower pricing

AWS SUPPORT PLANS

There are four AWS support plans available:

- Basic – no support (access to forums only)
- Developer – business hours support via email
- Business – 24×7 email, chat and phone support
- Enterprise – 24×7 email, chat and phone support

Enterprise support comes with a Technical Account Manager (TAM)

Developer allows one person to open unlimited cases

Business and Enterprise allow unlimited contacts to open unlimited cases

The table below highlights the features of each support plan (make sure you know these for the exam):

	Basic	Developer	Business	Enterprise
Customer Service and Communities	24x7 access to customer service, documentation, whitepapers, and support forums	24x7 access to customer service, documentation, whitepapers, and support forums	24x7 access to customer service, documentation, whitepapers, and support forums	24x7 access to customer service, documentation, whitepapers, and support forums
Best Practices	Access to 7 core Trusted Advisor checks	Access to 7 core Trusted Advisor checks	Access to full set of Trusted Advisor checks	Access to full set of Trusted Advisor checks
Health status and Notifications	Access to Personal Health Dashboard	Access to Personal Health Dashboard	Access to Personal Health Dashboard & Health API	Access to Personal Health Dashboard & Health API
Technical Support		Business hours** access to Cloud Support Associates via email	24x7 access to Cloud Support Engineers via email, chat & phone	24x7 access to Sr. Cloud Support Engineers via email, chat & phone
Who Can Open Cases		One primary contact/ Unlimited cases	Unlimited contacts/ Unlimited cases (IAM supported)	Unlimited contacts/ Unlimited cases (IAM supported)
Case Severity/ Response Times*		General guidance: < 24 business hours System impaired: < 12 business hours	General guidance: < 24 hours System impaired: < 12 hours Production system impaired: < 4 hours Production system down: < 1 hour	General guidance: < 24 hours System impaired: < 12 hours Production system impaired: < 4 hours Production system down: < 1 hour Business-critical system down: < 15 minutes

RESOURCE GROUPS AND TAGGING

Tags are key / value pairs that can be attached to AWS resources

Tags contain metadata (data about data)

Tags can sometimes be inherited – e.g. resources created by Auto Scaling, CloudFormation or Elastic Beanstalk

Resource groups make it easy to group resources using the tags that are assigned to them. You can group resources that share one or more tags

Resource groups contain general information, such as:

- Region
- Name

- Health Checks

And also specific information, such as:

- Public & private IP addresses (for EC2)
- Port configurations (for ELB)
- Database engine (for RDS)

AWS ORGANIZATIONS AND CONSOLIDATED BILLING

AWS organizations allows you to consolidate multiple AWS accounts into an organization that you create and centrally manage

Available in two feature sets:

- Consolidated Billing
- All features

Includes root accounts and organizational units

Policies are applied to root accounts or OUs

Consolidated billing includes:

- Paying Account – independent and cannot access resources of other accounts
- Linked Accounts – all linked accounts are independent

Limit of 20 linked accounts (by default)

One bill for multiple AWS accounts

Easy to track charges and allocate costs

Volume pricing discounts can be applied to resources

Billing alerts enabled on the Paying account include data for all Linked accounts (or can be created per Linked account)

Consolidated billing allows you to get volume discounts on all of your accounts

Unused reserved instances (RIs) for EC2 are applied across the group

CloudTrail is on a per account basis and per region basis but can be aggregated into a single bucket in the paying account

Best practices:

- Always enable multi-factor authentication (MFA) on the root account
- Always use a strong and complex password on the root account
- The Paying account should be used for billing purposes only. Do not deploy resources into the Paying account

AWS QUICK STARTS

Quick Starts are built by AWS solutions architects and partners to help you deploy popular solutions on AWS, based on AWS best practices for security and high availability

These reference deployments implement key technologies automatically on the AWS Cloud, often with a single click and in less than an hour

Leverages CloudFormation

AWS COST CALCULATORS AND TOOLS

- **AWS Cost Explorer** – enables you to visualize your usage patterns over time and to identify your underlying cost drivers
- **AWS Simple Monthly calculator** – shows you how much you would pay in AWS if you move your resources
- **Total Cost of Ownership (TCO) calculator** – use to compare the cost of running your applications in an on-premise or colocation environment against AWS

AWS Cost Explorer

The AWS Cost Explorer is a free tool that allows you to view charts of your costs

You can view cost data for the past 13 months and forecast how much you are likely to spend over the next three months

Cost Explorer can be used to discover patterns in how much you spend on AWS resources over time and to identify cost problem areas

Cost Explorer can help you to identify service usage statistics such as:

- Which services you use the most
- View metrics for which AZ has the most traffic
- Which linked account is used the most

AWS Simple Monthly Calculator

The AWS Simple Monthly Calculator helps customers and prospects estimate their monthly AWS bill more efficiently

With the AWS Simple Monthly Calculator you can add services in different regions

The calculator includes support for most AWS services and you can include additional costs such as data ingress/egress charges, data storage charges, and retrieval fees

It is possible to select EC2 dedicated hosts and reserved instances with various pricing models

Support can also be added

TCO Calculator

The TCO calculator is a free tool provided by AWS that allows you to estimate the cost savings of using the AWS Cloud vs. using an on-premised data center

The TCO calculator therefore helps you to reduce Total Cost of Ownership (TCO) by avoiding large capital expenditures on hardware and infrastructure

The TCO calculator can also provide directional guidance on cost savings

The TCO calculator works by you inputting cost elements of your current/or estimated on-premises data center, and comparing those cost requirements with how much it would cost on AWS

Elements can be added/modified as you move through the process to best estimate the cost savings

BILLING AND PRICING PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

To reward customers for using their services, what are two ways AWS reduce prices? (choose 2)

A. Volume based discounts when you use more services

B. Reduction in inbound data transfer charges

C. Reduced cost for reserved capacity

D. Discounts for using a wider variety of services

E. Removal of termination fees for customers who spend more

Question 2:

How do AWS charge for Amazon CloudFront? (choose 2)

A. Data transfer out

B. Data transfer in

C. Number of requests

D. Number of users

E. Uptime

Question 3:

Which pricing model should you use for EC2 instances that will be used in a lab environment for several hours on a weekend and must run uninterrupted?

A. On-Demand

B. Reserved

C. Spot

D. Dedicated Instance

Question 4:

Which types of pricing policies does AWS offer? (choose 2)

A. Pay-as-you-go

B. Enterprise license agreement (ELA)

C. Non-peak hour discounts

D. Global usage discounts

E. Save when you reserve

Question 1 answer: A,C

Explanation:

- AWS provide volume based discount so that when you use more services you reduce the cost per service. You can also reserve capacity by locking in to fixed 1 or 3 year contracts to get significant discounts
- You never pay for inbound data transfer
- You don't get discounts for using a variety of services, only when you use more services
- There are never termination fees with AWS

Question 2 answer: *A,C*

Explanation:

- With Amazon CloudFront the basic elements you are charged for include the amount of data transfer out and the number of requests. There are additional chargeable items such as invalidation requests, field-level encryption requests, and custom SSL certificates

Question 3 answer: *A*

Explanation:

- Spot instances are good for short term requirements as they can be very economical. However, you may find that the instance is terminated if the spot market price moves
- On-Demand is the best choice for this situation as it is the most economical option that will ensure no interruptions
- Reserved instances are good for long-term, static requirements as you must lock-in for 1 or 3 years in return for a decent discount
- Dedicated instances are EC2 instances that run on hardware dedicated to a single customer

Question 4 answer: *A,E*

Explanation:

- Amazon pricing includes options for pay-as-you-go, save when you reserve and pay less by using more
- Amazon does not offer ELAs, non-peak hour discounts, or global usage discounts

CLOUD SECURITY

GENERAL

As an AWS customer you inherit all the best practices of AWS policies, architecture, and operational processes

The AWS Cloud enables a shared responsibility model

AWS manages security OF the cloud, you are responsible for security IN the cloud

You retain control of the security you choose to implement to protect your own content, platform, applications, systems, and networks no differently than you would in an on-site data center

BENEFITS OF AWS SECURITY

Keep Your Data Safe – the AWS infrastructure puts strong safeguards in place to help

Protect your 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 s 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 your data safe

COMPLIANCE

AWS Cloud Compliance enables you to understand the robust controls in place at AWS to maintain security and data protection in the cloud

As systems are built on top of AWS Cloud infrastructure, compliance responsibilities will be shared

Compliance programs include:

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

AWS WAF & AWS SHIELD

AWS WAF

AWS WAF is a web application firewall

Protects against common exploits that could compromise application availability, compromise security or consume excessive resources

AWS Shield

AWS Shield is a managed Distributed Denial of Service (DDoS) protection service

Safeguards web application running on AWS with always-on detection and automatic inline mitigations

Helps to minimize application downtime and latency

Two tiers – Standard and Advanced

AWS KEY MANAGEMENT SERVICE (KMS)

AWS Key Management Service gives you centralized control over the encryption keys used to protect your data

You can create, import, rotate, disable, delete, define usage policies for, and audit the use of encryption keys used to encrypt your data

AWS Key Management Service is integrated with most other AWS services making it easy to encrypt the data you store in these services with encryption keys you control

AWS KMS is integrated with AWS CloudTrail which provides you the ability to audit who used which keys, on which resources, and when

AWS KMS enables developers to easily encrypt data, whether through 1-click encryption in the AWS Management Console, or using the AWS SDK to easily add encryption in their application code

https://aws.amazon.com/kms/features/

AWS CLOUDHSM

AWS CloudHSM is a cloud-based hardware security module (HSM) that enables you to easily generate and use your own encryption keys on the AWS Cloud

With CloudHSM, you can manage your own encryption keys using FIPS 140-2 Level 3 validated HSMs

CloudHSM offers you the flexibility to integrate with your applications using industry-standard APIs, such as PKCS#11, Java Cryptography Extensions (JCE), and Microsoft CryptoNG (CNG) libraries

https://aws.amazon.com/cloudhsm/features/

AWS ARTIFACT

AWS Artifact is your go-to, central resource for compliance-related information that matters to you.

It provides on-demand access to AWS' security and compliance reports and select online agreements

Reports available in AWS Artifact include our Service Organization Control (SOC) reports, Payment Card Industry (PCI) reports, and certifications from accreditation bodies across geographies and compliance verticals that validate the implementation and operating effectiveness of AWS security controls

Agreements available in AWS Artifact include the Business Associate Addendum (BAA) and the Nondisclosure Agreement (NDA)

https://aws.amazon.com/artifact/

AWS INSPECTOR AND AWS TRUSTED ADVISOR

AWS Inspector

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

Inspector automatically assesses applications for vulnerabilities or deviations from best practices

Uses an agent installed on EC2 instances

Instances must be tagged

AWS Trusted Advisor

Trusted Advisor is an online resource that helps to reduce cost, increase performance and improve security by optimizing your AWS environment

Trusted Advisor provides real time guidance to help you provision your resources following best practices

Advisor will advise you on Cost Optimization, Performance, Security, and Fault Tolerance

Trusted Advisor scans your AWS infrastructure and compares is to AWS best practices in five categories:

- Cost Optimization
- Performance
- Security
- Fault Tolerance
- Service Limits

Trusted Advisor comes in two versions:

- Core Checks and Recommendations (free)
  - Access to the 7 core checks to help increase security and performance
  - Checks include S3 bucket permissions, Security Groups, IAM use, MFA on root account, EBS public snapshots, RDS public snapshots
- Full Trusted Advisor Benefits (business and enterprise support plans)
  - Full set of checks to help optimize your entire AWS infrastructure
  - Advises on security, performance, cost, fault tolerance and service limits
  - Additional benefits include weekly update notifications, alerts, automated actions with CloudWatch and programmatic access using the AWS Support API

AWS PERSONAL HEALTH DASHBOARD

AWS Personal Health Dashboard provides alerts and remediation guidance when AWS is experiencing events that may impact you

Personal Health Dashboard gives you a personalized view into the performance and availability of the AWS services underlying your AWS resources

The dashboard displays relevant and timely information to help you manage events in progress

Also provides proactive notification to help you plan for scheduled activities

Alerts are triggered by changes in the health of AWS resources, giving you event visibility, and guidance to help quickly diagnose and resolve issues

You get a personalized view of the status of the AWS services that power your applications, enabling you to quickly see when AWS is experiencing issues that may impact you

Also provides forward looking notifications, and you can set up alerts across multiple channels, including email and mobile notifications, so you receive timely and relevant information to help plan for scheduled changes that may affect you

Alerts include remediation details and specific guidance to enable you to take immediate action to address AWS events impacting your resources

Can integrate with Amazon CloudWatch Events, enabling you to build custom rules and select targets such as AWS Lambda functions to define automated remediation actions

The AWS Health API allows you to integrate health data and notifications with your existing in-house or third-party IT Management tools

PENETRATION TESTING

Penetration testing is the practice of testing one's own application's security for vulnerabilities by simulating an attack. AWS allows penetration testing, however you must request permission from AWS

- Permission is required for all penetration tests
- You must complete and submit the AWS Vulnerability / Penetration Testing Request Form to request authorization for penetration testing to or originating from any AWS resources
- There is a limited set of resources on which penetration testing can be performed

AWS policy only permits testing of the following resources:

- EC2
- RDS
- Aurora
- CloudFront
- API Gateway
- Lambda
- LightSail
- DNS Zone Walking

In case an account is or may be compromised, AWS recommend that the following steps are taken:

1. Change your AWS root account password

2. Change all IAM user's passwords
3. Delete or rotate all programmatic (API) access keys
4. Delete any resources in your account that you did not create
5. Respond to any notifications you received from AWS through the AWS Support Center and/or contact AWS Support to open a support case

CLOUD SECURITY PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Which AWS service gives you centralized control over the encryption keys used to protect your data?

A. AWS STS

B. AWS KMS

C. AWS DMS

D. Amazon EBS

Question 2:

Which services are involved with security? (choose 2)

A. AWS CloudHSM

B. AWS DMS

C. AWS KMS

D. AWS SMS

E. Amazon ELB

Question 3:

How can a security compliance officer retrieve AWS compliance documentation such as a SOC 2 report?

A. Using AWS Artifact

B. Using AWS Trusted Advisor

C. Using AWS Inspector

D. Using the AWS Personal Health Dashboard

Question 4:

Which information security standard applies to entities that store, process or transmit credit cardholder data?

A. ISO 27001

B. HIPAA

C. NIST

D. PCI DSS

Question 1 answer: B

Explanation:

- AWS Key Management Service gives you centralized control over the encryption keys used to protect your data. You can create, import, rotate, disable, delete, define usage policies for, and audit the use of encryption keys used to encrypt your data
- The AWS Security Token Service (STS) is a web service that enables you to request temporary, limited-privilege credentials for AWS Identity and Access Management (IAM) users
- AWS Database Migration Service (DMS) helps you migrate databases to AWS quickly and securely
- Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use withAmazon EC2instances in the AWS Cloud

Question 2 answer: A,C

Explanation:

- AWS Key Management Service gives you centralized control over the encryption keys used to protect your data
- AWS CloudHSM is a cloud-based hardware security module (HSM) that enables you to easily generate and use your own encryption keys on the AWS Cloud
- AWS Database Migration Service and Server Migration Service are used for migration
- Amazon Elastic Load Balancing is used for distributing incoming connections to pools of EC2 instances

Question 3 answer: A

Explanation:

- AWS Artifact, available in the console, is a self-service audit artifact retrieval portal that provides our customers with on-demand access to AWS' compliance documentation and AWS agreements
- You can use AWS Artifact Reports to download AWS security and compliance documents, such as AWS ISO certifications, Payment Card Industry (PCI), and System and Organization Control (SOC) reports
- AWS Trusted Advisor is an online resource to help you reduce cost, increase performance, and improve security by optimizing your AWS environment
- Inspector is an automated security assessment service that helps improve the security and compliance of applications deployed on AWS
- AWS Personal Health Dashboard provides alerts and remediation guidance when AWS is experiencing events that may impact you

Question 4 answer: *D*

Explanation:

- The Payment Card Industry Data Security Standard (PCI DSS) is a proprietary information security standard administered by thePCI Security Standards Council
- AWS enables covered entities and their business associates subject to the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA) to use the secure AWS environment to process, maintain, and store protected health information
- TheNational Institute of Standards and Technology (NIST) 800-53 security controls are generally applicable to US Federal Information Systems
- ISO/IEC 27001:2013 is a security management standard that specifies security management best practices and comprehensive security controls following the ISO/IEC 27002 best practice guidance

SHARED RESPONSIBILITY MODEL

The shared responsibility model defines what you (as an AWS account holder/user) and AWS are responsible for when it comes to security and compliance

Security and Compliance is a shared responsibility between AWS and the customer. This shared model can help relieve customer's operational burdens as 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 service operates

The customer assumes responsibility and management of the guest operating system (including updates and security patches), other associated application software as well as the configuration of the AWS provided security group firewall

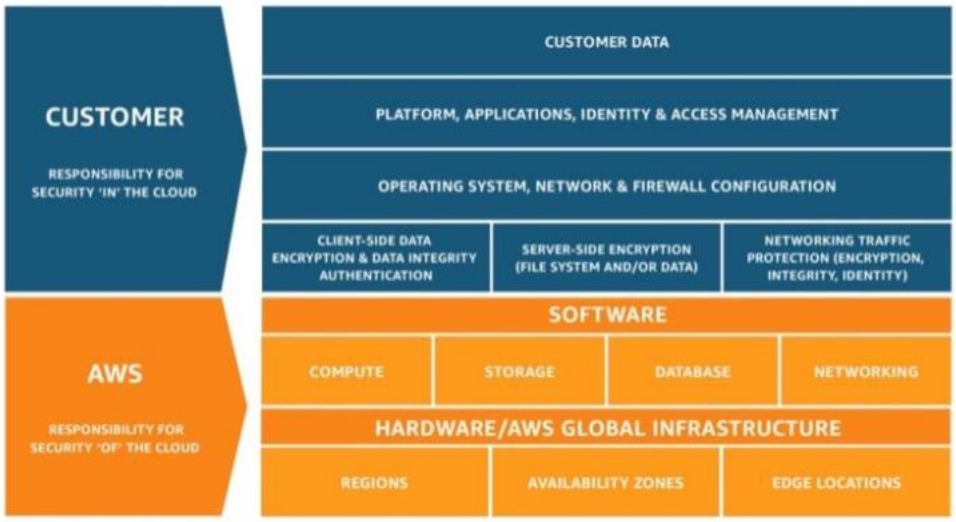
AWS are responsible for "Security **of** the Cloud"

- AWS is responsible for protecting the infrastructure that runs all of the services offered in the AWS Cloud
- This infrastructure is composed of the hardware, software, networking, and facilities that run AWS Cloud services

Customers are responsible for "Security **in** the Cloud"

- For EC2 this includes network level security (NACLs, security groups), operating system patches and updates, IAM user access management, and client and server side data encryption

The following diagram shows the split of responsibilities between AWS and the customer:



Inherited Controls – Controls which a customer fully inherits from AWS

- Physical and Environmental controls

Shared Controls – Controls which apply to both the infrastructure layer and customer layers, but in completely separate contexts or perspectives

In a shared control, AWS provides the requirements for the infrastructure and the customer must provide their own control implementation within their use of AWS services.

Examples of shared controls include:

- **Patch Management** – AWS is responsible for patching and fixing flaws within the infrastructure, but customers are responsible for patching their guest OS and applications
- **Configuration Management** – AWS maintains the configuration of its infrastructure devices, but a customer is responsible for configuring their own guest operating systems, databases, and applications
- **Awareness & Training** – AWS trains AWS employees, but a customer must train their own employees

Customer Specific – Controls which are solely the responsibility of the customer based on the application they are deploying within AWS services.

Examples of customer specific controls include:

- Service and Communications Protection or Zone Security which may require a customer to route or zone data within specific security environments

SHARED RESPONSIBILITY MODEL PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Under the AWS shared responsibility model what is AWS responsible for? (choose 2)

A. Physical security of the data center

B. Replacement and disposal of disk drives

C. Configuration of security groups

D. Patch management of operating systems

E. Encryption of customer data

Question 2:

Which statement is correct in relation to the AWS Shared Responsibility Model?

A. Customers are responsible for security of the cloud

B. AWS are responsible for encrypting customer data

C. Customers are responsible for patching storage systems

D. AWS are responsible for the security of regions and availability zones

Question 3:

Under the AWS shared responsibility model what is the customer responsible for? (choose 2)

A. Physical security of the data center

B. Replacement and disposal of disk drives

C. Configuration of security groups

D. Patch management of infrastructure

E. Encryption of customer data

Question 4:

Under the AWS Shared Responsibility Model, who is responsible for what? (choose 2)

A. Customers are responsible for compute infrastructure

B. AWS are responsible for network and firewall configuration

C. Customers are responsible for networking traffic protection

D. AWS are responsible for networking infrastructure

E. Customers are responsible for edge locations

Question 1 answer: A,B

Explanation:

- AWS are responsible for "Security **of** the Cloud"
- Customers are responsible for "Security **in** the Cloud"
- AWS are responsible for items such as the physical security of the DC, replacement of old disk drives, and patch management of the infrastructure
- Customers are responsible for items such as configuring security groups, network ACLs, patching their operating systems and encrypting their data

Question 2 answer: D

Explanation:

- AWS are responsible for "Security of the Cloud". AWS is responsible for protecting the infrastructure that runs all of the services offered in the AWS Cloud. This infrastructure is

composed of the hardware, software, networking, and facilities that run AWS Cloud services, and this includes regions, availability zones and edge locations

- Customers are responsible for "Security in the Cloud". This includes encrypting customer data, patching operating systems but not patching or maintaining the underlying infrastructure

Question 3 answer: *C,E*

Explanation:

- AWS are responsible for "Security **of** the Cloud"
- Customers are responsible for "Security **in** the Cloud"
- AWS are responsible for items such as the physical security of the DC, replacement of old disk drives, and patch management of the infrastructure
- Customers are responsible for items such as configuring security groups, network ACLs, patching their operating systems and encrypting their data

Question 4 answer: *C,D*

Explanation:

- Customers are responsible for networking traffic protection
- AWS are responsible for networking infrastructure
- AWS are responsible for compute infrastructure
- Customers are responsible for network and firewall configuration
- AWS are responsible for edge locations

ARCHITECTING FOR THE CLOUD

THE CLOUD COMPUTING DIFFERENCE

Architecting for the Cloud is one of the key subjects tested on the Cloud Practitioner exam. The information on this page has been extracted from the AWS whitepaper "Architecting for The Cloud: Best Practices" which can be downloaded from this link.

Also, please read the following AWS Blog article: https://aws.amazon.com/blogs/apn/the-5-pillars-of-the-aws-well-architected-framework/

Cloud computing differs from a traditional environment in the following ways:

IT assets become programmable resources

On AWS, servers, databases, storage, and higher-level application components can be instantiated within seconds

You can treat these as temporary and disposable resources, free from the inflexibility and constraints of a fixed and finite IT infrastructure

This resets the way you approach change management, testing, reliability, and capacity planning

Global, available and unlimited capacity

Using the global infrastructure of AWS, you can deploy your application to the AWS Region that best meets your requirements

For global applications, you can reduce latency to end users around the world by using the Amazon CloudFront content delivery network

It is also much easier to operate production applications and databases across multiple data centers to achieve high availability and fault tolerance

Higher level managed services

AWS customers also have access to a broad set of compute, storage, database, analytics, application, and deployment services

These services are instantly available to developers and can reduce dependency on in-house specialized skills and allow organizations to deliver new solutions faster

These services are managed by AWS, which can lower operational complexity and cost

Security built-in

The AWS cloud provides governance capabilities that enable continuous monitoring of configuration changes to your IT resources

Since AWS assets are programmable resources, your security policy can be formalized and embedded with the design of your infrastructure

DESIGN PRINCIPLES

Scalability

Systems that are expected to grow over time need to be built on top of a scalable architecture

Scaling Vertically

Scaling vertically takes place through an increase in the specifications of an individual resource (e.g., upgrading a server with a larger hard drive or a faster CPU)

On Amazon EC2, this can easily be achieved by stopping an instance and resizing it to an instance type that has more RAM, CPU, IO, or networking capabilities

Scaling Horizontally

Scaling horizontally takes place through an increase in the number of resources (e.g., adding more hard drives to a storage array or adding more servers to support an application)

This is a great way to build Internet-scale applications that leverage the elasticity of cloud computing

The table below provides more information on the differences between horizontal and vertical scaling:

Horizontal Scaling	Vertical Scaling
Add more instances as demand increases	Add more CPU and/or RAM to existing instances as demand increases
No downtime required to scale up or down	Requires a restart to scale up or down
Automatic using services such as AWS Auto-Scaling	Would require scripting or automation tools to automate
Unlimited scalability	Scalability limited by maximum instance size

Stateless applications:

- A stateless application is an application that needs no knowledge of previous interactions and stores no session information
- A stateless application can scale horizontally since any request can be serviced by any of the available compute resources (e.g., EC2 instances, AWS Lambda functions)

Stateless components:

- Most applications need to maintain some kind of state information
- For example, web applications need to track whether a user is signed in, or else they might present personalized content based on previous actions
- Web applications can use HTTP cookies to store information about a session at the client's browser (e.g., items in the shopping cart)
- Consider only storing a unique session identifier in a HTTP cookie and storing more detailed user session information server-side
- DynamoDB is often used for storing session state to maintain a stateless architecture
- For larger files a shared storage system can be used such as S3 or EFS
- SWF can be used for a multi-step workflow

Stateful components:

- Databases are stateful
- Many legacy applications are stateful
- Load balancing with session affinity can be used for horizontal scaling of stateful components
- Session affinity is however not guaranteed and existing sessions do not benefit from newly launched nodes

Distributed processing:

- Use cases that involve processing of very large amounts of data (e.g., anything that can't be handled by a single compute resource in a timely manner) require a distributed processing approach
- By dividing a task and its data into many small fragments of work, you can execute each of them in any of a larger set of available compute resources

DISPOSABLE RESOURCES INSTEAD OF FIXED SERVERS

Think of servers and other components as temporary resources

Launch as many as you need, and use them only for as long as you need them

An issue with fixed, long-running servers is that of configuration drift (where change and software patches are applied over time)

This problem can be solved with the "immutable infrastructure" pattern where a server is never updated but instead is replaced with a new one as required

Instantiating compute resources

You don't want to manually set up new resources with their configuration and code

Use automated, repeatable processes that avoid long lead times and are not prone to human error

Bootstrapping:

- Execute automated bootstrapping actions to modify default configurations
- This includes scripts that install software or copy data to bring that resource to a particular state
- You can parameterize configuration details that vary between different environments

Golden Images:

- Some resource types can be launched from a golden image
- Examples are EC2 instances, RDS instances and EBS volumes
- A golden image is a snapshot of a particular state for that resource
- Compared to bootstrapping golden images provide faster start times and remove dependencies to configuration services or third-party repositories

Infrastructure are Code:

- AWS assets are programmable, so you can apply techniques, practices, and tools from software development to make your whole infrastructure reusable, maintainable, extensible, and testable

AUTOMATION

In a traditional IT infrastructure, you often have to manually react to a variety of events

When deploying on AWS there is a lot of opportunity for automation

This improves both your system's stability and the efficiency of your organization

Examples of automations using AWS services include:

- AWS Elastic Beanstalk – the fastest and simplest way to get an application up and running on AWS
- Amazon EC2 Auto Recovery – You can create an Amazon CloudWatch alarm that monitors an Amazon EC2 instance and automatically recovers it if it becomes impaired
- Auto Scaling – With Auto Scaling, you can maintain application availability and scale your Amazon EC2 capacity up or down automatically according to conditions you define

- Amazon CloudWatch Alarms – You can create a CloudWatch alarm that sends an Amazon Simple Notification Service (Amazon SNS) message when a particular metric goes beyond a specified threshold for a specified number of periods
- Amazon CloudWatch Events – The CloudWatch service delivers a near real-time stream of system events that describe changes in AWS resources
- AWS OpsWorks Lifecycle events – AWS OpsWorks supports continuous configuration through lifecycle events that automatically update your instances' configuration to adapt to environment changes
- AWS Lambda Scheduled events – These events allow you to create a Lambda function and direct AWS Lambda to execute it on a regular schedule

LOOSE COUPLING

As application complexity increases, a desirable attribute of an IT system is that it can be broken into smaller, loosely coupled components

This means that IT systems should be designed in a way that reduces interdependencies—a change or a failure in one component should not cascade to other components

Design principles include:

- **Well-defined interfaces** – reduce interdependencies in a system by enabling interaction only through specific, technology-agnostic interfaces (e.g. RESTful APIs)
- **Service discovery** – disparate resources must have a way of discovering each other without prior knowledge of the network topology
- **Asynchronous integration** – this is another form of loose coupling where an interaction does not need an immediate response (think SQS queue or Kinesis)
- **Graceful failure** – build applications such that they handle failure in a graceful manner (reduce the impact of failure and implement retries)

SERVICES, NOT SERVERS

With traditional IT infrastructure, organizations have to build and operate a wide variety of technology components

AWS offers a broad set of compute, storage, database, analytics, application, and deployment services that help organizations move faster and lower IT costs

Managed services:

- On AWS, there is a set of services that provide building blocks that developers can consume to power their applications

- These managed services include databases, machine learning, analytics, queuing, search, email, notifications, and more

Serverless architectures:

- Another approach that can reduce the operational complexity of running applications is that of the serverless architectures
- It is possible to build both event-driven and synchronous services for mobile, web, analytics, and the Internet of Things (IoT) without managing any server infrastructure

DATABASES

With traditional IT infrastructure, organizations were often limited to the database and storage technologies they could use

With AWS, these constraints are removed by managed database services that offer enterprise performance at open source cost

Relational Databases

Relational databases (often called RDBS or SQL databases) normalize data into well-defined tabular structures known as tables, which consist of rows and columns

They provide a powerful query language, flexible indexing capabilities, strong integrity controls, and the ability to combine data from multiple tables in a fast and efficient manner

Amazon RDS is a relational database service

Scalability:

- Relational databases can scale vertically (e.g. upgrading to a larger RDS DB instance)
- For read-heavy use cases, you can scale horizontally using read replicas
- For scaling write capacity beyond a single instance data partitioning or sharding is required

High Availability:

- For production DBs, Amazon recommend the use of RDS Multi-AZ which creates a synchronously replicated standby in another AZ
- With Multi-AZ RDS can failover to the standby node without administrative intervention

Anti-Patterns:

- If your application primarily indexes and queries data with no need for joins or complex transactions consider a NoSQL database instead

- If you have large binary files (audio, video, and image), it will be more efficient to store the actual files in S3 and only hold the metadata for the files in your database

NoSQL Databases

NoSQL is a term used to describe databases that trade some of the query and transaction capabilities of relational databases for a more flexible data model that seamlessly scales horizontally

NoSQL databases utilize a variety of data models, including graphs, key-value pairs, and JSON documents

DynamoDB is Amazon's NoSQL database service

Scalability:

- NoSQL database engines will typically perform data partitioning and replication to scale both the reads and the writes in a horizontal fashion

High Availability:

- DynamoDB synchronously replicates data across three facilities in an AWS region for fault tolerance

Anti-Patterns:

- If your schema cannot be denormalized and your application requires joins or complex transactions, consider a relational database instead
- If you have large binary files (audio, video, and image), consider storing the files in Amazon S3 and storing the metadata for the files in your database

Data Warehouse

A data warehouse is a specialized type of relational database, optimized for analysis and reporting of large amounts of data

It can be used to combine transactional data from disparate sources making them available for analysis and decision-making

Amazon Redshift is a managed data warehouse service that is designed to operate at less than a tenth the cost of traditional solutions

Scalability:

- Amazon Redshift achieves efficient storage and optimum query performance through a combination of massively parallel processing (MPP), columnar data storage, and targeted data compression encoding schemes

- RedShift is particularly suited to analytic and reporting workloads against very large data sets

High Availability:

- Redshift has multiple features that enhance the reliability of your data warehouse cluster
- Multi-node clusters replicate data to other nodes within the cluster
- Data is continuously backed up to S3
- RedShift continuously monitors the health of the cluster and re-replicates data from failed drives and replaces nodes as necessary

Anti-Patterns:

- Because Amazon Redshift is a SQL-based relational database management system (RDBMS), it is compatible with other RDBMS applications and business intelligence tools
- Although Amazon Redshift provides the functionality of a typical RDBMS, including online transaction processing (OLTP) functions, it is not designed for these workloads

Search

Applications that require sophisticated search functionality will typically outgrow the capabilities of relational or NoSQL databases

A search service can be used to index and search both structured and free text format and can support functionality that is not available in other databases, such as customizable result ranking, faceting for filtering, synonyms, stemming, etc.

Scalability:

- Both Amazon CloudSearch and Amazon ES use data partitioning and replication to scale horizontally

High Availability:

- Both services provide features that store data redundantly across Availability Zones

REMOVING SINGLE POINTS OF FAILURE

A system is highly available when it can withstand the failure of an individual or multiple components

Automate recovery and reduce disruption at every layer of your architecture

Introducing Redundancy

Single points of failure can be removed by introducing redundancy

In standby redundancy when a resource fails, functionality is recovered on a secondary resource using a process called failover, which typically take some time to complete

In active redundancy, requests are distributed to multiple redundant compute resources, and when one of them fails, the rest can simply absorb a larger share of the workload

Detect Failure

Build as much automation as possible in both detecting and reacting to failure

Services like ELB and Route53 mask failure by routing traffic to healthy endpoint

Auto Scaling can be configured to automatically replace unhealthy nodes

You can also replace unhealthy nodes using the EC2 auto- recovery, OpsWorks and Elastic Beanstalk

Durable Data Storage

Design your architecture to protect both data availability and integrity

Data replication is the technique that introduces redundant copies of data

It can help horizontally scale read capacity, but it also increase data durability and availability

Replication can take place in a few different modes:

- Synchronous replication – transactions are acknowledged only after data has been durably stored in both the primary and replica instance. Can be used to protect data integrity (low RPO) and scaling read capacity (with strong consistency)
- Asynchronous replication – changes on the primary node are not immediately reflected on its replicas. Can be used to horizontally scale the system's read capacity (with replication lag), and data durability (with some data loss)
- Quorum-based replication – combines synchronous and asynchronous replication and is good for large-scale distributed database systems

Automated Multi-Data Center Resilience

With traditional infrastructure, failing over between data centers is performed using a disaster recovery plan

Long distances between data centers mean that latency makes synchronous replication impractical

Failovers often lead to data loss and costly data recovery processes

On AWS it is possible to adopt a simpler, more efficient protection from this type of failure

Each AWS region contains multiple distinct locations called Availability Zones (AZs).

Each AZ is engineered to be isolated from failures in other AZs

An AZ is a data center, and in some cases, an AZ consists of multiple data centers

AZs within a region provide inexpensive, low-latency network connectivity to other zones in the same region

This allows you to replicate your data across data centers in a synchronous manner so that failover can be automated and be transparent for your users

Fault Isolation and Traditional Horizontal Scaling

Though the active redundancy pattern is great for balancing traffic and handling instance or Availability Zone disruptions, it is not sufficient if there is something harmful about the requests themselves

If a particular request happens to trigger a bug that causes the system to fail over, then the caller may trigger a cascading failure by repeatedly trying the same request against all instances

One fault-isolating improvement you can make to traditional horizontal scaling is called sharding

Similar to the technique traditionally used with data storage systems, instead of spreading traffic from all customers across every node, you can group the instances into shards

In this way, you are able to reduce the impact on customers in direct proportion to the number of shards you have

Optimize for Cost

Just by moving existing architectures into the cloud, organizations can reduce capital expenses and drive savings as a result of the AWS economies of scale

By iterating and making use of more AWS capabilities there is further opportunity to create cost-optimized cloud architectures

Right Sizing:

- In some cases, you should select the cheapest type that suits your workload's requirements
- In other cases, using fewer instances of a larger instance type might result in lower total cost or better performance
- Benchmark and select the right instance type depending on how your workload utilizes CPU, RAM, network, storage size, and I/O

- Reduce cost by selecting the right storage solution for your needs
- E.g. S3 offers a variety of storage classes, including Standard, Reduced Redundancy, and Standard-Infrequent Access
- EC2, RDS, and ES support different EBS volume types (magnetic, general purpose SSD, provisioned IOPS SSD) that you should evaluate

Elasticity:

- Plan to implement Auto Scaling for as many EC2 workloads as possible, so that you horizontally scale up when needed and scale down automatically to reduce cost
- Automate turning off non-production workloads when not in use
- Where possible, replace EC2 workloads with AWS managed services that don't require you to take any capacity decisions. For example:
    - ELB
    - CloudFront
    - SQS
    - Kinesis Firehose
    - Lambda
    - SES
    - CloudSearch
- Or use services for which you can modify capacity as and when need. For example:
    - DynamoDB
    - RDS
    - Elasticsearch Service

Take Advantage of the Variety of Purchasing Options:

- EC2 On-Demand instance pricing gives you maximum flexibility with no long term commitments
- There are two more ways to pay for Amazon EC2 instances that can help you reduce spend: Reserved Instances and Spot Instances

Reserved Capacity

EC2 Reserved Instances allow you to reserve Amazon EC2 computing capacity in exchange for a significantly discounted hourly rate compared to On- Demand instance pricing

This is ideal for applications with predictable minimum capacity requirements

Spot Instances

For less steady workloads, you can consider the use of Spot Instances

EC2 Spot Instances allow you to bid on spareEC2 computing capacity

Since Spot Instances are often available at a discount compared to On-Demand pricing, you can significantly reduce the cost of running your applications

Spot Instances are ideal for workloads that have flexible start and end times

If the Spot market price increases above your bid price, your instance will be terminated automatically and you will not be charged for the partial hour that your instance has run

As a result, Spot Instances are great for workloads that have tolerance to interruption

CACHING

Caching is a technique that stores previously calculated data for future use

This technique is used to improve application performance and increase the cost efficiency of an implementation

It can be applied at multiple layers of an IT architecture

Application Data Caching

Applications can be designed so that they store and retrieve information from fast, managed, in-memory caches

Cached information may include the results of I/O-intensive database queries or the outcome of computationally intensive processing

Edge Caching

Copies of static content and dynamic content can be cached at Amazon CloudFront, which is a content delivery network (CDN) consisting of multiple edge locations around the world

Edge caching allows content to be served by infrastructure that is closer to viewers, lowering latency and giving you the high, sustained data transfer rates needed to deliver large popular objects to end users at scale

SECURITY

Most of the security tools and techniques that you might already be familiar with in a traditional IT infrastructure can be used in the cloud

At the same time, AWS allows you to improve your security in a variety of ways

AWS is a platform that allows you to formalize the design of security controls in the platform itself

Utilize AWS Features for Defence in Depth

Network level security includes building a VPC topology that isolates parts of the infrastructure through the use of subnets, security groups, and routing controls

Services like AWS WAF, a web application firewall, can help protect web applications from SQL injection and other vulnerabilities in application code

For access control, you can use IAM to define a granular set of policies and assign them to users, groups, and AWS resources

Finally, the AWS platform offers a breadth of options for protecting data, whether it is in transit or at rest with encryption

Offload Security Responsibility to AWS

AWS operates under a shared security responsibility model, where AWS is responsible for the security of the underlying cloud infrastructure and you are responsible for securing the workloads you deploy in AWS

Reduce Privileged Access

When you treat servers as programmable resources, you can capitalize on that for benefits in the security space as well

Eliminate the need for guest operating system access to production environments

If an instance experiences an issue you can automatically or manually terminate and replace it

In a traditional environment, service accounts would often be assigned long-term credentials stored in a configuration file

On AWS, you can instead use IAM roles to grant permissions to applications running on Amazon EC2 instances through the use of short-term credentials

Security as Code

Traditional security frameworks, regulations, and organizational policies define security requirements related to things such as firewall rules, network access controls, internal/external subnets, and operating system hardening

You can implement these in an AWS environment as well, but you now have the opportunity to capture them all in a script that defines a “Golden Environment.”

This means you can create an AWS CloudFormation script that captures your security policy and reliably deploys it

Security best practices can now be reused among multiple projects and become part of your continuous integration pipeline

You can perform security testing as part of your release cycle, and automatically discover application gaps and drift from your security policy

Real-Time Auditing

Testing and auditing your environment is key to moving fast while staying safe

Traditional approaches that involve periodic checks are not sufficient, especially in agile environments where change is constant

On AWS, it is possible to implement continuous monitoring and automation of controls to minimize exposure to security risks

Services like AWS Config, Amazon Inspector, and AWS Trusted Advisor continually monitor for compliance or vulnerabilities

With AWS Config rules you will also know if some component was out of compliance even for a brief period of time

You can implement extensive logging for your applications (using Amazon CloudWatch Logs) and for the actual AWS API calls by enabling AWS CloudTrail

Logs can then be stored in an immutable manner and automatically processed to either notify or even take action on your behalf, protecting your organization from non-compliance

You can use AWS Lambda, Amazon EMR, the Amazon Elasticsearch Service, or third- party tools from the AWS Marketplace to scan logs to detect things like unused permissions, overuse of privileged accounts, usage of keys, anomalous logins, policy violations, and system abuse

ARCHITECTING FOR THE CLOUD PRACTICE QUESTIONS

Answers and explanations are provided below after the last question in this section.

Question 1:

Which type of scaling does AWS Auto Scaling provide?

A. Vertical

B. Linear

C. Horizontal

D. Incremental

Question 2:

How can a Solutions Architect reduce the latency between end-users and applications or content? (choose 2)

A. Deploy applications in multiple AZs

B. Deploy applications in regions closest to the end-users

C. Use S3 Transfer Acceleration to improve application performance

D. Use Amazon CloudFront to cache content closer to end-users

E. Use larger EC2 instance types for the applications

Question 3:

Which of the following constitute the five pillars for the AWS Well-Architected Framework? (choose 2)

A. Operational excellence, security, and reliability

B. Operational excellence, elasticity and scalability

C. Cost prioritization, and cost optimization

D. Data consistency, and cost optimization

E. Performance efficiency, and cost optimization

Question 4:

Assuming you have configured them correctly, which AWS services can scale automatically without intervention? (choose 2)

A. Amazon RDS

B. Amazon EC2

C. Amazon S3

D. Amazon DynamoDB

E. Amazon EBS

Question 1 answer: C

Explanation:

- AWS Auto Scaling scales horizontally by adding additional compute instances

Question 2 answer: B,D

Explanation:

- To reduce latency, which corresponds with the distance over which network communications travel, you should aim to host your applications closer to your end-users. This means deploying them in the closest regions
- Deploying in multiple AZs may create resiliency but won't change latency much as AZs are geographically close to each other
- S3 Transfer Acceleration is used to improve upload speeds for S3 objects and does not affect application performance
- CloudFormation is used for deploying resources through code ("infrastructure as code")
- Using a larger instance type for your application may improve application performance but will not reduce latency

Question 3 answer: *A,E*

Explanation:

- The five pillars of the AWS Well-Architected Framework are operational excellence, security, reliability, performance efficiency, and cost optimization

Question 4 answer: *C,D*

Explanation:

- Both S3 and DynamoDB automatically scale as demand dictates. In the case of DynamoDB you can either configure the on-demand or provisioned capacity mode. With on-demand capacity mode DynamoDB automatically adjusts the read and write throughput for you
- EC2 cannot scale automatically. You need to use Auto Scaling to scale the number of EC2 instances deployed
- EBS and RDS do not scale automatically. You must intervene to adjust volume sizes and database instance types to scale these resources

ADDITIONAL TOOLS AND SERVICES

There are Additional AWS Services & Tools that may feature on the exam. Often you do not need to know these at a deep level but do need to understand what they are and what they are used for.

On this page I have listed some high-level details and links for more information for some of these services and tools.

Exam tip: Before sitting the exam it would be wise to go through the AWS console and pick out any services you're not familiar with and do a bit of reading up on them using the AWS documentation.

COMPUTE

Amazon Elastic Container Service for Kubernetes (EKS)

- Amazon Elastic Container Service for Kubernetes (EKS) is a managed Kubernetes service that makes it easy for you to run Kubernetes on AWS without needing to install, operate, and maintain your own Kubernetes control plane
- EKS is certified Kubernetes conformant, so existing applications running on upstream Kubernetes are compatible with Amazon EKS
- EKS automatically manages the availability and scalability of the Kubernetes control plane nodes that are responsible for starting and stopping containers, scheduling containers on virtual machines, storing cluster data, and other tasks
- EKS automatically detects and replaces unhealthy control plane nodes for each cluster
- Generally available but only in limited regions currently
- https://aws.amazon.com/eks/features/

AWS Batch

- With AWS Batch, you simply package the code for your batch jobs, specify their dependencies, and submit your batch job using the AWS Management Console, CLIs, or SDKs
- AWS Batch allows you to specify execution parameters and job dependencies, and facilitates integration with a broad range of popular batch computing workflow engines and languages (e.g., Pegasus WMS, Luigi, and AWS Step Functions)
- AWS Batch efficiently and dynamically provisions and scales Amazon EC2 and Spot Instances based on the requirements of your jobs. AWS Batch provides default job queues and compute environment definitions that enable you to get started quickly
- https://aws.amazon.com/batch/features/

AWS Elastic Beanstalk

- AWS Elastic Beanstalk is the fastest and simplest way to get web applications up and running on AWS
- Developers simply upload their application code and the service automatically handles all the details such as resource provisioning, load balancing, auto-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), Auto Scaling, and Elastic Load Balancing to easily support applications that need to scale to serve millions of users
- https://digitalcloud.training/certification-training/aws-solutions-architect-associate/compute/aws-elastic-beanstalk/
- https://aws.amazon.com/elasticbeanstalk/details/

STORAGE

AWS Storage Gateway

- AWS Storage Gateway is a hybrid cloud storage service that connects your existing on-premises environments with the AWS Cloud
- Its features make it easy for you to run hybrid cloud workloads at any stage of your cloud adoption, whether it's getting started with cloud backups, running cloud processing workflows for data generated by on-premises machines, or performing a one-time migration of block volume data or databases
- Storage Gateway seamlessly connects to your local production or backup applications with NFS, SMB, iSCSI, or iSCSI-VTL, so you can adopt AWS Cloud storage without needing to modify your applications
- Its protocol conversion and device emulation enables you to access block data on volumes managed by Storage Gateway on top of Amazon S3, store files as native Amazon S3 objects, and keep virtual tape backups online in a Virtual Tape Library backed by S3 or move the backups to a tape archive tier on Amazon Glacier
- https://digitalcloud.training/certification-training/aws-solutions-architect-associate/storage/aws-storage-gateway/
- https://aws.amazon.com/storagegateway/features/

DATABASE

Amazon Elasticache

- Amazon ElastiCache offers fully managed Redis and Memcached
- Seamlessly deploy, run, and scale popular open source compatible in-memory data stores
- Amazon ElastiCache can be used to significantly improve latency and throughput for many read-heavy application workloads (such as social networking, gaming, media sharing and Q&A portals) or compute-intensive workloads (such as a recommendation engine) by allowing you to store the objects that are often read in cache
- Amazon ElastiCache simplifies and offloads the management, monitoring, and operation of in-memory cache environments, enabling you to focus on the differentiating parts of your applications
- Pay only for the resources you consume based on node hours used
- https://aws.amazon.com/elasticache/features/

Amazon Neptune

- 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
- With Amazon Neptune, you can create sophisticated, interactive graph applications that can query billions of relationships in milliseconds
- SQL queries for highly connected data are complex and hard to tune for performance. Instead, Amazon Neptune allows you to use the popular graph query languages Apache TinkerPop Gremlin and W3C's SPARQL to execute powerful queries that are easy to write and perform well on connected data
- https://aws.amazon.com/neptune/features/

MIGRATION

AWS Migration Hub

- AWS Migration Hub provides a single location to track the progress of application migrations across multiple AWS and partner solutions
- Using Migration Hub allows you to choose the AWS and partner migration tools that best fit your needs, while providing visibility into the status of migrations across your portfolio of applications
- For example, you might use AWS Database Migration Service, AWS Server Migration Service, and partner migration tools such as ATADATA ATAmotion, CloudEndure Live Migration, or RiverMeadow Server Migration SaaS to migrate an application comprised of a database, virtualized web servers, and a bare metal server

- Using Migration Hub, you can view the migration progress of all the resources in the application
- https://aws.amazon.com/migration-hub/features/

AWS Database Migration Service

- AWS Database Migration Service helps you migrate databases to AWS quickly and securely
- The source database remains fully operational during the migration, minimizing downtime to applications that rely on the database
- The AWS Database Migration Service can migrate your data to and from most widely used commercial and open-source databases
- AWS Database Migration Service supports homogenous migrations such as Oracle to Oracle, as well as heterogeneous migrations between different database platforms, such as Oracle or Microsoft SQL Server to Amazon Aurora
- With AWS Database Migration Service, 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
- https://aws.amazon.com/dms/

AWS Server Migration Service

- AWS Server Migration Service (SMS) is an agentless service which makes it easier and faster for you to migrate thousands of on-premises workloads to AWS
- AWS SMS allows you to automate, schedule, and track incremental replications of live server volumes, making it easier for you to coordinate large-scale server migrations
- https://aws.amazon.com/server-migration-service/

NETWORKING & CONTENT DELIVERY

Amazon API Gateway

- Amazon API Gateway is a fully managed service that makes it easy for developers to create, publish, maintain, monitor, and secure APIs at any scale
- With a few clicks in the AWS Management Console, you can create an API that acts as a "front door" for applications to access data, business logic, or functionality from your back-end services
- Back-end services may include Amazon Elastic Compute Cloud (Amazon EC2), code running on AWS Lambda, or any web application
- https://aws.amazon.com/api-gateway/features/

AWS Direct Connect

- 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 datacenter, office, or colocation environment, which in many cases 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 industry standard 802.1q VLANs, this dedicated connection can be partitioned into multiple virtual interfaces
- This allows you to use the same connection to access public resources such as objects stored in Amazon S3 using 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
- https://aws.amazon.com/directconnect/features/

DEVELOPER TOOLS

AWS CodeStar

- AWS CodeStar enables you to quickly develop, build, and deploy applications on AWS. AWS CodeStar provides a unified user interface, enabling you to easily manage your software development activities in one place
- With AWS CodeStar, you can set up your entire continuous delivery toolchain in minutes, allowing you to start releasing code faster. AWS CodeStar makes it easy for your whole team to work together securely, allowing you to easily manage access and add owners, contributors, and viewers to your projects
- With AWS CodeStar, you can use a variety of project templates to start developing applications on Amazon EC2, AWS Lambda, and AWS Elastic Beanstalk
- AWS CodeStar projects support many popular programming languages including Java, JavaScript, PHP, Ruby, and Python
- https://aws.amazon.com/codestar/features/

AWS CodeCommit

- AWS CodeCommit is a fully-managed source control service that hosts secure Git-based repositories
- It makes it easy for teams to collaborate on code in a secure and highly scalable ecosystem.

- CodeCommit eliminates the need to operate your own source control system or worry about scaling its infrastructure
- You can use CodeCommit to securely store anything from source code to binaries, and it works seamlessly with your existing Git tools
- https://aws.amazon.com/codecommit/features/

AWS CodeBuild

- AWS CodeBuild is a fully managed continuous integration service that compiles source code, runs tests, and produces software packages that are ready to deploy
- With CodeBuild, you don't need to provision, manage, and scale your own build servers. CodeBuild scales continuously and processes multiple builds concurrently, so your builds are not left waiting in a queue
- You can get started quickly by using pre-packaged build environments, or you can create custom build environments that use your own build tools
- With CodeBuild, you are charged by the minute for the compute resources you use
- https://aws.amazon.com/codebuild/features/

AWS CodeDeploy

- AWS CodeDeploy is a fully managed deployment service that automates software deployments to a variety of compute services such as Amazon EC2, AWS Lambda, and your on-premises servers
- AWS CodeDeploy makes it easier for you to rapidly release new features, helps you avoid downtime during application deployment, and handles the complexity of updating your applications
- You can use AWS CodeDeploy to automate software deployments, eliminating the need for error-prone manual operations. The service scales to match your deployment needs, from a single Lambda function to thousands of EC2 instances
- https://aws.amazon.com/codedeploy/features/

AWS CodePipeline

- AWS CodePipeline is a fully managed continuous delivery service that helps you automate your release pipelines for fast and reliable application and infrastructure updates
- CodePipeline automates the build, test, and deploy phases of your release process every time there is a code change, based on the release model you define
- This enables you to rapidly and reliably deliver features and updates
- You can easily integrate AWS CodePipeline with third-party services such as GitHub or with your own custom plugin
- https://aws.amazon.com/codepipeline/features/

AWS X-Ray

- AWS X-Ray helps developers analyze and debug production, distributed applications, such as those built using a microservices architecture
- With X-Ray, you can understand how your application and its underlying services are performing to identify and troubleshoot the root cause of performance issues and errors
- X-Ray provides an end-to-end view of requests as they travel through your application, and shows a map of your application's underlying components
- You can use X-Ray to analyze both applications in development and in production, from simple three-tier applications to complex microservices applications consisting of thousands of service
- https://aws.amazon.com/xray/features/

MANAGEMENT TOOLS

AWS CloudFormation

- AWS CloudFormation provides a common language for you to describe and provision all the infrastructure resources in your cloud environment
- CloudFormation allows you to use a simple text file to model and provision, in an automated and secure manner, all the resources needed for your applications across all regions and accounts
- This file serves as the single source of truth for your cloud environment
- You can use JSON or YAML to describe what AWS resources you want to create and configure
- https://aws.amazon.com/cloudformation/features/
- https://digitalcloud.training/certification-training/aws-solutions-architect-associate/management-tools/aws-cloudformation/

AWS Config

- AWS Config is a service that enables you to assess, audit, and evaluate the configurations of your AWS resources
- Config continuously monitors and records your AWS resource configurations and allows you to automate the evaluation of recorded configurations against desired configurations
- With Config, you can review changes in configurations and relationships between AWS resources, dive into detailed resource configuration histories, and determine your overall compliance against the configurations specified in your internal guidelines
- This enables you to simplify compliance auditing, security analysis, change management, and operational troubleshooting
- https://aws.amazon.com/config/features/

AWS OpsWorks

- AWS OpsWorks is a configuration management service that provides managed instances of Chef and Puppet
- Chef and Puppet are automation platforms that allow you to use code to automate the configurations of your servers
- OpsWorks lets you use Chef and Puppet to automate how servers are configured, deployed, and managed across your Amazon EC2 instances or on-premises compute environments
- OpsWorks has three offerings, AWS Opsworks for Chef Automate, AWS OpsWorks for Puppet Enterprise, and AWS OpsWorks Stacks

AWS Service Catalog

- AWS Service Catalog allows organizations to create and manage catalogs of IT services that are approved for use on AWS
- These IT services can include everything from virtual machine images, servers, software, and databases to complete multi-tier application architectures
- AWS Service Catalog allows you to centrally manage commonly deployed IT services, and helps you achieve consistent governance and meet your compliance requirements, while enabling users to quickly deploy only the approved IT services they need
- Uses CloudFormation templates
- https://aws.amazon.com/servicecatalog/features/

AWS Systems Manager

- AWS Systems Manager gives you visibility and control of your infrastructure on AWS
- Systems Manager provides a unified user interface so you can view operational data from multiple AWS services and allows you to automate operational tasks across your AWS resources
- With Systems Manager, you can group resources, like Amazon EC2 instances, Amazon S3 buckets, or Amazon RDS instances, by application, view operational data for monitoring and troubleshooting, and take action on your groups of resources
- Systems Manager simplifies resource and application management, shortens the time to detect and resolve operational problems, and makes it easy to operate and manage your infrastructure securely at scale
- https://aws.amazon.com/systems-manager/features/

AWS Managed Services

- AWS Managed Services provides ongoing management of your AWS infrastructure so you can focus on your applications
- By implementing best practices to maintain your infrastructure, AWS Managed Services helps to reduce your operational overhead and risk

- AWS Managed Services automates common activities such as change requests, monitoring, patch management, security, and backup services, and provides full-lifecycle services to provision, run, and support your infrastructure
- AWS Managed Services delivers consistent operations management and predictable results by following ITIL® best practices, and provides tooling and automation to increase efficiency, and reduce your operational overhead and risk
- https://aws.amazon.com/managed-services/#

ANALYTICS

Amazon Athena

- Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 using standard SQL
- Athena is serverless, so there is no infrastructure to manage, and you pay only for the queries that you run
- With a few clicks in the AWS Management Console, customers can point Athena at their data stored in S3 and begin using standard SQL to run ad-hoc queries and get results in seconds
- You can use Athena to process logs, perform ad-hoc analysis, and run interactive queries
- Athena scales automatically – executing queries in parallel – so results are fast, even with large datasets and complex queries
- https://aws.amazon.com/athena/features/

Amazon EMR

- Amazon Elastic Map Reduce (EMR) provides a managed Hadoop framework that makes it easy, fast, and cost-effective to process vast amounts of data across dynamically scalable Amazon EC2 instances
- You can also run other popular distributed frameworks such as Apache Spark, HBase, Presto, and Flink in Amazon EMR, and interact with data in other AWS data stores such as Amazon S3 and Amazon DynamoDB
- Amazon EMR securely and reliably handles a broad set of big data use cases, including log analysis, web indexing, data transformations (ETL), machine learning, financial analysis, scientific simulation, and bioinformatic
- https://aws.amazon.com/emr/features/

Amazon CloudSearch

- Amazon CloudSearch is a managed service in the AWS Cloud that makes it simple and cost-effective to set up, manage, and scale a search solution for your website or application.

- Amazon CloudSearch supports 34 languages and popular search features such as highlighting, autocomplete, and geospatial search
- https://aws.amazon.com/cloudsearch/

Amazon Elasticsearch

- Amazon Elasticsearch Service, is a fully managed service that makes it easy for you to deploy, secure, operate, and scale Elasticsearch to search, analyze, and visualize data in real-time
- With Amazon Elasticsearch Service you get easy-to-use APIs and real-time analytics capabilities to power use-cases such as log analytics, full-text search, application monitoring, and clickstream analytics, with enterprise-grade availability, scalability, and security
- https://aws.amazon.com/elasticsearch-service/features/

Amazon Kinesis

- Amazon Kinesis makes it easy to collect, process, and analyze real-time, streaming data so you can get timely insights and react quickly to new information
- There are four types of Kinesis service:
    - Kinesis Video Streams makes it easy to securely stream video from connected devices to AWS for analytics, machine learning (ML), and other processing
    - Kinesis Data Streams enables you to build custom applications that process or analyze streaming data for specialized needs
    - Kinesis Data Firehose is the easiest way to load streaming data into data stores and analytics tools
    - Amazon Kinesis Data Analytics is the easiest way to process and analyze real-time, streaming data
- https://aws.amazon.com/kinesis/
- https://digitalcloud.training/certification-training/aws-solutions-architect-associate/analytics/amazon-kinesis/

AWS Data Pipeline

- AWS Data Pipeline is a web service that helps you reliably process and move data between different AWS compute and storage services, as well as on-premises data sources, at specified intervals
- With AWS Data Pipeline, you can regularly access your data where it's stored, transform and process it at scale, and efficiently transfer the results to AWS services such as Amazon S3, Amazon RDS, Amazon DynamoDB, and Amazon EMR
- AWS Data Pipeline helps you easily create complex data processing workloads that are fault tolerant, repeatable, and highly available
- https://aws.amazon.com/datapipeline/

AWS Glue

- AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data for analytics
- You can create and run an ETL job with a few clicks in the AWS Management Console
- You simply point AWS Glue to your data stored on AWS, and AWS Glue discovers your data and stores the associated metadata (e.g. table definition and schema) in the AWS Glue Data Catalog
- Once cataloged, your data is immediately searchable, queryable, and available for ETL
- AWS Glue generates the code to execute your data transformations and data loading processes
- https://aws.amazon.com/glue/features/

MEDIA SERVICES

Amazon Elastic Transcoder

- Amazon Elastic Transcoder is media transcoding in the cloud
- It is designed to be a highly scalable, easy to use and a cost effective way for developers and businesses to convert (or "transcode") media files from their source format into versions that will playback on devices like smartphones, tablets and PCs
- https://aws.amazon.com/elastictranscoder/

SECURITY, IDENTITY AND COMPLIANCE

Amazon Cognito

- Amazon Cognito lets you add user sign-up, sign-in, and access control to your web and mobile apps quickly and easily
- Amazon Cognito scales to millions of users and supports sign-in with social identity providers, such as Facebook, Google, and Amazon, and enterprise identity providers via SAML 2.0
- https://aws.amazon.com/cognito/details/

AWS Certificate Manager

- AWS Certificate Manager is a service that lets you easily provision, manage, and deploy public and private Secure Sockets Layer/Transport Layer Security (SSL/TLS) certificates for use with AWS services and your internal connected resources

- SSL/TLS certificates are used to secure network communications and establish the identity of websites over the Internet as well as resources on private networks
- AWS Certificate Manager removes the time-consuming manual process of purchasing, uploading, and renewing SSL/TLS certificates
- https://aws.amazon.com/certificate-manager/features/

AWS CloudHSM

- AWS CloudHSM is a cloud-based hardware security module (HSM) that enables you to easily generate and use your own encryption keys on the AWS Cloud
- With CloudHSM, you can manage your own encryption keys using FIPS 140-2 Level 3 validated HSMs
- CloudHSM offers you the flexibility to integrate with your applications using industry-standard APIs, such as PKCS#11, Java Cryptography Extensions (JCE), and Microsoft CryptoNG (CNG) libraries
- https://aws.amazon.com/cloudhsm/features/

AWS Directory Service

- AWS Directory Service for Microsoft Active Directory, also known as AWS Managed Microsoft AD, enables your directory-aware workloads and AWS resources to use managed Active Directory in the AWS Cloud
- AWS Managed Microsoft AD is built on actual Microsoft Active Directory and does not require you to synchronize or replicate data from your existing Active Directory to the cloud
- You can use standard Active Directory administration tools and take advantage of built-in Active Directory features, such as Group Policy and single sign-on (SSO)
- With AWS Managed Microsoft AD, you can easily join Amazon EC2 and Amazon RDS for SQL Server instances to your domain, and use AWS Enterprise IT applications such as Amazon WorkSpaces with Active Directory users and groups
- https://aws.amazon.com/directoryservice/features/

AWS Artifact

- AWS Artifact is your go-to, central resource for compliance-related information that matters to you.
- It provides on-demand access to AWS' security and compliance reports and select online agreements
- Reports available in AWS Artifact include our Service Organization Control (SOC) reports, Payment Card Industry (PCI) reports, and certifications from accreditation bodies across geographies and compliance verticals that validate the implementation and operating effectiveness of AWS security controls
- Agreements available in AWS Artifact include the Business Associate Addendum (BAA) and the Nondisclosure Agreement (NDA)

- https://aws.amazon.com/artifact/

MACHINE LEARNING

Amazon Rekognition

- Amazon Rekognition makes it easy to add image and video analysis to your applications
- You just provide an image or video to the Rekognition API, and the service can identify the objects, people, text, scenes, and activities, as well as detect any inappropriate content
- Amazon Rekognition also provides highly accurate facial analysis and facial recognition on images and video that you provide
- You can detect, analyze, and compare faces for a wide variety of user verification, people counting, and public safety use cases
- https://aws.amazon.com/rekognition/

Amazon SageMaker

- Amazon SageMaker is a fully-managed platform that enables developers and data scientists to quickly and easily build, train, and deploy machine learning models at any scale
- Amazon SageMaker removes all the barriers that typically slow down developers who want to use machine learning
- https://aws.amazon.com/sagemaker/features/

Amazon Comprehend

- Amazon Comprehend is a natural language processing (NLP) service that uses machine learning to find insights and relationships in text
- The service identifies the language of the text; extracts key phrases, places, people, brands, or events; understands how positive or negative the text is; analyzes text using tokenization and parts of speech; and automatically organizes a collection of text files by topic
- Using these APIs, you can analyze text and apply the results in a wide range of applications including voice of customer analysis, intelligent document search, and content personalization for web applications
- https://aws.amazon.com/comprehend/features/

Amazon Transcribe

- Amazon Transcribe is an automatic speech recognition (ASR) service that makes it easy for developers to add speech-to-text capability to their applications

- Using the Amazon Transcribe API, you can analyze audio files stored in Amazon S3 and have the service return a text file of the transcribed speech
- Amazon Transcribe can be used for lots of common applications, including the transcription of customer service calls and generating subtitles on audio and video content
- The service can transcribe audio files stored in common formats, like WAV and MP3, with time stamps for every word so that you can easily locate the audio in the original source by searching for the text
- https://aws.amazon.com/transcribe/

MOBILE SERVICES

AWS AppSync

- AWS AppSync makes it easy to build data-driven mobile and browser-based apps that deliver responsive, collaborative experiences by keeping the data updated when devices are connected, enabling the app to use local data when offline, and synchronizing the data when the devices reconnect
- AWS AppSync uses the open standard GraphQL query language so you can request, change, and subscribe to the exact data you need with just a few lines of code
- https://aws.amazon.com/appsync/product-details/

AWS Device Farm

- AWS Device Farm is an app testing service that lets you test and interact with your Android, iOS, and web apps on many devices at once, or reproduce issues on a device in real time
- View video, screenshots, logs, and performance data to pinpoint and fix issues and increase quality before shipping your app
- https://aws.amazon.com/device-farm/

APPLICATION INTEGRATION

AWS Step Functions

- AWS Step Functions lets you coordinate multiple AWS services into serverless workflows so you can build and update apps quickly
- Using Step Functions, you can design and run workflows that stitch together services such as AWS Lambda and Amazon ECS into feature-rich applications

- Workflows are made up of a series of steps, with the output of one step acting as input into the next
- https://aws.amazon.com/step-functions/features/

Amazon MQ

- Amazon MQ is a managed message broker service for Apache ActiveMQ that makes it easy to set up and operate message brokers in the cloud
- Message brokers allow different software systems–often using different programming languages, and on different platforms–to communicate and exchange information
- Messaging is the communications backbone that connects and integrates the components of distributed applications, such as order processing, inventory management, and order fulfillment for e-commerce
- https://aws.amazon.com/amazon-mq/features/

Amazon SQS

- Amazon Simple Queue Service (SQS) is a fully managed message queuing service that enables you to decouple and scale microservices, distributed systems, and serverless applications
- SQS eliminates the complexity and overhead associated with managing and operating message oriented middleware, and empowers developers to focus on differentiating work
- Using SQS, you can send, store, and receive messages between software components at any volume, without losing messages or requiring other services to be available
- https://aws.amazon.com/sqs/features/

Amazon SWF

- Amazon SWF helps developers build, run, and scale background jobs that have parallel or sequential steps
- You can think of Amazon SWF as a fully-managed state tracker and task coordinator in the Cloud
- https://aws.amazon.com/swf/

INTERNET OF THINGS

AWS IoT Core

- AWS IoT Core is a managed cloud service that lets connected devices easily and securely interact with cloud applications and other devices. AWS IoT Core can support billions of

devices and trillions of messages and can process and route those messages to AWS endpoints and to other devices reliably and securely. With AWS IoT Core, your applications can keep track of and communicate with all your devices, all the time, even when they aren't connected

- https://aws.amazon.com/iot-core/features/

DESKTOP & APP STREAMING

Amazon Workspaces

- Amazon WorkSpaces is a managed, secure cloud desktop service. You can use Amazon WorkSpaces to provision either Windows or Linux desktops in just a few minutes and quickly scale to provide thousands of desktops to workers across the globe
- Amazon WorkSpaces offers you an easy way to provide a secure, managed, cloud-based virtual desktop experience to your end-users. Unlike traditional on-premises Virtual Desktop Infrastructure (VDI) solutions, you don't have to worry about procuring, deploying, and managing a complex environment – Amazon WorkSpaces takes care of the heavy lifting and provides a fully managed service
- https://aws.amazon.com/workspaces/features/

CONCLUSION

Hopefully these training notes have helped you to gain a complete understanding of the facts you need to know to pass the AWS Certified Cloud Practitioner exam.

The exam covers a broad set of technologies and it's vital to ensure you are armed with the knowledge to answer whatever questions come up in your certification exam, so we recommend reviewing these training notes until you're confident in all areas.

Don't forget to claim your bonus offer:
Gain access to a free 65-question practice exam on our interactive exam simulator. Simply send an email to info@digitalcloud.training with "CCP-BONUS" in the subject line and a copy of your purchase receipt attached. If you have already purchased our full set of practice questions, please note that these questions are already included.

Sign up today to get access to the full set of practice questions
The Digital Cloud Training practice questions are the closest to the actual exam question format you can find and the only exam-difficulty questions on the market. If you can pass our exams, you're well set to smash the real thing! Click the link below to get signed up:
https://digitalcloud.training/aws-certified-cloud-practitioner-practice-tests-2019/

Feel free to reach out with any questions you may have:

Join our private Facebook group to ask questions and share knowledge and exam tips with the community: https://www.facebook.com/groups/awscertificationqa

For technical support, reach out via email support@digitalcloud.training

Best wishes for your certification journey!

OTHER BOOKS BY THIS AUTHOR

AWS CERTIFIED SOLUTIONS ARCHITECT ASSOCIATE TRAINING NOTES

This book is based on the latest version of the Amazon Web Services (AWS) Certified Solutions Architect Associate (SAA-C01) exam blueprint that was released in February 2018.

The SAA-C01 exam covers a broad set of AWS services and the aim of this AWS Solutions Architect Associate study guide is to provide a detailed list of the facts you need to know before you sit the exam. This will shortcut your study time and maximize your chance of passing the exam first time.

The Solutions Architect – Associate certification is extremely valuable in the Cloud Computing industry today. Preparing to answer the associate level scenario-based questions requires a significant commitment in time and effort.

AWS Solutions Architect and successful IT instructor, Neal Davis, has consolidated the information you need to be successful. Master the details of the AWS Cloud so you can achieve exam success.

This book will help you prepare for your AWS Certified Solutions Architect – Associate exam in the following ways:

• Deep dive into the SAA-C01 exam objectives with over 240 pages of detailed facts, tables, and diagrams – everything you need to know!
• Familiarize yourself with the exam question format with the practice questions included in each section
• Use our online exam simulator to evaluate progress and ensure you're ready for the real AWS exam.

AWS CERTIFIED SOLUTIONS ARCHITECT ASSOCIATE PRACTICE QUESTIONS

AVAILABLE ON KINDLE ONLY

The **AWS Solutions Architect Associate certification** is extremely valuable in the Cloud Computing industry today and preparing to answer the difficult scenario-based questions requires a significant commitment in time and effort.

The **latest SAA-C01 exam** is composed entirely of scenario-based questions that test your knowledge and experience working with Amazon Web Services. Our practice tests are patterned to reflect the difficulty of the AWS exam and are the closest to the real AWS exam experience available anywhere.

There are **6 practice exams with 65 questions each** covering the five domains of the AWS exam blueprint. Each set of questions is repeated once without answers and explanations, and once with answers and explanations, so you get to choose from two methods of preparation:

- To simulate the exam experience and assess your exam readiness, use the "**PRACTICE QUESTIONS ONLY**" sets.
- To use the practice questions as a learning tool, use the "**PRACTICE QUESTIONS, ANSWERS & EXPLANATIONS**" sets to view the answers and read the in-depth explanations as you move through the questions.

With more than 20 years of experience in the IT industry, **Neal Davis** is a true expert in virtualization and cloud computing. Neal's practice tests have been used by over 20,000 students and are highly regarded for their quality and similarity to the real AWS exam.
These Practice Questions will prepare you for your AWS exam in the following ways:

- **Master the new exam pattern**: All 390 practice questions are based on the SAA-C01 exam blueprint and use the question format of the real AWS exam
- **6 sets of exam-difficulty practice questions**: Presented with and without answers so you can study or simulate an exam
- **Ideal exam prep tool that will shortcut your study time**: Assess your exam readiness to maximize your chance of passing the AWS exam first time.

The exam covers a broad set of technologies and it's vital to ensure you are armed with the knowledge to answer whatever questions come up in your certification exam, so we recommend reviewing these practice questions until you're confident in all areas and **ready to ace your AWS exam**.

ABOUT THE AUTHOR

Neal Davis is the founder of Digital Cloud Training, an AWS Cloud Solutions Architect and a successful IT instructor. With more than 20 years of experience in the tech industry, Neal is a true expert in virtualization and cloud computing. His passion is to help others achieve career success by offering in-depth AWS certification training resources.

Neal started DIGITAL CLOUD TRAINING to provide a variety of certification training resources for Amazon Web Services (AWS) certifications that represent a higher standard of quality than is otherwise available in the market. With over 15,000 students currently enrolled in digitalcloud.training courses, Neal's focus is on creating additional course content and growing his student base.

Save valuable time by getting straight to the facts you need to know to be successful and ensure you pass your AWS Certified Cloud Practitioner exam first time

---

This book is based on the CLF-C01 exam blueprint and provides a deep dive into the subject matter in a concise and easy-to-read format so you can fast-track your time to success.

The Cloud Practitioner certification is a great first step into the world of Cloud Computing and requires a foundational knowledge of the AWS Cloud, its architectural principles, value proposition, billing and pricing, key services and more.

AWS Solutions Architect and successful instructor, Neal Davis, has consolidated the information you need to be successful from numerous training sources and AWS FAQ pages to save you time.

In addition to the book, you are provided with access to a 65-question practice exam on an interactive exam simulator to evaluate your progress and ensure you're prepared for the style and difficulty of the real AWS exam.

This book can help you prepare for your AWS exam in the following ways:

- Deep dive into the CLF-C01 exam objectives with over 150 pages of detailed facts, tables, and diagrams – everything you need to know!
- Familiarize yourself with the exam question format with the practice questions included in each section
- Use our online exam simulator to evaluate progress and ensure you're ready for the real thing

DigitalCloud TRAINING