SECTION 1

Let's Get Started!

© Digital Cloud Training | https://digitalcloud.training

The SOA-C02 Exam

The SOA-C02 Exam

Level: Associate

Length: 130 minutes

Format: 65 questions

Exam Labs: Exam labs are paused from March 28th 2023

Cost: $150 USD

Delivery Method: Testing center or online

Scoring:

- Scaled score between 100 – 1000
- Minimum passing score of 720

© Digital Cloud Training | https://digitalcloud.training

The SOA-C02 Exam

Question format:

- Multiple-choice: Has one correct response and three incorrect responses
- Multiple-response: Has two or more correct responses out of five or more options

© Digital Cloud Training | https://digitalcloud.training

The SOA-C02 Exam

Domain 1: Monitoring, Logging, and Remediation

- Task Statement 1.1: Implement metrics, alarms, and filters by using AWS monitoring and logging services
- Task Statement 1.2: Remediate issues based on monitoring and availability metrics

Domain 2: Reliability and Business Continuity

- Task Statement 2.1: Implement scalability and elasticity
- Task Statement 2.2: Implement high availability and resilient environments
- Task Statement 2.3: Implement backup and restore strategies

© Digital Cloud Training | https://digitalcloud.training

The SOA-C02 Exam

Domain 3: Deployment, Provisioning, and Automation

- Task Statement 3.1: Provision and maintain cloud resources
- Task Statement 3.2: Automate manual or repeatable processes

Domain 4: Security and Compliance

- Task Statement 4.1: Implement and manage security and compliance policies
- Task Statement 4.2: Implement data and infrastructure protection strategies

© Digital Cloud Training | https://digitalcloud.training

The SOA-C02 Exam

Domain 5: Networking and Content Delivery

- Task Statement 5.1: Implement networking features and connectivity
- Task Statement 5.2: Configure domains, DNS services, and content delivery
- Task Statement 5.3: Troubleshoot network connectivity issues

Domain 6: Cost and Performance Optimization

- Task Statement 6.1: Implement cost optimization strategies
- Task Statement 6.2: Implement performance optimization strategies

© Digital Cloud Training | https://digitalcloud.training

SECTION 2

Create a Free Tier Account on AWS

© Digital Cloud Training | https://digitalcloud.training

AWS Account Overview

© Digital Cloud Training | https://digitalcloud.training

AWS Account Overview

It's an IAM best practice to create **individual users** and to avoid using the **Root** account

Unique **email address** required

The Root user has **full control** over the account

© Digital Cloud Training | https://digitalcloud.training

AWS Account Overview

Authentication: IAM principals authenticate to IAM using the console, API, or CLI

Authorization: IAM principals can then create resources across AWS Regions

All AWS **identities** and **resources** are created within the AWS account

© Digital Cloud Training | https://digitalcloud.training

Create your AWS Free Tier Account

© Digital Cloud Training | https://digitalcloud.training

What you need...

Credit card for setting up the account and paying any bills

Unique email address for this account

Check if you can use a **dynamic alias** with an existing email address

john@gmail.com

john+ACCOUNT-ALIAS-1@gmail.com

john+ACCOUNT-ALIAS-2@gmail.com

AWS account name / alias

Phone to receive an **SMS** verification code

© Digital Cloud Training | https://digitalcloud.training

Configure Account and Create a Budget and Alarm

© Digital Cloud Training | https://digitalcloud.training

Account Configuration

- Configure **Account Alias**
- Enable access to billing for **IAM users**
- Update **billing preferences**
- Create a **budget and alarm**

© Digital Cloud Training | https://digitalcloud.training

AWS Identity and Access Management (AWS IAM)

© Digital Cloud Training | https://digitalcloud.training

AWS IAM

IAM Principals must be **authenticated** to send requests (with a few exceptions)

© Digital Cloud Training | https://digitalcloud.training

AWS IAM Principals

A **principal** is a person or application that can make a request for an **action** or **operation** on an AWS **resource**

© Digital Cloud Training | https://digitalcloud.training

Principals and Authorizing Requests

Request to perform operation on AWS resource

IAM User

arn:aws:iam::111122223333:user/Paul

Amazon Resource Name (ARN)

IAM Role

arn:aws:iam::111122223333:role/my-app-role

Amazon Resource Name (ARN)

Request context – information about the request (actions, operations, principal, resource)

↓

Authentication – user name/password, access keys etc.

↓

Authorization – values from request context used to determine if request allowed or denied

↓

Approval – actions or operations approved on resources

© Digital Cloud Training | https://digitalcloud.training

IAM Authentication Methods

© Digital Cloud Training | https://digitalcloud.training

Authentication for Applications

© Digital Cloud Training | https://digitalcloud.training

IAM User Groups

IAM Users and Roles are **IAM Identities**

IAM Group – "Operations"

User groups are used for **collecting** users with common needs and then applying **IAM permissions policies** to them

IAM Group – "Developers"

User groups are **NOT** IAM principals, and **cannot** be **authenticated** or **authorized**; they are used purely for applying permissions

© Digital Cloud Training | https://digitalcloud.training

Create an IAM User and Group

© Digital Cloud Training | https://digitalcloud.training

Install Tools and Configure AWS CLI

© Digital Cloud Training | https://digitalcloud.training

Install Tools and Configure AWS CLI

- ✓ Download the code (from the *next* lesson)
- ✓ Install Visual Studio Code
- ✓ Install and Configure the AWS CLI
- ✓ Access AWS CloudShell

© Digital Cloud Training | https://digitalcloud.training

SECTION 3

AWS Fundamentals

© Digital Cloud Training | https://digitalcloud.training

Amazon Virtual Private Cloud (VPC)

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

A **VPC** is a logically isolated portion of the AWS cloud within a region

Subnets are created within **AZs**

Main Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	igw-id

The **route table** is used to configure the VPC router

An **Internet Gateway** is used to connect to the Internet

You can launch **EC2 instances** into your VPC subnets

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

Each **VPC** has a different block of IP addresses

CIDR stands for Classless Interdomain Routing

Each subnet has a block of **IP addresses** from the CIDR block

You can create **multiple VPCs** within each region

© Digital Cloud Training | https://digitalcloud.training

AWS Public/Private Services and Endpoints

© Digital Cloud Training | https://digitalcloud.training

Create a Custom VPC

© Digital Cloud Training | https://digitalcloud.training

Custom VPC

Main Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	igw-id

Private-RT Route Table

Destination	Target
10.0.0.0/16	Local

© Digital Cloud Training | https://digitalcloud.training

Security Groups and Network ACLs

© Digital Cloud Training | https://digitalcloud.training

Stateful vs Stateless Firewalls

PROTOCOL	SOURCE IP	DESTINATION IP	SOURCE PORT	DESTINATION PORT
HTTP	10.1.1.1	10.2.1.10	65188	80
HTTP	10.2.1.10	10.1.1.1	80	65188

A stateful firewall allows the return traffic automatically

A stateless firewall checks for an allow rule for both connections

© Digital Cloud Training | https://digitalcloud.training

Security Groups and Network ACLs

NACLs apply at the **subnet** level

NACLs apply only to traffic **entering** / **exiting** the subnet

© Digital Cloud Training | https://digitalcloud.training

Security Groups & Network Access Control Lists (NACLs)

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

© Digital Cloud Training | https://digitalcloud.training

Create a Security Group

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 and EBS Overview

© Digital Cloud Training | https://digitalcloud.training

Amazon Elastic Compute Cloud (EC2)

EC2 instances run Windows, Linux, or MacOS

An **EC2 instance** is a virtual server

EC2 hosts are **managed by AWS**

EC2 Host Server

A selection of **instance types** come with varying combinations of CPU, memory, storage and networking

© Digital Cloud Training | https://digitalcloud.training

Public, Private, and Elastic IP addresses

Type	Description
Public IP address	Lost when the instance is stopped Used in Public Subnets No charge Associated with a private IP address on the instance Cannot be moved between instances
Private IP address	Retained when the instance is stopped Used in Public and Private Subnets
Elastic IP address	Static Public IP address You are charged if not used Associated with a private IP address on the instance Can be moved between instances and Elastic Network Adapters

© Digital Cloud Training | https://digitalcloud.training

Network Interfaces (ENI, ENA, EFA)

© Digital Cloud Training | https://digitalcloud.training

Network Interfaces (ENI, ENA, EFA)

Elastic network interface

- Basic adapter type for when you don't have any high-performance requirements
- Can use with all instance types

Elastic network adapter

- Enhanced networking performance
- Higher bandwidth and lower inter-instance latency
- Must choose supported instance type

Elastic Fabric Adapter

- Use with High Performance Computing and MPI and ML use cases
- Tightly coupled applications
- Can use with all instance types

© Digital Cloud Training | https://digitalcloud.training

Amazon Elastic Block Store (EBS)

The volume is automatically replicated **within the AZ**

© Digital Cloud Training | https://digitalcloud.training

Amazon EBS Volume Types

	General Purpose SSD		Provisioned IOPS SSD		
Volume type	gp3	gp2	io2 Block Express ‡	io2	io1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.999% durability (0.001% annual failure rate)		99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	• Low-latency interactive apps • Development and test environments		Workloads that require sub-millisecond latency, and sustained IOPS performance or more than 64,000 IOPS or 1,000 MiB/s of throughput	• Workloads that require sustained IOPS performance or more than 16,000 IOPS • I/O-intensive database workloads	
Volume size	1 GiB - 16 TiB		4 GiB - 64 TiB	4 GiB - 16 TiB	
Max IOPS per volume (16 KiB I/O)	16,000		256,000	64,000 †	
Max throughput per volume	1,000 MiB/s	250 MiB/s *	4,000 MiB/s	1,000 MiB/s †	
Amazon EBS Multi-attach	Not supported		Not supported	Supported	
Boot volume	Supported				

© Digital Cloud Training | https://digitalcloud.training

Amazon EBS vs Instance Store

EBS volumes are **attached** over the **network**

EBS volumes offer persistent storage

Instance Store volumes are physically attached to the host and offer high - performance

Instance Store volumes are **ephemeral** (non-persistent)

© Digital Cloud Training | https://digitalcloud.training

Launching and Managing EC2 and EBS

© Digital Cloud Training | https://digitalcloud.training

Launching an EC2 Instance

EC2 Instance

Select an **instance type**

An **AMI** defines the configuration of the instance

A **snapshot** is a point-in-time backup of an instance

Amazon Machine Image (AMI)

EBS Snapshot

Family	Type	vCPUs	Memory (GiB)
General purpose	t2.micro	1	1
Compute optimized	c5n.large	2	5.25
Memory optimized	r5ad.large	2	16
Storage optimized	d2.xlarge	4	30.5
GPU instances	g2.2xlarge	8	15

Linux

Microsoft Windows

You can customize your instance and create a **custom AMI**

The **instance type** defines the hardware profile (and cost)

Customized AMI

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

Changing the Instance Type

1. **Stop** the Instance

2. Select **Change Instance Type**

3. Choose the new instance type

Note: Stopping and starting the instance will also move it to different underlying hardware

© Digital Cloud Training | https://digitalcloud.training

EC2 Instance Lifecycle

© Digital Cloud Training | https://digitalcloud.training

EC2 Instance Lifecycle

© Digital Cloud Training | https://digitalcloud.training

EC2 Instance Lifecycle

Stopping EC2 instances

- EBS backed instances only
- No charge for stopped instances
- EBS volumes remain attached (chargeable)
- Data in RAM is lost
- Instance is migrated to a different host
- Private IPv4 addresses and IPv6 addresses retained; public IPv4 addresses released
- Associated Elastic IPs retained

© Digital Cloud Training | https://digitalcloud.training

EC2 Instance Lifecycle

Hibernating EC2 instances

- Applies to on-demand or reserved Linux instances
- Contents of RAM saved to EBS volume
- Must be enabled for hibernation when launched
- Specific prerequisites apply
- When started (after hibernation):
  - The EBS root volume is restored to its previous state
  - The RAM contents are reloaded
  - The processes that were previously running on the instance are resumed
  - Previously attached data volumes are reattached and the instance retains its instance ID

© Digital Cloud Training | https://digitalcloud.training

EC2 Instance Lifecycle

Rebooting EC2 instances

- Equivalent to an OS reboot
- DNS name and all IPv4 and IPv6 addresses retained
- Does not affect billing

Retiring EC2 instances

- Instances may be retired if AWS detects irreparable failure of the underlying hardware that hosts the instance
- When an instance reaches its scheduled retirement date, it is stopped or terminated by AWS

© Digital Cloud Training | https://digitalcloud.training

EC2 Instance Lifecycle

Terminating EC2 instances

- Means deleting the EC2 instance
- Cannot recover a terminated instance
- By default root EBS volumes are deleted

Recovering EC2 instances

- CloudWatch can be used to monitor system status checks and recover instance if needed
- Applies if the instance becomes impaired due to underlying hardware / platform issues
- Recovered instance is identical to original instance

© Digital Cloud Training | https://digitalcloud.training

Nitro Instances and Nitro Enclaves

© Digital Cloud Training | https://digitalcloud.training

AWS Nitro System

- Nitro is the underlying platform for the next generation of EC2 instances
- Support for many virtualized and bare metal instance types
- Breaks functions into specialized hardware with a Nitro Hypervisor
- Specialized hardware includes:
  - Nitro cards for VPC
  - Nitro cards for EBS
  - Nitro for Instance Storage
  - Nitro card controller
  - Nitro security chip
  - Nitro hypervisor
  - Nitro Enclaves

© Digital Cloud Training | https://digitalcloud.training

AWS Nitro System

- Improves performance, security and innovation:
  - Performance close to bare metal for virtualized instances
  - Elastic Network Adapter and Elastic Fabric Adapter
  - More bare metal instance types
  - Higher network performance (e.g. 100 Gbps)
  - High Performance Computing (HPC) optimizations
  - Dense storage instances (e.g. 60 TB)

© Digital Cloud Training | https://digitalcloud.training

AWS Nitro Enclaves

- Isolated compute environments
- Runs on isolated and hardened virtual machines
- No persistent storage, interactive access, or external networking
- Uses cryptographic attestation to ensure only authorized code is running
- Integrates with AWS Key Management Service (KMS)
- Protect and securely process highly sensitive data:
  - Personally identifiable information (PII)
  - Healthcare data
  - Financial data
  - Intellectual Property data

© Digital Cloud Training | https://digitalcloud.training

Serverless Services

© Digital Cloud Training | https://digitalcloud.training

Serverless Services and Event-Driven Architecture

© Digital Cloud Training | https://digitalcloud.training

Serverless Services

- With serverless there are **no instances** to manage
- You don’t need to provision hardware
- There is no management of operating systems or software
- Capacity provisioning and patching is handled automatically
- Provides automatic scaling and high availability
- Can be very cheap!

© Digital Cloud Training | https://digitalcloud.training

AWS Lambda

© Digital Cloud Training | https://digitalcloud.training

AWS Lambda

- AWS 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 AWS Lambda:
  - No servers to manage
  - Continuous scaling
  - Millisecond billing
  - Integrates with almost all other AWS services

© Digital Cloud Training | https://digitalcloud.training

AWS Lambda

- Primary use cases for AWS Lambda:
  - Data processing
  - Real-time file processing
  - Real-time stream processing
  - Build serverless backends for web, mobile, IOT, and 3rd party API requests

© Digital Cloud Training | https://digitalcloud.training

Decoupling with SQS Queues

App tier must keep up with workload or failure will occur

© Digital Cloud Training | https://digitalcloud.training

Amazon SNS Notifications

Event **producer** sends one message to one SNS **topic**

Each **subscriber** will get all the messages

© Digital Cloud Training | https://digitalcloud.training

AWS Storage: S3, EFS, FSx

© Digital Cloud Training | https://digitalcloud.training

File Storage vs Object Storage

File Share

- Data stored in directories
- A hierarchy of directories can be formed
- File systems are mounted to an operating system
- Function like local storage

Object Store

`http://bucket.s3.aws-region.amazonaws.com`

- Data stored in buckets
- Flat namespace (no hierarchy)
- Hierarchy can be mimicked with prefixes
- Accessed by **REST API** and cannot be mounted

© Digital Cloud Training | https://digitalcloud.training

Amazon Simple Storage Service (S3)

A **bucket** is a container for objects

S3 Bucket

An **object** is a file you upload

You can store millions of **objects** in a **bucket**

Accessing objects in a bucket:

`https://bucket.s3.aws-region.amazonaws.com/key`

`https://s3.aws-region.amazonaws.com/bucket/key`

The **HTTP protocol** is used with a **REST API** (e.g. GET, PUT, POST, SELECT, DELETE)

© Digital Cloud Training | https://digitalcloud.training

Amazon Simple Storage Service (S3)

- You can store any type of file in S3
- Files can be anywhere from 0 bytes to 5 TB
- There is unlimited storage available
- S3 is a universal namespace so **bucket names** must be **unique globally**
- However, you create your buckets within a **REGION**
- It is a best practice to create buckets in regions that are physically closest to your users to reduce latency
- There is no hierarchy for objects within a bucket
- Delivers strong read-after-write consistency

© Digital Cloud Training | https://digitalcloud.training

Buckets, Folders, and Objects

The object name is the **Key**
the data is the **Value**

A **Folder** is a shared **prefix**
for grouping objects

© Digital Cloud Training | https://digitalcloud.training

Buckets, Folders, and Objects

- Folders **can** be created within folders
- Buckets **cannot** be created within other buckets
- An objects consists of:
  - Key (the name of the object)
  - Version ID
  - Value (actual data)
  - Metadata
  - Subresources
  - Access control information

© Digital Cloud Training | https://digitalcloud.training

What is File Storage?

Mount command:
mount –t nfs servername:folderpath /mountpoint

Appliances are known as **Network Attached Storage (NAS)** devices for both SMB / CIFS and NFS

© Digital Cloud Training | https://digitalcloud.training

Amazon Elastic File System (EFS)

© Digital Cloud Training | https://digitalcloud.training

FSx

Amazon FSx

- Amazon FSx provides fully managed third-party file systems
- Amazon FSx provides you with two file systems to choose from:
  - **Amazon FSx for Windows File Server** for Windows-based applications
  - **Amazon FSx for Lustre** for compute-intensive workloads

© Digital Cloud Training | https://digitalcloud.training

Amazon FSx for Windows File Server

With **multi-AZ** a standby can be created in another AZ

© Digital Cloud Training | https://digitalcloud.training

FSx

Amazon FSx for Lustre

- High-performance file system optimized for fast processing of workloads such as:
  - Machine learning
  - High performance computing (HPC)
  - Video processing
  - Financial modeling
  - Electronic design automation (EDA)
- Works natively with S3, letting you transparently access your S3 objects as files
- Your S3 objects are presented as files in your file system, and you can write your results back to S3
- Provides a POSIX-compliant file system interface

DigitalCloud TRAINING

Amazon FSx for Lustre

Use for **cloud bursting** and data **migration**

S3 objects can be viewed as files in FSx; result data can be written back to S3

© Digital Cloud Training | https://digitalcloud.training

Working with Files on EFS

© Digital Cloud Training | https://digitalcloud.training

Amazon EFS

© Digital Cloud Training | https://digitalcloud.training

Working with Objects on S3

© Digital Cloud Training | https://digitalcloud.training

Using the REST API with Amazon S3

Example request to **download** an image from Amazon S3

```
GET /sunrise.jpg HTTP/1.1
Host: myphotos.s3.us-east-1.amazonaws.com
Date: Mon, 6 Jun 2022 22:32:00 GMT
```

Example request to **upload** an image to Amazon S3

```
PUT /sunset.jpg HTTP/1.1
Host: myphotos.s3.us-east-1.amazonaws.com
Date: Tue, 7 Jun 2022 17:50:00 GMT
```

© Digital Cloud Training | https://digitalcloud.training

AWS Databases: RDS, Aurora, and ElastiCache

© Digital Cloud Training | https://digitalcloud.training

Amazon Relational Database Service (RDS)

RDS runs on EC2 instances, so you choose an **instance type**

RDS is a managed, **relational database**

RDS supports the following database engines:

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

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS

- Relational databases are known as Structured Query Language (SQL) databases
- RDS used for Online Transaction Processing (OLTP) use cases
- Easy to setup, highly available, fault tolerant, and scalable
- Common use cases include online stores and banking systems
- You can encrypt your Amazon RDS instances and snapshots at rest by enabling the encryption option for your Amazon RDS DB instance (during creation)
- Encryption uses AWS Key Management Service (KMS)

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS – Scaling up (vertically)

© Digital Cloud Training | https://digitalcloud.training

Amazon Aurora

- Amazon Aurora is an AWS database offering in the RDS family
- Amazon Aurora is a MySQL and PostgreSQL-compatible relational database built for the cloud
- Amazon Aurora is up to five times faster than standard MySQL databases and three times faster than standard PostgreSQL databases
- Amazon Aurora features a distributed, fault-tolerant, self-healing storage system that auto-scales up to 128TB per database instance

© Digital Cloud Training | https://digitalcloud.training

Amazon Aurora Key Features

Aurora Feature	Benefit
High performance and scalability	Offers high performance, self-healing storage that scales up to 128TB, point-in-time recovery and continuous backup to S3
DB compatibility	Compatible with existing MySQL and PostgreSQL open source databases
Aurora Replicas	In-region read scaling and failover target – up to 15 (can use Auto Scaling)
MySQL Read Replicas	Cross-region cluster with read scaling and failover target – up to 5 (each can have up to 15 Aurora Replicas)
Global Database	Cross-region cluster with read scaling (fast replication / low latency reads). Can remove secondary and promote
Multi-Master	Scales out writes within a region
Serverless	On-demand, autoscaling configuration for Amazon Aurora - does not support read replicas or public IPs (can only access through VPC or Direct Connect - not VPN)

© Digital Cloud Training | https://digitalcloud.training

Amazon ElastiCache

- Fully managed implementations **Redis** and **Memcached**
- ElastiCache is a **key/value** store
- In-memory database offering high performance and low latency
- Can be put in front of databases such as RDS and DynamoDB
- ElastiCache nodes run on Amazon EC2 instances, so you must choose an instance family/type

© Digital Cloud Training | https://digitalcloud.training

Amazon ElastiCache

Feature	Memcached	Redis (cluster mode disabled)	Redis (cluster mode enabled)
Data persistence	No	Yes	Yes
Data types	Simple	Complex	Complex
Data partitioning	Yes	No	Yes
Encryption	No	Yes	Yes
High availability (replication)	No	Yes	Yes
Multi-AZ	Yes, place nodes in multiple AZs. No failover or replication	Yes, with auto-failover. Uses read replicas (0-5 per shard)	Yes, with auto-failover. Uses read replicas (0-5 per shard)
Scaling	Up (node type); out (add nodes)	Up (node type); out (add replica)	Up (node type); out (add shards)
Multithreaded	Yes	No	No
Backup and restore	No (and no snapshots)	Yes, automatic and manual snapshots	Yes, automatic and manual snapshots

© Digital Cloud Training | https://digitalcloud.training

Amazon ElastiCache Use Cases

- Data that is relatively **static** and **frequently accessed**
- Applications that are tolerant of stale data
- Data is slow and expensive to get compared to cache retrieval
- Require push-button scalability for memory, writes and reads
- Often used for storing session state

© Digital Cloud Training | https://digitalcloud.training

Amazon ElastiCache Examples

Use Case	Benefit
Web session store	In cases with load-balanced web servers, store web session information in Redis so if a server is lost, the session info is not lost, and another web server can pick it up
Database caching	Use Memcached in front of AWS RDS to cache popular queries to offload work from RDS and return results faster to users
Leaderboards	Use Redis to provide a live leaderboard for millions of users of your mobile app
Streaming data dashboards	Provide a landing spot for streaming sensor data on the factory floor, providing live real-time dashboard displays

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Security

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Security

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

Amazon RDS Security

- Encryption **at rest** can be enabled – includes DB storage, backups, read replicas and snapshots
- You can only enable encryption for an Amazon RDS DB instance when you create it, not after the DB instance is created
- DB instances that are encrypted can't be modified to disable encryption
- Uses AES 256 encryption and encryption is transparent with minimal performance impact
- RDS for Oracle and SQL Server is also supported using Transparent Data Encryption (TDE) (may have performance impact)
- AWS KMS is used for managing encryption keys

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Security

- You can't have:
  - An **encrypted** read replica of an **unencrypted** DB instance
  - An **unencrypted** read replica of an **encrypted** DB instance
- Read replicas of encrypted primary instances are encrypted
- The same KMS key is used if in the same Region as the primary
- If the read replica is in a different Region, a different KMS key is used
- You can't restore an unencrypted backup or snapshot to an encrypted DB instance

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Security

© Digital Cloud Training | https://digitalcloud.training

Create and Manage RDS Database

© Digital Cloud Training | https://digitalcloud.training

SECTION 4

Monitoring, Logging, and Remediation

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Overview

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch

CloudWatch is used for performance monitoring, alarms, log collection and automated actions

Use cases / benefits include:

- Collect performance metrics from AWS and on-premises systems
- Automate responses to operational changes
- Improve operational performance and resource optimization
- Derive actionable insights from logs
- Get operational visibility and insight

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch

CloudWatch Core Features:

- **CloudWatch Metrics** – services send time-ordered data points to CloudWatch
- **CloudWatch Alarms** – monitor metrics and initiate actions
- **CloudWatch Logs** – centralized collection of system and application logs
- **CloudWatch Events** – stream of system events describing changes to AWS resources and can trigger actions

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Metrics

- Metrics are sent to CloudWatch for many AWS services
- EC2 metrics are sent every **5 minutes** by default (free)
- Detailed EC2 monitoring sends every **1 minute** (chargeable)
- Unified CloudWatch Agent sends system-level metrics for EC2 and on-premises servers
- System-level metrics include memory and disk usage

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Metrics

- Can publish custom metrics using CLI or API
- Custom metrics are one of the following resolutions:
  - **Standard resolution** – data having a one-minute granularity
  - **High resolution** – data at a granularity of one second
- AWS metrics are standard resolution by default

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Alarms

Two types of alarms

- **Metric alarm** – performs one or more actions based on a single metric
- **Composite alarm** – uses a rule expression and takes into account multiple alarms
- You can create a custom metric using the API or CLI command set-alarm-state
- Metric alarm states:
  - OK – Metric is within a threshold
  - ALARM – Metric is outside a threshold
  - INSUFFICIENT_DATA – not enough data

© Digital Cloud Training | https://digitalcloud.training

Metrics, Namespaces, and Dimensions

© Digital Cloud Training | https://digitalcloud.training

CloudWatch Metrics

- Metrics are the fundamental concept in CloudWatch
- A metric represents a time-ordered set of data points that are published to CloudWatch
- AWS services send metrics to CloudWatch
- You can also send your own custom metrics to CloudWatch
- Metrics exist within a region
- Metrics cannot be deleted but automatically expire after 15 months
- Metrics are uniquely defined by a name, a namespace, and zero or more dimensions
- Time stamps can be up to two weeks in the past and up to two hours into the future

© Digital Cloud Training | https://digitalcloud.training

CloudWatch Metrics

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)

© Digital Cloud Training | https://digitalcloud.training

CloudWatch Namespaces

- A namespace is a container for CloudWatch metrics
- Metrics in different namespaces are isolated from each other
- You must specify a namespace for each data point you publish to CloudWatch
- You can specify a namespace name when you create a metric

Service	Namespace
Amazon API Gateway	AWS/ApiGateway
Amazon CloudFront	AWS/CloudFront
AWS CloudHSM	AWS/CloudHSM
Amazon CloudWatch Logs	AWS/Logs
AWS CodeBuild	AWS/CodeBuild
Amazon Cognito	AWS/Cognito
Amazon DynamoDB	AWS/DynamoDB
Amazon EC2	AWS/EC2
AWS Elastic Beanstalk	AWS/ElasticBeanstalk

© Digital Cloud Training | https://digitalcloud.training

CloudWatch Dimensions

- A dimension is a name/value pair that is part of the identity of a metric
- You can assign up to 10 dimensions to a metric
- Dimensions are categories for the characteristics of each metric

CloudWatch Statistics

- Statistics are metric data aggregations over specified periods of time
- CloudWatch provides statistics based on the metric data points provided by your custom data or provided by other AWS services to CloudWatch

Statistic	Description
`Minimum`	The lowest value observed during the specified period. You can use this value to determine low volumes of activity for your application.
`Maximum`	The highest value observed during the specified period. You can use this value to determine high volumes of activity for your application.
`Sum`	All values submitted for the matching metric added together. This statistic can be useful for determining the total volume of a metric.
`Average`	The value of `Sum` / `SampleCount` during the specified period. By comparing this statistic with the `Minimum` and `Maximum`, you can determine the full scope of a metric and how close the average use is to the `Minimum` and `Maximum`. This comparison helps you to know when to increase or decrease your resources as needed.
`SampleCount`	The count (number) of data points used for the statistical calculation.
`pNN.NN`	The value of the specified percentile. You can specify any percentile, using up to two decimal places (for example, p95.45). Percentile statistics are not available for metrics that include any negative values. For more information, see Percentiles.

© Digital Cloud Training | https://digitalcloud.training

CloudWatch API Actions (Metrics)

GetMetricData

- Retrieve as many as 500 different metrics in a single request

PutMetricData

- Publishes metric data points to Amazon CloudWatch
- CloudWatch associates the data points with the specified metric
- If the specified metric does not exist, CloudWatch creates the metric

GetMetricStatistics

- Gets statistics for the specified metric
- CloudWatch aggregates data points based on the length of the period that you specify
- Maximum number of data points returned from a single call is 1,440

© Digital Cloud Training | https://digitalcloud.training

Create a Custom Metric and Alarm

© Digital Cloud Training | https://digitalcloud.training

The Unified CloudWatch Agent

© Digital Cloud Training | https://digitalcloud.training

The Unified CloudWatch Agent

The unified CloudWatch agent enables you to do the following:

- Collect internal system-level metrics from Amazon **EC2 instances** across operating systems
- Collect system-level metrics from **on-premises servers**
- Retrieve custom metrics from your applications or services using the StatsD and collectd protocols
- Collect logs from Amazon EC2 instances and on-premises servers (Windows / Linux)

© Digital Cloud Training | https://digitalcloud.training

The Unified CloudWatch Agent

- Agent must be installed on the server
- Can be installed on:
  - Amazon EC2 instances
  - On-premises servers
  - Linux, Windows Server, or macOS
- You can use the following installation methods:
  - Command line
  - AWS Systems Manager
  - AWS CloudFormation
- List of metrics collected:

https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/metrics-collected-by-CloudWatch-agent.html

© Digital Cloud Training | https://digitalcloud.training

The Unified CloudWatch Agent

Installation Process:

1. Create IAM roles to enable metric collection
2. Download the agent package
3. Update CloudWatch agent configuration file
4. Start the agent

- When starting the agent, you must attach the IAM role (EC2), or specify a named profile (on-premises)
- You can optionally integrate with AWS Systems Manager

© Digital Cloud Training | https://digitalcloud.training

Deploy CloudWatch Agent

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Logs

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Logs

- Gather and centralize application and system logs in CloudWatch
- Define expiration policies and KMS encryption
- Send to:
  - Amazon S3 (export)
  - Kinesis Data Streams
  - Kinesis Data Firehose

Unified CloudWatch Agent installed on EC2 and on-premises servers

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Logs Features

- CloudWatch Logs Insights can be used to interactively search and analyze log data
- Monitor applications and systems using log data
- Create alarms based on API activity captured by CloudTrail
- By default, logs are kept indefinitely and never expire
- You can adjust the retention policy for each log group, keeping the indefinite retention, or choosing a retention period between 10 years and one day
- CloudWatch Logs can be used to log information about the DNS queries that Route 53 receives

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudWatch Logs Concepts

- **Log events** – a record of activity recorded by a monitored resource
- **Log streams** – a sequence of log events that share the same source
- **Log groups** – groups of log streams that share retention, monitoring, and access control settings
- **Metric filters** – used to extract metric observations from ingested events and transform them to data points in a CloudWatch metric
- **Retention settings** – specify how long log events are kept in CloudWatch Logs

© Digital Cloud Training | https://digitalcloud.training

CloudWatch Logs and Filters

© Digital Cloud Training | https://digitalcloud.training

CloudWatch Dashboards

© Digital Cloud Training | https://digitalcloud.training

Amazon SNS

© Digital Cloud Training | https://digitalcloud.training

Amazon SNS Notifications

Event **producer** sends one message to one SNS **topic**

Each **subscriber** will get all the messages

© Digital Cloud Training | https://digitalcloud.training

Amazon SNS

- Amazon SNS is a highly available, durable, secure, fully managed pub/sub messaging service
- Amazon SNS provides topics for high-throughput, push-based, many-to-many messaging
- Publisher systems can fan out messages to many subscriber endpoints:
- Endpoints include:
  - Amazon SQS queues
  - AWS Lambda functions
  - HTTP/S webhooks
  - Mobile push
  - SMS
  - Email

© Digital Cloud Training | https://digitalcloud.training

Amazon SNS

- Multiple recipients can be grouped using Topics
- A topic is an “access point” for allowing recipients to dynamically subscribe for identical copies of the same notification
- One topic can support deliveries to multiple endpoint types
- Simple APIs and easy integration with applications
- Flexible message delivery over multiple transport protocols

© Digital Cloud Training | https://digitalcloud.training

Amazon SNS + Amazon SQS Fan-Out

- You can subscribe one or more Amazon SQS queues to an Amazon SNS topic
- Amazon SQS manages the subscription and any necessary permissions
- When you publish a message to a topic, Amazon SNS sends the message to every subscribed queue

© Digital Cloud Training | https://digitalcloud.training

Configure Notifications

© Digital Cloud Training | https://digitalcloud.training

AWS CloudTrail

© Digital Cloud Training | https://digitalcloud.training

AWS CloudTrail

- CloudTrail logs **API activity** for auditing
- By default, management events are logged and retained for 90 days
- A **CloudTrail Trail** logs any events to S3 for indefinite retention
- Trail can be within Region or all Regions
- CloudWatch Events can triggered based on API calls in CloudTrail
- Events can be streamed to CloudWatch Logs

© Digital Cloud Training | https://digitalcloud.training

AWS CloudTrail

© Digital Cloud Training | https://digitalcloud.training

CloudTrail – Types of Events

- **Management events** provide information about management operations that are performed on resources in your AWS account
- **Data events** provide information about the resource operations performed on or in a resource
- **Insights events** identify and respond to unusual activity associated with write API calls by continuously analyzing CloudTrail management events

© Digital Cloud Training | https://digitalcloud.training

CloudTrail – Types of Events

Example data events:

- Amazon S3 object-level API activity (for example, GetObject, DeleteObject, and PutObject API operations)
- AWS Lambda function invocation activity (for example, InvokeFunction API operations)
- Amazon DynamoDB object-level API activity on tables (for example, PutItem, DeleteItem, and UpdateItem)

Example management events:

- Creating an Amazon S3 bucket
- Creating and managing AWS IAM resources
- Registering devices
- Configuring routing table rules

© Digital Cloud Training | https://digitalcloud.training

Create a CloudTrail Trail

© Digital Cloud Training | https://digitalcloud.training

Amazon EventBridge

© Digital Cloud Training | https://digitalcloud.training

Amazon EventBridge

EventBridge used to be known as **CloudWatch Events**

© Digital Cloud Training | https://digitalcloud.training

Amazon EventBridge Example 1

© Digital Cloud Training | https://digitalcloud.training

Amazon EventBridge Example 1

Event pattern Info

Event source
AWS service or EventBridge partner as source

AWS services

AWS service
The name of the AWS service as the event source

EC2

Event type
The type of events as the source of the matching pattern

EC2 Instance State-change Notification

- ○ Any state
- ◉ Specific state(s)

terminated ×

- ◉ Any instance
- ○ Specific instance Id(s)

Event pattern
Event pattern, or filter to match the events

```
1 {
2   "source": ["aws.ec2"],
3   "detail-type": ["EC2 Instance State-change Notificatio
4   "detail": {
5     "state": ["terminated"]
6   }
7 }
```

Copy | Test pattern | Edit pattern


```
{
   "version": "0",
   "id": "6a7e8feb-b491-4cf7-a9f1-bf3703467718",
   "detail-type": "EC2 Instance State-change Notification",
   "source": "aws.ec2",
   "account": "111122223333",
   "time": "2017-12-22T18:43:48Z",
   "region": "us-west-1",
   "resources": [
      "arn:aws:ec2:us-west-1:123456789012:instance/i-1234567890abcdef0"
  ],
  "detail": {
    "instance-id": "i-1234567890abcdef0",
    "state": "terminated"
    }
  }
```


© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

Amazon EventBridge Example 2

```
{
   "source": ["aws.cloudtrail"],
   "detail-type": ["AWS API Call via CloudTrail"],
   "detail": {
     "eventSource": ["cloudtrail.amazonaws.com"],
     "eventName": ["s3:PutBucketPolicy"]
   }
 }
```

© Digital Cloud Training | https://digitalcloud.training

Create Event Bus and Rule

© Digital Cloud Training | https://digitalcloud.training

AWS Config

© Digital Cloud Training | https://digitalcloud.training

AWS Config

- Evaluate your AWS resource configurations for desired settings
- Get a snapshot of the current configurations of resources that are associated with your AWS account
- Retrieve configurations of resources that exist in your account
- Retrieve historical configurations of one or more resources
- Receive a notification whenever a resource is created, modified, or deleted
- View relationships between resources

© Digital Cloud Training | https://digitalcloud.training

AWS Config

AWS Config evaluates the **configuration** against desired configurations

Example Services:

Amazon EC2

Elastic Load Balancing

Amazon RDS

AWS Config

Amazon S3

Amazon Simple Notification Service

Send notifications with **SNS**

CloudWatch Events

Alert via **CW Events** when changes occur

Automatic **remediation**

Systems Manager Automation

Configuration changes occur and information is sent to **AWS Config**

© Digital Cloud Training | https://digitalcloud.training

AWS Config

Example Rule	Description
s3-bucket-versioning-enabled	Checks whether versioning is enabled for your S3 buckets
restricted-ssh	Checks whether security groups that are in use disallow unrestricted incoming SSH traffic
rds-instance-public-access-check	Checks whether the Amazon Relational Database Service (RDS) instances are not publicly accessible
cloudtrail-enabled	Checks whether AWS CloudTrail is enabled in your AWS account
approved-amis-by-id	Checks whether running instances are using specified AMIs

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager

- AWS Systems Manager is a secure end-to-end management solution for resources on AWS, on premises, and on other clouds
- Systems Manager is an operations hub for operations management, application management, change management and node management
- You can manage Amazon EC2 instances and on-premises servers and virtual machines (VMs), including VMs hosted by other cloud providers
- The SSM Agent must be installed on managed instances and servers
- IAM permissions must be configured to allow management
- For hybrid environments an IAM service role is used with hybrid activation

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager

Systems Manager components include:

- Automation
- Run Command
- Inventory
- Patch Manager
- Compliance
- Session Manager
- Parameter Store

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager Automation

Documents define the actions to perform (YAML or JSON)

Automates **IT operations** and **management** tasks across AWS resources

Documents

Automation

Amazon RDS

```yaml
description: Creates an RDS Snapshot for an RDS instance.
schemaVersion: '0.3'
assumeRole: "{{AutomationAssumeRole}}"
parameters:
  DBInstanceIdentifier:
    type: String
    description: (Required) The DBInstanceId ID of the RDS Instance.
  DBSnapshotIdentifier:
    type: String
    description: (Optional) The DBSnapshotIdentifier ID.
    default: ''
  InstanceTags:
    type: String
    default: ''
    description: (Optional) Tags to create for instance.
  SnapshotTags:
    type: String
    default: ''
    description: (Optional) Tags to create for snapshot
```

This automation, takes a snapshot of an RDS DB instance

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager Run Command

Document types include **command**, **automation**, **package** etc.

Runs commands on managed EC2 instances

Documents

Run Command

Amazon EC2

```
{
  "schemaVersion": "1.2",
  "description": "List missing Microsoft Windows updates.",
  "parameters": {
    "UpdateLevel": {
      "type": "String",
      "default": "Important",
      "description": "Important: .......",
      "allowedValues": [
        "None",
        "All",
        "Important",
        "Optional"
      ]
    },
```

This command checks for missing updates

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager Inventory

Inventory

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager Patch Manager

- Helps you select and deploy operating system and software patches automatically across large groups of Amazon EC2 or on-premises instances
- Patch baselines:
  - Set rules to auto-approve select categories of patches to be installed
  - Specify a list of patches that override these rules and are automatically approved or rejected
- You can also schedule maintenance windows for your patches so that they are only applied during predefined times
- Systems Manager helps ensure that your software is up-to-date and meets your compliance policies

Patch Manager

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager Compliance

- AWS Systems Manager lets you scan your managed instances for patch compliance and configuration inconsistencies
- You can collect and aggregate data from multiple AWS accounts and Regions, and then drill down into specific resources that aren't compliant
- By default, AWS Systems Manager displays data about patching and associations
- You can also customize the service and create your own compliance types based on your requirements (must use the AWS CLI, AWS Tools for Windows PowerShell, or the SDKs)

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager Session Manager

- Secure remote management of your instances at scale without logging into your servers
- Replaces the need for bastion hosts, SSH, or remote PowerShell
- Integrates with IAM for granular permissions
- All actions taken with Systems Manager are recorded by AWS CloudTrail
- Can store session logs in an S3 bucket and output to CloudWatch Logs
- Requires IAM permissions for EC2 instance to access SSM, S3, and CloudWatch Logs

Doesn't require open ports: 22,5985/5986

No need for bastion hosts

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager Parameter Store

- Parameter Store provides secure, hierarchical storage for configuration data management and secrets management
- Highly scalable, available, and durable
- Store data such as passwords, database strings, and license codes as parameter values
- Store values as plaintext (unencrypted data) or ciphertext (encrypted data)
- Reference values by using the unique name that you specified when you created the parameter
- No native rotation of keys (difference with AWS Secrets Manager which does it automatically)

© Digital Cloud Training | https://digitalcloud.training

SSM Automation and Config Rules

© Digital Cloud Training | https://digitalcloud.training

AWS Health API and Dashboards

© Digital Cloud Training | https://digitalcloud.training

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
- Also provides proactive notification to help you plan for scheduled activities

© Digital Cloud Training | https://digitalcloud.training

AWS Service Health Dashboard

Not personalized information so may not be relevant to you

No proactive notification of scheduled activities

Current Status - Jun 7, 2020 PDT

Amazon Web Services publishes our most up-to-the-minute information on service availability in the table below. Check back here any time to get current status information, or subscribe to an RSS feed to be notified of interruptions to each individual service. If you are experiencing a real-time, operational issue with one of our services that is not described below, please inform us by clicking on the "Contact Us" link to submit a service issue report. All dates and times are Pacific Time (PST/PDT).

North America | South America | Europe | Africa | Asia Pacific | Middle East | Contact Us

Recent Events	Details	RSS
No recent events.		

Remaining Services	Details	RSS
Alexa for Business (N. Virginia)	Service is operating normally	
Amazon API Gateway (Montreal)	Service is operating normally	
Amazon API Gateway (N. California)	Service is operating normally	
Amazon API Gateway (N. Virginia)	Service is operating normally	
Amazon API Gateway (Ohio)	Service is operating normally	
Amazon API Gateway (Oregon)	Service is operating normally	
Amazon AppStream 2.0 (N. Virginia)	Service is operating normally	
Amazon AppStream 2.0 (Oregon)	Service is operating normally	
Amazon Athena (Montreal)	Service is operating normally	
Amazon Athena (N. Virginia)	Service is operating normally	
Amazon Athena (Ohio)	Service is operating normally	
Amazon Athena (Oregon)	Service is operating normally	

Shows current status information on service availability

© Digital Cloud Training | https://digitalcloud.training

SECTION 5

Reliability and Business Continuity

© Digital Cloud Training | https://digitalcloud.training

Elasticity: Scaling Up vs Out

© Digital Cloud Training | https://digitalcloud.training

Scaling Up (vertical scaling)

Scaling Up (vertical scaling)

Scaling up means **adding** resources to the instance

Limitation is that you have a **single point of failure** (SPOF)

DigitalCloud TRAINING

Scaling Out (horizontal scaling)

Scaling out provides greater **resiliency**

Scaling out can be used to add almost unlimited capacity

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

Scaling Up (vertical scaling)

© Digital Cloud Training | https://digitalcloud.training

Scaling Out (horizontal scaling)

© Digital Cloud Training | https://digitalcloud.training

AWS Auto Scaling

© Digital Cloud Training | https://digitalcloud.training

AWS Auto Scaling

- Automatically scale AWS services including:
- Amazon EC2 – Launch or terminate EC2 instances
- Amazon EC2 Spot Fleets
- Amazon ECS – Adjust ECS service desired count up/down
- Amazon DynamoDB – increase provisioned RCUs/WCUs
- Amazon Aurora – Adjust the number of Read Replicas

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Auto Scaling

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Auto Scaling

- EC2 Auto Scaling **launches** and **terminates** instances dynamically
- Scaling is horizontal (scales out)
- Provides **elasticity** and **scalability**
- Responds to EC2 status checks and CloudWatch metrics
- Can scale based on demand (performance) or on a schedule
- Scaling policies define how to respond to changes in demand
- Auto Scaling groups define collections of EC2 instances that are scaled and managed together

© Digital Cloud Training | https://digitalcloud.training

Configuration of an Auto Scaling Group

A **Launch Template** specifies the EC2 instance configuration

Launch Template

- AMI and instance type
- EBS volumes
- Security groups
- Key pair
- IAM instance profile
- User data
- Shutdown behavior
- Termination protection
- Placement group name
- Capacity reservation
- Tenancy
- Purchasing option (e.g. Spot)

Launch Config

- AMI and instance type
- EBS volumes
- Security groups
- Key pair
- Purchasing option (e.g. Spot)
- IAM instance profile
- User data

Launch Configurations are replaced by launch templates and have fewer features

Configure purchase options – On-demand vs Spot

Configure VPC and Subnets

Attach Load Balancer

Configure health checks EC2 & ELB

Group size and scaling policies

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Auto Scaling

- Health checks
  - EC2 = EC2 status checks
  - ELB = Uses the ELB health checks **in addition** to EC2 status checks
- Health check grace period
  - How long to wait before checking the health status of the instance
  - Auto Scaling does not act on health checks until grace period expires

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Auto Scaling

Types of Auto Scaling:

- **Dynamic** – automatically scales based on demand
- **Predictive** – uses Machine Learning to predict
- **Scheduled** – scales based on a schedule

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Auto Scaling

Types of Policy:

- **Target tracking scaling**—Increase and decrease the current capacity of the group based on an Amazon CloudWatch metric and a target value. E.g., keeping the aggregate CPU usage of your ASG at 70%
- **Step scaling**—Increase and decrease the current capacity of the group based on a set of scaling adjustments, known as step adjustments, that vary based on the size of the alarm breach
- **Simple scaling**—Increase and decrease the current capacity of the group based on a single scaling adjustment, with a cooldown period between each scaling activity

DigitalCloud TRAINING

Auto Scaling - Monitoring

- Group metrics (ASG)
  - Data points about the Auto Scaling group
  - 1-minute granularity
  - No charge
  - Must be enabled
- Basic monitoring (Instances)
  - 5-minute granularity
  - No Charge
- Detailed monitoring (Instances)
  - 1-minute granularity
  - Charges apply

© Digital Cloud Training | https://digitalcloud.training

Additional Scaling Settings

- **Cooldowns** – Used with simple scaling policy to prevent Auto Scaling from launching or terminating before effects of previous activities are visible. Default value is 300 seconds (5 minutes)

- **Termination Policy** – Controls which instances to terminate first when a scale-in event occurs

- **Termination Protection** – Prevents Auto Scaling from terminating protected instances

- **Standby State** – Used to put an instance in the InService state into the Standby state, update or troubleshoot the instance

© Digital Cloud Training | https://digitalcloud.training

Additional Scaling Settings

- **Lifecycle Hooks** – Used to perform custom actions by pausing instances as the ASG launches or terminates them
- Use case:
  - Run a script to download and install software after launching
  - Pause an instance to process data before a scale-in (termination)

Amazon Elastic Load Balancing

© Digital Cloud Training | https://digitalcloud.training

Amazon Elastic Load Balancing

© Digital Cloud Training | https://digitalcloud.training

Amazon Elastic Load Balancing

EC2 Auto Scaling **launches** instance 5

© Digital Cloud Training | https://digitalcloud.training

Types of Elastic Load Balancer (ELB)

Application Load Balancer

- Operates at the request level
- Routes based on the content of the request (layer 7)
- Supports path-based routing, host-based routing, query string parameter-based routing, and source IP address-based routing
- Supports instances, IP addresses, Lambda functions and containers as targets

Network Load Balancer

- Operates at the connection level
- Routes connections based on IP protocol data (layer 4)
- Offers ultra high performance, low latency and TLS offloading at scale
- Can have a static IP / Elastic IP
- Supports UDP and static IP addresses as targets

© Digital Cloud Training | https://digitalcloud.training

ELB Use Cases

Application Load Balancer

- Web applications with L7 routing (HTTP/HTTPS)
- Microservices architectures (e.g. Docker containers)
- Lambda targets

Network Load Balancer

- TCP and UDP based applications
- Ultra-low latency
- Static IP addresses
- VPC endpoint services

© Digital Cloud Training | https://digitalcloud.training

Create an ASG and ALB with the CLI

© Digital Cloud Training | https://digitalcloud.training

Create a Lifecycle Hook

© Digital Cloud Training | https://digitalcloud.training

Session State and Session Stickiness

© Digital Cloud Training | https://digitalcloud.training

Without Sticky Sessions

© Digital Cloud Training | https://digitalcloud.training

With Sticky Sessions

© Digital Cloud Training | https://digitalcloud.training

Storing Session State

ElastiCache is also a popular solution for storing **session-state data**

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Backup and Recovery

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Automated Backups

Restore can be to **any** point in time during the **retention period**

Restored DB instances are associated with the **default** DB parameter and option groups

Restore

New DB Instance

The DB instance must be in the available state for backups to occur

RDS uploads transaction logs for DB instances to Amazon S3 every **five minutes**

Backup

Amazon S3

DB Instance

Retention is **0–35** days

Automated backups can be **replicated** to any AWS Region

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Manual Backups (Snapshot)

- Backs up the entire DB instance, not just individual databases
- For single-AZ DB instances there is a brief suspension of I/O
- For Multi-AZ SQL Server, I/O activity is briefly suspended on primary
- For Multi-AZ MariaDB, MySQL, Oracle and PostgreSQL the snapshot is taken from the standby
- Snapshots do not expire (no retention period)

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS Maintenance Windows

- Operating system and DB patching can require taking the database offline
- These tasks take place during a maintenance window
- By default, a weekly maintenance window is configured
- You can choose your own maintenance window

© Digital Cloud Training | https://digitalcloud.training

Amazon Aurora Backups

- Aurora backups are continuous and incremental so you can quickly restore to any point within the backup retention period
- For Amazon Aurora DB clusters, the default backup retention period is one day regardless of how the DB cluster is created
- You can't disable automated backups on Aurora. The backup retention period for Aurora is managed by the DB cluster
- You can also use AWS Backup to manage backups of Amazon Aurora DB clusters
- You can also take DB cluster snapshots which capture a snapshot of the entire DB cluster
- Manual snapshots aren't subject to the backup retention period and do not expire
- For very long-term backups, you can export snapshot data to Amazon S3

© Digital Cloud Training | https://digitalcloud.training

Amazon Aurora Backups

- With Amazon Aurora, you can copy automated backups or manual DB cluster snapshots
- You can copy a snapshot within the same AWS Region, you can copy a snapshot across AWS Regions, and you can copy shared snapshots
- Amazon RDS deletes automated backups in several situations:
  - At the end of their retention period
  - When you disable automated backups for a DB cluster
  - When you delete a DB cluster
- If you want to keep an automated backup for a longer period, copy it to create a manual snapshot

© Digital Cloud Training | https://digitalcloud.training

RDS/Aurora Read Replicas and Multi-AZ

© Digital Cloud Training | https://digitalcloud.training

RDS Disaster Recovery (DR) and Scaling Out (Horizontally)

Multi-AZ creates a passive standby. Primarily used for **disaster recovery**

Application servers can **read** from the **read replica** and **write** to the **primary**

Read Replicas are used for scaling database **queries** (reads)

© Digital Cloud Training | https://digitalcloud.training

Aurora Fault Tolerance and Aurora Replicas

Aurora Fault Tolerance

- Fault tolerance across 3 AZs
- Single logical volume
- Aurora Replicas scale-out read requests
- Up to 15 Aurora Replicas with **sub-10ms** replica lag
- Aurora Replicas are independent endpoints
- Can **promote** Aurora Replica to be a new primary or create new primary
- Set priority (tiers) on Aurora Replicas to control order of promotion
- Can use **Auto Scaling** to add replicas

© Digital Cloud Training | https://digitalcloud.training

Cross-Region Replica with Aurora MySQL

© Digital Cloud Training | https://digitalcloud.training

Aurora Global Database

Applications can connect to the cluster **Reader Endpoint**

Replication uses the **Aurora storage layer**

© Digital Cloud Training | https://digitalcloud.training

Aurora Multi-Master

Aurora Multi-Master

- All nodes allow reads/writes
- Available for MySQL only
- Up to four read/write nodes
- Single Region only
- Cannot have cross-Region replicas
- Can work with active-active and active-passive workloads
- Can restart read/write DB instance without impacting other instances

© Digital Cloud Training | https://digitalcloud.training

Aurora Serverless

Each **Aurora Capacity Unit** (ACU) is 2 GB of memory plus CPU

Warm pool of capacity

ACU assigned ->

Capacity seamlessly scales **up** and **down**

© Digital Cloud Training | https://digitalcloud.training

Aurora Serverless Use Cases

- Infrequently used applications
- New applications
- Variable workloads
- Unpredictable workloads
- Development and test databases
- Multi-tenant applications

© Digital Cloud Training | https://digitalcloud.training

DNS and Amazon Route 53

© Digital Cloud Training | https://digitalcloud.training

Amazon Route Features

Amazon Route 53

Domain Registration

.net
.com
.org

Hosted zone

example.com
dctlabs.com

Health Checks

EC2 Instances

Traffic Flow

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

Amazon Route 53 Hosted Zones

Name	Type	Value	TTL
example.com	A	8.1.2.1	60
dev.example.com	A	8.1.2.2	60

Amazon Route 53

This is an example of a **public hosted zone**

A **hosted zone** represents a set of records belonging to a domain

example.com

What's the address for example.com?

Address is 8.1.2.1

HTTP GET to 8.1.2.1

Region

VPC

Availability Zone

Public subnet

Web Server: 8.1.2.1

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

Amazon Route 53 Hosted Zones

Name	Type	Value	TTL
db.mycompany.local	A	10.0.0.10	60
app.mycompany.local	A	10.0.0.11	60

Amazon Route 53

This is an example of a **private hosted zone**

Enable DNS Hostnames = Enabled
DNS Resolution = Enabled

Association with VPC

mycompany.local

What's the address for db.mycompany.local?

Address is 10.0.0.10

VPC

Public subnet

Private subnet

Connection to 10.0.0.10

DB Server: 10.0.0.10

© Digital Cloud Training | https://digitalcloud.training

Migration to/from Route 53

- You can migrate from **another DNS provider** and can import records
- You can migrate a hosted zone to **another AWS account**
- You can migrate from Route 53 to **another registrar**
- You can also associate a Route 53 hosted zone with a **VPC in another account**
  - Authorize association with VPC in the second account.
  - Create an association in the second account

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 Routing Policies

Routing Policy	What it does
Simple	Simple DNS response providing the IP address associated with a name
Failover	If primary is down (based on health checks), routes to secondary destination
Geolocation	Uses geographic location you're in (e.g. Europe) to route you to the closest region
Geoproximity	Routes you to the closest region within a geographic area
Latency	Directs you based on the lowest latency route to resources
Multivalue answer	Returns several IP addresses and functions as a basic load balancer
Weighted	Uses the relative weights assigned to resources to determine which to route to
IP-Based	Route traffic based on the IP addresses that the traffic originates from

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 – Simple Routing Policy

Name	Type	Value	TTL
simple.dctlabs.com	A	1.1.1.1 2.2.2.2	60
simple2.dctlabs.com	A	3.3.3.3	60

Amazon Route 53 – Weighted Routing Policy

Name	Type	Value	Health	Weight
weighted.dctlabs.com	A	1.1.1.1	ID	60
weighted.dctlabs.com	A	2.2.2.2	ID	20
weighted.dctlabs.com	A	3.3.3.3	ID	20

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 – Latency Routing Policy

Name	Type	Value	Health	Region
latency.dctlabs.com	A	1.1.1.1	ID	ap-southeast-1
latency.dctlabs.com	A	2.2.2.2	ID	us-east-1
latency.dctlabs.com	A	alb-id	ID	ap-southeast-2

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 – Failover Routing Policy

Name	Type	Value	Health	Record Type
failover.dctlabs.com	A	1.1.1.1	ID	Primary
failover.dctlabs.com	A	alb-id		Secondary

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 – Geolocation Routing Policy

Name	Type	Value	Health	Geolocation
geolocation.dctlabs.com	A	1.1.1.1	ID	Singapore
geolocation.dctlabs.com	A	2.2.2.2	ID	Default
geolocation.dctlabs.com	A	alb-id	ID	Oceania

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 – Multivalue Routing Policy

Name	Type	Value	Health	Multi Value
multivalue.dctlabs.com	A	1.1.1.1	ID	Yes
multivalue.dctlabs.com	A	2.2.2.2	ID	Yes
multivalue.dctlabs.com	A	3.3.3.3	ID	Yes

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 – IP-Based Routing Policy

Create a **CIDR collection** and **CIDR location**

CIDR location

EU-1

Valid characters: a-z, 0-9, - _

CIDR blocks

List of IP addresses from /0 to /24 for IPv4 and /0 to /48 for IPv6.

1.2.3.0/24

Configure routing to targets based on **CIDR location**

Name	Type	Value	Routing	Differentiator
dctlabs.com	A	1.1.1.1	IP-Based	EU-1
dctlabs.com	A	2.2.2.2	IP-Based	EU-2

Create an **IP-based** routing policy

TTL (seconds) Info

300 | 1m | 1h | 1d

Recommended values: 60 to 172800 (two days)

Routing policy Info

IP-based

IP-based

EU-1

Health check ID - *optional* Info

Choose health check

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

Route 53 Routing Policies and Health Checks

© Digital Cloud Training | https://digitalcloud.training

RPO, RTO, and DR Strategies

© Digital Cloud Training | https://digitalcloud.training

RPO and RTO

Recovery Point Objective (RPO)

- Measurement of the amount of data that can be acceptably lost
- Measured in seconds, minutes, or hours
- Example:
  - You can acceptably lose 2 hours of data in a database (2hr RPO)
  - This means backups must be taken every 2 hours

© Digital Cloud Training | https://digitalcloud.training

RPO and RTO

Recovery Time Objective (RTO)

- Measurement of the amount of time it takes to restore after a disaster event
- Measured in seconds, minutes, or hours
- Example:
  - The IT department expect it to take 4 hours to bring applications online after a disaster
  - This would be an RTO of 4 hours

© Digital Cloud Training | https://digitalcloud.training

RPO and RTO

© Digital Cloud Training | https://digitalcloud.training

Achievable RPOs

Recovery Point Objective	Technique
Milliseconds - Seconds	Synchronous replication
Seconds - Minutes	Asynchronous replication
Minutes to hours	Snapshots, cloud backup, D2D
Hours to days	Offsite / traditional backups / tape backups

The **RPO** is determined by how you take a backup of data

© Digital Cloud Training | https://digitalcloud.training

Achievable RTOs

Recovery Time Objective	Technique
Milliseconds / Seconds	Fault tolerance
Seconds - Minutes	High availability, load balancing, auto scaling
Minutes to hours	Cross-site recovery (cloud), automated recovery
Hours to days	Cross-site recovery (cloud/on-premises), manual recovery

The **RTO** is determined by how quickly you can recover

© Digital Cloud Training | https://digitalcloud.training

DR Strategies

Backup and Restore

- Low priority workloads
- Provision/restore after event
- Cost $

Pilot Light

- Data replicated
- Services idle/off
- Resources activated after event
- Cost $$

Warm standby

- Minimum resources always running
- Business critical workloads
- Scale up/out after event
- Cost $$$

Multi-site active/active

- Zero downtime
- Near zero data loss
- Mission critical workloads
- Cost $$$$

© Digital Cloud Training | https://digitalcloud.training

AWS Backup and DLM

© Digital Cloud Training | https://digitalcloud.training

AWS Backup

- Fully managed service for automating data protection
- Centralized backup management for many services:
  - Compute services such as EC2
  - Storage services such as EBS, S3, FSx, and EFS
  - Database services such as RDS, DynamoDB, and DocumentDB
  - Hybrid applications like VMware workloads on-premises or on VMware Cloud or Outposts
- Policy-based data protection
- Managed through the console, API, and CLI
- Backups are encrypted with AWS KMS keys
- Also provide protection from ransomware

© Digital Cloud Training | https://digitalcloud.training

Amazon EBS Snapshots

© Digital Cloud Training | https://digitalcloud.training

Amazon Data Lifecycle Manager (DLM)

- DLM automates the creation, retention, and deletion of EBS snapshots and EBS-backed AMIs

- DLM helps with the following:
  - Protects valuable data by enforcing a regular backup schedule
  - Create standardized AMIs that can be refreshed at regular intervals
  - Retain backups as required by auditors or internal compliance
  - Reduce storage costs by deleting outdated backups
  - Create disaster recovery backup policies that back up data to isolated accounts

© Digital Cloud Training | https://digitalcloud.training

Implement EBS Volume Lifecycle

© Digital Cloud Training | https://digitalcloud.training

Amazon S3: Durability, Availability, and Lifecycle

© Digital Cloud Training | https://digitalcloud.training

Durability and Availability in S3

Durability

Durability is protection against:

- Data loss
- Data corruption
- S3 offers 11 9s durability (99.999999999)

If you store 10 million objects, then you expect to lose one object every 10,000 years!

Availability

Availability is a measurement of:

- The amount of time the data is available to you
- Expressed as a percent of time per year
- E.g. 99.99%

© Digital Cloud Training | https://digitalcloud.training

S3 Storage Classes

	S3 Standard	S3 Intelligent Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
Designed for durability	99.999999999%	99.999999999%	99.999999999%	99.999999999%	99.999999999%	99.999999999%	99.999999999%
Designed for availability	99.99%	99.9%	99.9%	99.5%	99.9%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99%	99.9%	99.9%
Availability Zones	≥3	≥3	≥3	1	≥3	≥3	≥3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	N/A	30 days	30 days	90 days	90 days	180 days
Retrieval fee	N/A	N/A	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	milliseconds	minutes or hours	hours
Storage type	Object	Object	Object	Object	Object	Object	Object
Lifecycle transitions	Yes	Yes	Yes	Yes	Yes	Yes	Yes

© Digital Cloud Training | https://digitalcloud.training

S3 Storage Class Use Cases

- S3 Standard
- S3 Intelligent Tiering
- S3 Standard-IA
- S3 One Zone-IA
- S3 Glacier IR
- S3 Glacier FR
- S3 Glacier Deep Archive

Require long-term storage for production data with low cost and immediate access for occasional requests

Copy of backup data required in a separate Region with minimal access latency

Frequently accessed data; ad-hoc needs; short-term requirements (<30 days)

Lowest cost required for long term data with instant retrieval

Lowest cost required for long-term archival of data for compliance purposes

© Digital Cloud Training | https://digitalcloud.training

S3 Replication

Cross-Region Replication (CRR)

Buckets must have **versioning** enabled

© Digital Cloud Training | https://digitalcloud.training

S3 Encryption

- S3 default encryption is enabled by default
- objects are encrypted using server-side encryption
- Server-side encryption is free and does not impact performance
- You can choose to use:
  - Amazon S3 managed keys (SSE-S3)
  - AWS KMS keys (SSE-KMS)
- Amazon S3 encrypts an object before saving it to disk and decrypts it when you download the objects

© Digital Cloud Training | https://digitalcloud.training

S3 Lifecycle Management

There are two types of actions:

- **Transition actions** - Define when objects transition to another storage class
- **Expiration actions** - Define when objects expire (deleted by S3)

© Digital Cloud Training | https://digitalcloud.training

S3 LM: Supported Transitions

https://docs.aws.amazon.com/AmazonS3/latest/userguide/lifecycle-transition-general-considerations.html

© Digital Cloud Training | https://digitalcloud.training

S3 Versioning and Lifecycle Management

© Digital Cloud Training | https://digitalcloud.training

SECTION 6

Deployment, Provisioning, and Automation

© Digital Cloud Training | https://digitalcloud.training

AWS CloudFormation

© Digital Cloud Training | https://digitalcloud.training

AWS CloudFormation

Infrastructure patterns are defined in a **template** file using **code**

CloudFormation **builds** your infrastructure according to the **template**

AWS CloudFormation

```yaml
Description:  This template deploys a VPC, with a pair of public and private subnets spread
  across two Availability Zones. It deploys an internet gateway, with a default
  route on the public subnets.

Parameters:
  EnvironmentName:
    Description: An environment name that is prefixed to resource names
    Type: String

  VpcCIDR:
    Description: Please enter the IP range (CIDR notation) for this VPC
    Type: String
    Default: 10.192.0.0/16

  PublicSubnet1CIDR:
    Description: Please enter the IP range (CIDR notation) for the public subnet in the first Availability Zone
    Type: String
    Default: 10.192.10.0/24

  PrivateSubnet1CIDR:
    Description: Please enter the IP range (CIDR notation) for the private subnet in the first Availability Zone
    Type: String
    Default: 10.192.20.0/24
```

© Digital Cloud Training | https://digitalcloud.training

AWS CloudFormation - Benefits

- Infrastructure is provisioned consistently, with fewer mistakes (human error)
- Less time and effort than configuring resources manually
- You can use version control and peer review for your CloudFormation templates
- Free to use (you're only charged for the resources provisioned)
- Can be used to manage updates and dependencies
- Can be used to rollback and delete the entire stack as well

© Digital Cloud Training | https://digitalcloud.training

AWS CloudFormation

Component	Description
Templates	The JSON or YAML text file that contains the instructions for building out the AWS environment
Stacks	The entire environment described by the template and created, updated, and deleted as a single unit
StackSets	AWS CloudFormation StackSets extends the functionality of stacks by enabling you to create, update, or delete stacks across multiple accounts and regions with a single operation
Change Sets	A summary of proposed changes to your stack that will allow you to see how those changes might impact your existing resources before implementing them

© Digital Cloud Training | https://digitalcloud.training

Deploy Stacks and Change Sets

© Digital Cloud Training | https://digitalcloud.training

Stacks, Stack Sets, and Nested Stacks

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Stacks

- Deployed resources based on **templates**
- You create, update and delete ***stacks*** using templates
- Stacks are deployed through the Management Console, CLI or APIs
- Stack creation errors:
  - Automatic rollback on error is enabled by default
  - You will be charged for resources provisioned even if there is an error

© Digital Cloud Training | https://digitalcloud.training

CloudFormation StackSets

- Create, update, or delete stacks across **multiple accounts** and **Regions** with a single operation
- An administrator account is the AWS account in which you create StackSets
- A stack set is managed by signing into the AWS administrator account in which it was created
- A target account is the account into which you create, update, or delete one or more stacks in your stack set

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Nested Stacks

- Nested stacks allow re-use of CloudFormation code for common use cases
- For example, a standard configuration for a load balancer, web server, application server etc.
- Create a standard template for each common use case and reference it from within your CloudFormation template

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Template Deep Dive

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Templates

- A template is a **YAML** or **JSON** template used to describe the end-state of the infrastructure you are provisioning or changing
- After creating the template, you upload it to CloudFormation directly or using Amazon S3
- CloudFormation reads the template and makes the API calls on your behalf
- The resulting resources are called a "Stack"
- **Logical IDs** – used to reference resources within the template
- **Physical IDs** – identify resources outside of AWS CloudFormation templates, but only after the resources have been created

© Digital Cloud Training | https://digitalcloud.training

Intrinsic Functions

- AWS CloudFormation provides several built-in functions that help you manage your stacks
- Use intrinsic functions in your templates to assign values to properties that are not available until runtime
- The full list can be found here:
- https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/intrinsic-function-reference.html

© Digital Cloud Training | https://digitalcloud.training

Intrinsic Functions - Ref

- The intrinsic function Ref returns the value of the specified parameter or resource
- When you specify a parameter's logical name, it returns the value of the parameter
- When you specify a resource's logical name, it returns a value that you can typically use to refer to that resource, such as a physical ID

© Digital Cloud Training | https://digitalcloud.training

Intrinsic Functions - Ref

- The following resource declaration for an Elastic IP address needs the instance ID of an EC2 instance and uses the Ref function to specify the instance ID of the MyEC2Instance resource:

```yaml
MyEIP:
  Type: "AWS::EC2::EIP"
  Properties:
    InstanceId: !Ref MyEC2Instance
```

© Digital Cloud Training | https://digitalcloud.training

Intrinsic Functions - Fn::GetAtt

- The Fn::GetAtt intrinsic function returns the value of an attribute from a resource in the template
- Full syntax (YAML):
  - *Fn::GetAtt: [logicalNameOfResource, attributeName]*
- Short form (YAML):
  - *!GetAtt logicalNameOfResource.attributeName*

© Digital Cloud Training | https://digitalcloud.training

Intrinsic Functions - Fn::GetAtt

The example template returns the following for the ELB named MyELB:

- SourceSecurityGroup.OwnerAlias
- SourceSecurityGroup.GroupName

```yaml
AWSTemplateFormatVersion: 2010-09-09
Resources:
  myELB:
    Type: AWS::ElasticLoadBalancing::LoadBalancer
    Properties:
      AvailabilityZones:
        - eu-west-1a
      Listeners:
        - LoadBalancerPort: '80'
          InstancePort: '80'
          Protocol: HTTP
  myELBIngressGroup:
    Type: AWS::EC2::SecurityGroup
    Properties:
      GroupDescription: ELB ingress group
      SecurityGroupIngress:
        - IpProtocol: tcp
          FromPort: '80'
          ToPort: '80'
          SourceSecurityGroupOwnerId: !GetAtt myELB.SourceSecurityGroup.OwnerAlias
          SourceSecurityGroupName: !GetAtt myELB.SourceSecurityGroup.GroupName
```

© Digital Cloud Training | https://digitalcloud.training

Intrinsic Functions - Fn::FindInMap

- The intrinsic function Fn::FindInMap returns the value corresponding to keys in a two-level map that is declared in the Mappings section
- Full syntax (YAML):
  - *Fn::FindInMap: [MapName, TopLevelKey, SecondLevelKey]*
- Short form (YAML):
  - *!FindInMap [MapName, TopLevelKey, SecondLevelKey]*

© Digital Cloud Training | https://digitalcloud.training

Intrinsic Functions - Fn::FindInMap

- The example shows how to use Fn::FindInMap for a template with a Mappings section that contains a single map (RegionMap) that associates AMIs with AWS regions:

```yaml
Mappings:
  RegionMap:
    us-east-1:
      HVM64: "ami-0ff8a91507f77f867"
      HVMG2: "ami-0a584ac55a7631c0c"
    us-west-1:
      HVM64: "ami-0bdb828fd58c52235"
      HVMG2: "ami-066ee5fd4a9ef77f1"
Resources:
  myEC2Instance:
    Type: "AWS::EC2::Instance"
    Properties:
      ImageId: !FindInMap
        - RegionMap
        - !Ref 'AWS::Region'
        - HVM64
      InstanceType: m1.small
```

© Digital Cloud Training | https://digitalcloud.training

Template Sections - Resources

- The required Resources section declares the AWS resources that you want to include in the stack, such as an Amazon EC2 instance or an Amazon S3 bucket
- This is a mandatory section
- Resources are declared and can reference each other

```yaml
Resources:
 MyEC2Instance:
   Type: "AWS::EC2::Instance"
   Properties:
     ImageId: "ami-0ff8a91507f77f867"
```

© Digital Cloud Training | https://digitalcloud.training

Template Sections - Parameters

- Use the *optional* Parameters section to customize your templates
- Parameters enable you to input custom values to your template each time you create or update a stack
- Useful for template reuse

```yaml
Parameters:
  InstanceTypeParameter:
    Type: String
    Default: t2.micro
    AllowedValues:
      - t2.micro
      - m1.small
      - m1.large
    Description: Enter t2.micro, m1.small, or m1.large. Default is t2.micro.
```

© Digital Cloud Training | https://digitalcloud.training

Template Sections - Mappings

- The *optional* Mappings section matches a key to a corresponding set of named values

```yaml
RegionMap:
    us-east-1:
      HVM64: ami-0ff8a91507f77f867
      HVMG2: ami-0a584ac55a7631c0c
    us-west-1:
      HVM64: ami-0bdb828fd58c52235
      HVMG2: ami-066ee5fd4a9ef77f1
```

Exam tip: with mappings you can set values based on a region. You can create a mapping that uses the region name as a key and contains the values you want to specify for each specific region

© Digital Cloud Training | https://digitalcloud.training

Template Sections - Outputs

- The *optional* Outputs section declares output values that you can import into other stacks (to create cross-stack references), return in response (to describe stack calls), or view on the AWS CloudFormation console
- In the following example YAML code, the output named StackVPC returns the ID of a VPC, and then exports the value for cross-stack referencing with the name VPCID appended to the stack's name

```yaml
Outputs:
  StackVPC:
    Description: The ID of the VPC
    Value: !Ref MyVPC
    Export:
      Name: !Sub "${AWS::StackName}-VPCID"
```

© Digital Cloud Training | https://digitalcloud.training

Template Sections - Conditions

- The *optional* Conditions section contains statements that define the circumstances under which entities are created or configured
- In the sample YAML code below, resources are created only if the EnvType parameter is equal to prod:

```yaml
Conditions:
  CreateProdResources: !Equals [ !Ref EnvType, prod ]
```

© Digital Cloud Training | https://digitalcloud.training

Template Sections - Transform

- The *optional* Transform section specifies one or more macros that AWS CloudFormation uses to process your template
- The transform section can be used to reference additional code stored in S3, such as Lambda code or reusable snippets of CloudFormation code

© Digital Cloud Training | https://digitalcloud.training

Template Sections - Transform

- The AWS::Serverless transform specifies the version of the AWS Serverless Application Model (AWS SAM) to use
- This model defines the AWS SAM syntax that you can use and how AWS CloudFormation processes it
- The AWS::Include transform works with template snippets that are stored separately from the main AWS CloudFormation template
- In the following example, the template uses AWS SAM syntax to simplify the declaration of a Lambda function and its execution role:

```yaml
Transform: AWS::Serverless-2016-10-31
Resources:
  MyServerlessFunctionLogicalID:
    Type: AWS::Serverless::Function
    Properties:
      Handler: index.handler
      Runtime: nodejs8.10
      CodeUri: 's3://testBucket/mySourceCode.zip'
```

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Helper Scripts (cfn-init and cfn-signal)

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Helper Scripts (cfn-init)

- The cfn-init helper script reads template metadata from the AWS::CloudFormation::Init key and acts accordingly to:
  - Fetch and parse metadata from AWS CloudFormation
  - Install packages
  - Write files to disk
  - Enable/disable and start/stop services
- Logs go to /var/log/cfn-init.log

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Helper Scripts (cfn-init)

- To install the applications the UserData property and Metadata property can be added to a template

```
"Properties": {
        "ImageId" ami-02354e95b39ca8dec
        "InstanceType"   : { "Ref" : "InstanceType" },
        "SecurityGroups" : [ {"Ref" : "WebServerSecurityGroup"} ],
        "KeyName"        : { "Ref" : "KeyName" },
        "UserData"       : { "Fn::Base64" : { "Fn::Join" : ["", [
          "#!/bin/bash -xe\n",
          "yum install -y aws-cfn-bootstrap\n",

          "# Install the files and packages from the metadata\n",
          "/opt/aws/bin/cfn-init ",
          "         --stack ", { "Ref" : "AWS::StackName" },
          "         --resource WebServerInstance ",
          "         --configsets InstallAndRun ",
          "         --region ", { "Ref" : "AWS::Region" }, "\n"
```

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Helper Scripts (cfn-signal)

- The cfn-signal helper script signals AWS CloudFormation to indicate whether Amazon EC2 instances have been successfully created or updated
- After installing software on EC2 instances, you can signal AWS CloudFormation when those software applications are ready
- You use the cfn-signal script in conjunction with a CreationPolicy or an Auto Scaling group with a WaitOnResourceSignals update policy

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Helper Scripts (cfn-signal)

- In the UserData property, the template runs the cfn-signal script to send a success signal with an exit code if all the services are configured and started successfully

```
"# Install the files and packages from the metadata\n",
"/opt/aws/bin/cfn-init ",
"         --stack ", { "Ref" : "AWS::StackName" },
"         --resource WebServerInstance ",
"         --configsets InstallAndRun ",
"         --region ", { "Ref" : "AWS::Region" }, "\n"
"# Signal the status from cfn-init\n",
  "/opt/aws/bin/cfn-signal -e $? ",
  "         --stack ", { "Ref" : "AWS::StackName" },
  "         --resource WebServerInstance ",
  "         --region ", { "Ref" : "AWS::Region" }, "\n"
```

© Digital Cloud Training | https://digitalcloud.training

CloudFormation Helper Scripts (cfn-init and cfn-signal)

Troubleshooting errors:

- Make sure the AMI has the CloudFormation helper scripts included
- Check that the cfn-init and cfn-signal commands have run successfully

© Digital Cloud Training | https://digitalcloud.training

Creation and Deletion Policies, DependsOn, and WaitConditions

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – CreationPolicy

- Use the CreationPolicy attribute when you want to wait on resource configuration actions before stack creation proceeds
- You can associate the CreationPolicy attribute with a resource to prevent its status from reaching create complete until AWS CloudFormation receives a specified number of success signals or the timeout period is exceeded
- To signal a resource, you can use the cfn-signal helper script or SignalResource API
- AWS CloudFormation publishes valid signals to the stack events so that you track the number of signals sent

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – CreationPolicy

The following CloudFormation resources support creation policies:

- AWS::AutoScaling::AutoScalingGroup
- AWS::EC2::Instance
- AWS::CloudFormation::WaitCondition

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – DeletionPolicy

- With the DeletionPolicy attribute you can preserve or (in some cases) backup a resource when its stack is deleted
- You specify a DeletionPolicy attribute for each resource that you want to control
- If a resource has no DeletionPolicy attribute, AWS CloudFormation deletes the resource by default
- Deletion policies can be specified as:
  - DeletionPolicy=Retain – preserves the resources
  - DeletionPolicy=Snapshot – takes a snapshot (e.g. for EC2, ElastiCache, RDS)
  - DeletionPolicy=Delete – default, attempts to delete the resources

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – DependsOn

- With the DependsOn attribute you can specify that the creation of a specific resource follows another
- When you add a DependsOn attribute to a resource, that resource is created *only after* the creation of the resource specified in the DependsOn attribute

```yaml
AWSTemplateFormatVersion: '2010-09-09'
Mappings:
  RegionMap:
    us-east-1:
      AMI: ami-0ff8a91507f77f867
    us-west-1:
      AMI: ami-0bdb828fd58c52235
Resources:
  Ec2Instance:
    Type: AWS::EC2::Instance
    Properties:
      ImageId:
        Fn::FindInMap:
        - RegionMap
        - Ref: AWS::Region
        - AMI
    DependsOn: myDB
  myDB:
    Type: AWS::RDS::DBInstance
    Properties:
      AllocatedStorage: '5'
      DBInstanceClass: db.t2.small
      Engine: MySQL
      EngineVersion: '5.5'
      MasterUsername: MyName
      MasterUserPassword: MyPassword
```

EC2 instance creation **depends on** the DB creation

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – WaitCondition

- Use a WaitCondition to ensure resources are ready
- You can use a wait condition for situations like the following:
  - To coordinate stack resource creation with configuration actions that are external to the stack creation
  - To track the status of a configuration process

```
"WaitCondition": {
    "Type": "AWS::CloudFormation::WaitCondition",
    "DependsOn": "WebServerGroup",
    "Properties": {
        "Handle": {
            "Ref": "WaitHandle"
        },
        "Timeout": "300",
        "Count": {
            "Ref": "WebServerCapacity"
        }
    }
}
```

WaitCondition that waits for the desired number of instances in a web server group (not shown)

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – UpdatePolicy and UpdateReplacePolicy

- Use the UpdatePolicy attribute to specify how AWS CloudFormation handles updates to the following resources:
  - AWS::AutoScaling::AutoScalingGroup,
  - AWS::ElastiCache::ReplicationGroup
  - AWS::Elasticsearch::Domain
  - AWS::Lambda::Alias
- Use the UpdateReplacePolicy attribute to retain or (in some cases) backup the existing physical instance of a resource when it is replaced during a stack update operation

© Digital Cloud Training | https://digitalcloud.training

Rollbacks and Stack Creation Failures

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – Rollbacks and Stack Creation Failures

Stack creation failures:

- By default everything will be deleted
- Can modify the OnFailure attribute for a stack
- OnFailure must be one of:
  - DO_NOTHING – leaves the resources in place (good for troubleshooting)
  - ROLLBACK – rolls the stack back
  - DELETE – deletes the resources

© Digital Cloud Training | https://digitalcloud.training

CloudFormation – Rollbacks and Stack Creation Failures

Stack update failures:

- A stack goes into the UPDATE_ROLLBACK_FAILED state when AWS CloudFormation cannot roll back all changes during an update
- The stack will automatically roll back to the previous known working state
- When a stack is in the UPDATE_ROLLBACK_FAILED state, you can continue to roll it back to a working state (UPDATE_ROLLBACK_COMPLETE)
- You can't update a stack that is in the UPDATE_ROLLBACK_FAILED state
- However, if you can continue to roll it back, you can return the stack to its original settings and then try to update it again

© Digital Cloud Training | https://digitalcloud.training

Deploy Advanced Template

© Digital Cloud Training | https://digitalcloud.training

AWS Resource Access Manager (RAM)

© Digital Cloud Training | https://digitalcloud.training

AWS RAM

- Shares resources:
  - Across AWS accounts
  - Within AWS Organizations or OUs
  - IAM roles and IAM users
- Resource shares are created with:
  - The AWS RAM Console
  - AWS RAM APIs
  - AWS CLI
  - AWS SDKs

© Digital Cloud Training | https://digitalcloud.training

AWS RAM

RAM can be used to share:

- AWS App Mesh
- Amazon Aurora
- AWS Certificate Manager Private Certificate Authority
- AWS CodeBuild
- Amazon EC2
- EC2 Image Builder
- AWS Glue
- AWS License Manager
- AWS Network Firewall
- AWS Outposts
- Amazon S3 on Outposts
- AWS Resource Groups
- Amazon Route 53
- AWS Systems Manager Incident Manager
- Amazon VPC

© Digital Cloud Training | https://digitalcloud.training

AWS Elastic Beanstalk

© Digital Cloud Training | https://digitalcloud.training

AWS Elastic Beanstalk

© Digital Cloud Training | https://digitalcloud.training

AWS Elastic Beanstalk

- Supports Java, .NET, PHP, Node.js, Python, Ruby, Go, and Docker web applications
- Supports the following languages and development stacks:
  - Apache Tomcat for Java applications
  - Apache HTTP Server for PHP applications
  - Apache HTTP Server for Python applications
  - Nginx or Apache HTTP Server for Node.js applications
  - Passenger or Puma for Ruby applications
  - Microsoft IIS 7.5, 8.0, and 8.5 for .NET applications
  - Java SE
  - Docker
  - Go

© Digital Cloud Training | https://digitalcloud.training

AWS Elastic Beanstalk

There are several **layers**

Applications:

- Contain environments, environment configurations, and application versions
- You can have multiple application versions held within an application

APPLICATION

© Digital Cloud Training | https://digitalcloud.training

AWS Elastic Beanstalk

Application version

- A specific reference to a section of deployable code
- The application version will point typically to an Amazon S3 bucket containing the code

© Digital Cloud Training | https://digitalcloud.training

AWS Elastic Beanstalk

Environments:

- An application version that has been deployed on AWS resources
- The resources are configured and provisioned by AWS Elastic Beanstalk
- The environment is comprised of all the resources created by Elastic Beanstalk and not just an EC2 instance with your uploaded code

© Digital Cloud Training | https://digitalcloud.training

Web Servers and Workers

- **Web servers** are standard applications that listen for and then process HTTP requests, typically over port 80
- **Workers** are specialized applications that have a background processing task that listens for messages on an Amazon SQS queue
- **Workers** should be used for long-running tasks

© Digital Cloud Training | https://digitalcloud.training

AWS Elastic Beanstalk

Elastic Beanstalk Deployment Scenarios

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates

- **All at once:** Deploys the new version to all instances simultaneously
- **Rolling:** Update a batch of instances, and then move onto the next batch once the first batch is healthy
- **Rolling with additional batch:** Like Rolling but launches new instances in a batch ensuring that there is full availability
- **Immutable:** Launches new instances in a new ASG and deploys the version update to these instances before swapping traffic to these instances once healthy
- **Blue/green:** Create a new "stage" environment and deploy updates there

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – All at Once Update

Update is applied to **all instances** simultaneously

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – All at Once Update

- Deploys the new version to all instances simultaneously
- All of your instances are out of service while the deployment takes place
- Fastest deployment
- Good for quick iterations in development environment
- You will experience an outage while the deployment is taking place - not ideal for mission-critical systems
- If the update fails, you need to roll back the changes by re-deploying the original version to all of your instances
- No additional cost

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Rolling Update

Update is applied to **a batch** of instances

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Rolling Update

- Update a few instances at a time (batch), and then move onto the next batch once the first batch is healthy (downtime for 1 batch at a time)
- Application is running both versions simultaneously
- Each batch of instances is taken out of service while the deployment takes place
- Your environment capacity will be reduced by the number of instances in a batch while the deployment takes place
- Not ideal for performance-sensitive systems
- If the update fails, you need to perform an additional rolling update to roll back the changes.
- No additional cost
- Long deployment time

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Rolling with Additional Batch Update

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Rolling with Additional Batch Update

- Like Rolling but launches new instances in a batch ensuring that there is full availability
- Application is running at capacity
- Can set the batch size
- Application is running both versions simultaneously
- Small additional cost
- Additional batch is removed at the end of the deployment
- Longer deployment
- Good for production environments

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Immutable Update

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Immutable Update

- Launches new instances in a new ASG and deploys the version update to these instances before swapping traffic to these instances once healthy
- Zero downtime
- New code is deployed to new instances using an ASG
- High cost as double the number of instances running during updates
- Longest deployment
- Quick rollback in case of failures
- Great for production environments

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Blue/green

© Digital Cloud Training | https://digitalcloud.training

Deploying Updates – Blue/green

- This is not a feature within Elastic Beanstalk
- You create a new "staging" environment and deploy updates there
- The new environment (green) can be validated independently, and you can roll back if there are issues
- Route 53 can be setup using weighted policies to redirect a a percentage of traffic to the staging environment
- Using Elastic Beanstalk, you can "swap URLs" when done with the environment test
- Zero downtime

© Digital Cloud Training | https://digitalcloud.training

Swap URLs

Create additional environment for our **Dev App v2**

Administrator

© Digital Cloud Training | https://digitalcloud.training

Swap URLs

© Digital Cloud Training | https://digitalcloud.training

Environment Tiers: Web Servers and Workers

© Digital Cloud Training | https://digitalcloud.training

Environment Tiers: Web Servers and Workers

- Determines how Elastic Beanstalk provisions resources based on what the application is designed to do
- Consists of Web Servers and Workers:
  - **Web servers** are standard applications that listen for and then process HTTP requests, typically over port 80
  - **Workers** are specialized applications that have a background processing task that listens for messages on an Amazon SQS queue
  - Workers should be used for *long-running* tasks

© Digital Cloud Training | https://digitalcloud.training

Deploy Beanstalk Application

© Digital Cloud Training | https://digitalcloud.training

AWS OpsWorks

© Digital Cloud Training | https://digitalcloud.training

AWS OpsWorks

- AWS OpsWorks is a configuration management service that provides managed instances of Chef and Puppet
- Updates include patching, updating, backup, configuration and compliance management

© Digital Cloud Training | https://digitalcloud.training

Run EventBridge Event on a Schedule

© Digital Cloud Training | https://digitalcloud.training

Create and Manage AMIs

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway – File Gateway

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway – File Gateway

- File gateway provides a virtual **on-premises file server**
- Store and retrieve files as objects in Amazon S3
- Use with on-premises applications, and EC2-based applications that need file storage in S3 for object-based workloads
- File gateway offers **SMB** or **NFS**-based access to data in Amazon S3 with local caching

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway - Volume Gateway

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway - Volume Gateway

- The volume gateway supports block-based volumes
- Block storage – iSCSI protocol
- **Cached Volume mode** – the entire dataset is stored on S3 and a cache of the most frequently accessed data is cached on-site
- **Stored Volume mode** – the entire dataset is stored on-site and is asynchronously backed up to S3 (EBS point-in-time snapshots). Snapshots are incremental and compressed

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway - Tape Gateway

© Digital Cloud Training | https://digitalcloud.training

AWS Storage Gateway - Tape Gateway

- Used for backup with popular backup software
- Each gateway is preconfigured with a media changer and tape drives. Supported by NetBackup, Backup Exec, Veeam etc.
- When creating virtual tapes, you select one of the following sizes: 100 GB, 200 GB, 400 GB, 800 GB, 1.5 TB, and 2.5 TB
- A tape gateway can have up to 1,500 virtual tapes with a maximum aggregate capacity of 1 PB
- All data transferred between the gateway and AWS storage is encrypted using SSL
- All data stored by tape gateway in S3 is encrypted server-side with Amazon S3-Managed Encryption Keys (SSE-S3)

© Digital Cloud Training | https://digitalcloud.training

Create DynamoDB Table with Data

© Digital Cloud Training | https://digitalcloud.training

Amazon DynamoDB

- Fully managed NoSQL database service
- Key/value store and document store
- Non-relational database
- Fully serverless
- Seamless scaling

Data is stored in **partitions** which are replicated across multiple AZs in a Region

DynamoDB Table

© Digital Cloud Training | https://digitalcloud.training

Amazon DynamoDB

- DynamoDB provides low latency (milliseconds)
- Microsecond latency can be achieved with DynamoDB Accelerator (DAX)
- All data is stored on SSD storage
- DynamoDB Global Tables synchronizes tables across Regions

© Digital Cloud Training | https://digitalcloud.training

SECTION 7

Security and Compliance

© Digital Cloud Training | https://digitalcloud.training

Advanced IAM: RBAC and ABAC

© Digital Cloud Training | https://digitalcloud.training

Role-Based Access Control (RBAC)

© Digital Cloud Training | https://digitalcloud.training

Role-Based Access Control (RBAC)

AWS Managed Policies:

- Administrator
- Billing
- Database administrator
- Data scientist
- Developer power user
- Network administrator
- Security auditor
- Support user
- System administrator
- View-only user

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "aws-portal:*Billing",
                "aws-portal:*Usage",
                "aws-portal:*PaymentMethods",
                "budgets:ViewBudget",
                "budgets:ModifyBudget",
                "ce:UpdatePreferences",
                "ce:CreateReport",
                "ce:UpdateReport",
                "ce:DeleteReport",
                "ce:CreateNotificationSubscription",
                "ce:UpdateNotificationSubscription",
                "ce:DeleteNotificationSubscription",
                "cur:DescribeReportDefinitions",
                "cur:PutReportDefinition",
                "cur:ModifyReportDefinition",
                "cur:DeleteReportDefinition",
                "purchase-orders:*PurchaseOrders"
            ],
            "Resource": "*"
        }
    ]
}
```

© Digital Cloud Training | https://digitalcloud.training

Attribute-Based Access Control (ABAC)

Tag Key	Tag Value
Department	DBAdmins

```
"Effect": "Allow",
"Action": [
    "rds:RebootDBInstance",
    "rds:StartDBInstance",
    "rds:StopDBInstance"
],
"Resource": "*",
"Condition": {
    "StringEquals": {
        "aws:PrincipalTag/Department": "DBAdmins",
        "rds:db-tag/Environment": "Production"
```

© Digital Cloud Training | https://digitalcloud.training

Identity Based vs Resource Based Policies

Identity-Based Policies

Resource-Based Policies

```
{
    "Version": "2012-10-17",
    "Id": "Policy1561964929358",
    "Statement":[
    {
        "Sid": "Stmt1561964454052",
        "Effect": "Allow",
        "Principal": {
            "AWS": "arn:aws:iam::515148227241:user/Paul"
        },
        "Action": "s3:*",
        "Resource": "arn:aws:s3:::dctcompany"
    }
    ]
}
```

© Digital Cloud Training | https://digitalcloud.training

Reading IAM Policies

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:*",
                "dynamodb:Describe*",
                "dynamodb:List*",
                "dynamodb:GetItem"
            ],
            "Resource": [
                "arn:aws:s3:::mys3bucket",
                "arn:aws:s3:::mys3bucket/*",
                "arn:aws:dynamodb:us-east-1:111122223333:table/mytable"
            ]
        }
    ]
}
```

The effect is either allow or deny

* is a wildcard (means everything)

Actions are the specific resource operations that the policy applies to

Resources are the specific resources that the policy allows or denies access to

© Digital Cloud Training | https://digitalcloud.training

Example Policy 1

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": "ec2:*",
            "Resource": "*",
            "Effect": "Allow",
            "Condition": {
                "StringEquals": {
                    "ec2:Region": "us-east-2"
                }
            }
        }
    ]
}
```

The policy is only effective in the us-east-2 Region due to the condition

© Digital Cloud Training | https://digitalcloud.training

Example Policy 2

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": ["ec2:TerminateInstances"],
            "Resource": ["*"]
        },
        {
            "Effect": "Deny",
            "Action": ["ec2:TerminateInstances"],
            "Condition": {
                "NotIpAddress": {
                    "aws:SourceIp": [
                        "192.0.2.0/24",
                        "203.0.113.0/24"
                    ]
                }
            },
            "Resource": ["*"]
        }
    ]
}
```

Instance termination is only possible if the IP address of the requester is within the specific ranges

© Digital Cloud Training | https://digitalcloud.training

IAM Policy Evaluation

© Digital Cloud Training | https://digitalcloud.training

Evaluation Logic

© Digital Cloud Training | https://digitalcloud.training

Steps for Authorizing Requests to AWS

1. Authentication – AWS authenticates the principal that makes the request

User

Identity-based policy

3. Evaluating all policies within the account

AWS IAM

Resource-based policy

Console

CLI

API

Request context:

- **Actions** – the actions or operations the principal wants to perform
- **Resources** – The AWS resource object upon which actions are performed
- **Principal** – The user, role, federated user, or application that sent the request
- **Environment data** – Information about the IP address, user agent, SSL status, or time of day
- **Resource data** – Data related to the resource that is being requested

s3:GetObject

S3 Bucket

4. Determining whether a request is **allowed** or **denied**

2. Processing the request context

DigitalCloud TRAINING

© Digital Cloud Training | https://digitalcloud.training

Types of Policy

- **Identity-based policies** – attached to users, groups, or roles
- **Resource-based policies** – attached to a resource; define permissions for a principal accessing the resource
- **IAM permissions boundaries** – set the maximum permissions an identity-based policy can grant an IAM entity
- **AWS Organizations service control policies (SCP)** – specify the maximum permissions for an organization or OU
- **Session policies** – used with AssumeRole* API actions

© Digital Cloud Training | https://digitalcloud.training

Evaluating Policies within an AWS Account

© Digital Cloud Training | https://digitalcloud.training

Determination Rules

1. By default, all requests are implicitly denied (though the root user has full access)
2. An explicit allow in an identity-based or resource-based policy overrides this default
3. If a permissions boundary, Organizations SCP, or session policy is present, it might override the allow with an implicit deny
4. An explicit deny in any policy overrides any allows

© Digital Cloud Training | https://digitalcloud.training

Cross-Account Access

© Digital Cloud Training | https://digitalcloud.training

Cross Account Access (IAM Role)

The trust policy condition requires the external ID

```
"Statement": {
  "Effect": "Allow",
  "Action": "sts:AssumeRole",
  "Principal": {"AWS": "3rd party AWS Account ID"},
  "Condition": {"StringEquals": {"sts:ExternalId": "12345"}}
```

© Digital Cloud Training | https://digitalcloud.training

MFA with the AWS CLI

© Digital Cloud Training | https://digitalcloud.training

Access Evaluation Tools

© Digital Cloud Training | https://digitalcloud.training

AWS Organizations and Control Tower

© Digital Cloud Training | https://digitalcloud.training

AWS Organizations

- 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

© Digital Cloud Training | https://digitalcloud.training

AWS Organizations

© Digital Cloud Training | https://digitalcloud.training

AWS Organizations API

- The AWS Organizations API can be used to automate organization and account creation
- The following API actions are useful to know:
  - CreateOrganization – creates an organization
  - CreateAccount – creates an account that is a member of the organization
  - CreatePolicy – creates a policy that can be attached to a root, OU, or individual AWS account
  - AttachPolicy - Attaches a policy to a root, an organizational unit (OU), or an individual account
  - InviteAccountToOrganization - Sends an invitation to another account to join your organization as a member account

© Digital Cloud Training | https://digitalcloud.training

Service Control Policies

SCPs control the maximum available permissions

AWS Organization

Root Account

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "*",
            "Resource": "*"
        }
    ]
}
```

Tag policy applied to enforce tag standardization

Management Account

Users in the **management** account are not restricted

Test

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "RequireMicroInstanceType",
            "Effect": "Deny",
            "Action": "ec2:RunInstances",
            "Resource": "arn:aws:ec2:*:*:instance/*",
            "Condition": {
                "StringNotEquals": {
                    "ec2:InstanceType": "t2.micro"
                }
            }
        }
}
```

Dev

Dev users can only launch **T2.micro** instances

Prod

Prod users **cannot** launch any instance other than t2.micro (denied above)

NOTE: SCPs do not grant **ANY** permissions, they control the **AVAILABLE** permissions

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

SCP Strategies and Inheritance

Deny List Strategy

- The FullAWSAccess SCP is attached to every OU and account
- Explicitly allows all permissions to flow down from the root
- Can explicitly override with a deny in an SCP
- This is the default setup

Note: An **explicit deny** overrides any kind of **allow**

Allow List Strategy

- The FullAWSAccess SCP is **removed** from every OU and account
- To allow a permission, SCPs with allow statements must be added to the account and every OU above it including root
- Every SCP in the hierarchy must explicitly allow the APIs you want to use

© Digital Cloud Training | https://digitalcloud.training

Service Control Policy Strategies

Deny List Strategy

Allow List Strategy

DynamoDB is **denied** as it's not allowed at every level

© Digital Cloud Training | https://digitalcloud.training

AWS Organizations vs Control Tower

AWS Organizations

- Manage multiple accounts
- Consolidated billing
- Organizational Units
- Service Control Policies
- Tag Policies
- Backup Policies

AWS Control Tower

- Extends capabilities of AWS Organizations
- Landing Zones (best practices)
- Federated Access (IAM Identity Center)
- Centralized Logging
- Account Factory (automation)
- Guardrails (governance rules)

© Digital Cloud Training | https://digitalcloud.training

AWS Control Tower

Directory source can be IAM Identity Center , SAML 2.0 IdP, or Microsoft AD

IAM Identity Center

AWS Control Tower

A **landing zone** is a well-architected multi-account baseline

AWS Organization

Landing Zone

Root

Security OU

Sandbox OU

Production OU

aws Mgmt

aws Audit

aws Dev/Test

aws Prod

aws Log Archive

Detective Guardrails are used for governance and compliance

Preventive Guardrails disallow API actions using **SCPs**

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

AWS Control Tower

- Control Tower creates a well-architected multi-account baseline based on best practices
- This is known as a landing zone
- Guardrails are used for governance and compliance:
  - **Preventive guardrails** are based on SCPs and disallow API actions
  - **Detective guardrails** are implemented using AWS Config rules and Lambda functions and monitor and govern compliance
- The root user in the management account can perform actions that guardrails would disallow

© Digital Cloud Training | https://digitalcloud.training

Organizations and SCPs

© Digital Cloud Training | https://digitalcloud.training

Service Control Policies

AWS Organization

Root Account

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "*",
            "Resource": "*"
        }
    ]
}
```

Management Account

OU1

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "RequireMicroInstanceType",
            "Effect": "Deny",
            "Action": "ec2:RunInstances",
            "Resource": "arn:aws:ec2:*:*:instance/*",
            "Condition": {
                "StringNotEquals": {
                    "ec2:InstanceType": "t2.micro"
                }
            }
        }
    ]
}
```

Dev

Dev users can only launch **t2.micro** instances

© Digital Cloud Training | https://digitalcloud.training

Directory Services and Federation

© Digital Cloud Training | https://digitalcloud.training

AWS Managed Microsoft AD

Managed implementation of **Microsoft Active Directory** running on Windows Server

HA pair of Windows Server **Domain Controllers** (DCs)

One or two-way **trust relationship**

Also Allows you to:

- Apply group policy
- Use single sign-on to apps and services
- Enable MFA with RADIUS

Example apps and services that support **authentication** and **authorization** using AWS Directory Service

Synchronize users and federate identities with **Azure**/**O365**

© Digital Cloud Training | https://digitalcloud.training

DigitalCloud TRAINING

AWS Managed Microsoft AD

- Fully managed AWS service
- Best choice if you have more than 5000 users and/or need a trust relationship set up
- Can perform schema extensions
- Can setup trust relationships to with on-premises Active Directories:
  - On-premise users and groups can access resources in either domain using SSO
  - Requires a VPN or Direct Connect connection
- Can be used as a standalone AD in the AWS cloud

© Digital Cloud Training | https://digitalcloud.training

AD Connector

Sign-in to **AWS applications** such as Amazon WorkSpaces, Amazon WorkDocs, and Amazon WorkMail by using your Active Directory credentials

Provides federated sign-in to the AWS Management Console by mapping Active Directory identities to IAM Roles

© Digital Cloud Training | https://digitalcloud.training

AD Connector

- Redirects directory requests to your on-premise Active Directory
- Best choice when you want to use an existing Active Directory with AWS services
- AD Connector comes in two sizes:
  - Small – designed for organizations up to 500 users
  - Large – designed for organizations up to 5000 users
- Requires a VPN or Direct Connect connection
- Join EC2 instances to your on-premise AD through AD Connector
- Login to the AWS Management Console using your on-premise AD DCs for authentication

© Digital Cloud Training | https://digitalcloud.training

IAM – SAML 2.0 Identity Federation

Active Directory is an LDAP **Identity Store**

Active Directory Federation Services is an **Identity Provider** (IdP)

© Digital Cloud Training | https://digitalcloud.training

IAM – Web Identity Federation

Any **Open ID Connect (OIDC)** compatible IdP supported

Social identity providers (IdPs)

App calls
sts:AssumeRoleWithWebIdentity

Mobile App

AWS Security Token Service (STS) returns **temporary security credentials**

AWS STS

App uses credentials to access DynamoDB

DynamoDB Table

AWS recommend to use **Cognito** for **web identity federation** in most cases

© Digital Cloud Training | https://digitalcloud.training

AWS IAM Identity Center

Identity sources can be Identity Center directory, Active Directory and standard providers using SAML 2.0

AWS IAM Identity Center is the successor to **AWS Single Sign-On** (SSO)

Corporate Office / DC

Active Directory (self-managed)

AD Connector or AWS Managed Microsoft AD

Enables centralized **permissions** management and **SSO**

AWS Identity Center

AWS Organizations

AWS Account A

AWS Account B

AWS Account C

AWS Account D

Azure

Azure AD

Built-in **SSO** integrations to business applications

Custom SAML 2.0 application

Adobe Creative Cloud	AppDynamics	Box
Confluence	Domo	Dropbox
G Suite	Github	GoToMeeting
Jira	NewRelic	Office365
PagerDuty	Salesforce	ServiceNow
Slack	SumoLogic	Tableau
Workplace by Facebook	Zendesk	Zoom

DigitalCloud TRAINING

© Digital Cloud Training | https://digitalcloud.training

Amazon Inspector and AWS Trusted Advisor

© Digital Cloud Training | https://digitalcloud.training

Amazon Inspector

- Runs assessments that check for security exposures and vulnerabilities in EC2 instances
- Can be configured to run on a schedule
- Agent must be installed on EC2 for host assessments
- Network assessments do not require an agent

© Digital Cloud Training | https://digitalcloud.training

Amazon Inspector

Network Assessments

- Assessments: Network configuration analysis to check for ports reachable from outside the VPC
- If the Inspector Agent is installed on your EC2 instances, the assessment also finds processes reachable on port
- Price based on the number of instance assessments

© Digital Cloud Training | https://digitalcloud.training

Amazon Inspector

Host Assessments

- Assessments: Vulnerable software (CVE), host hardening (CIS benchmarks), and security best practices
- Requires an agent (auto-install with SSM Run Command)
- Price based on the number of instance assessments

© Digital Cloud Training | https://digitalcloud.training

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**

© Digital Cloud Training | https://digitalcloud.training

Encryption Primer

© Digital Cloud Training | https://digitalcloud.training

Encryption In Transit vs At Rest

Encryption In Transit

Data is protected by **SSL/TLS** in transit

Encryption At Rest

Amazon S3 **encrypts** the object as it is **written** to the bucket

© Digital Cloud Training | https://digitalcloud.training

Asymmetric Encryption

- Asymmetric encryption is also known as public key cryptography
- Messages encrypted with the public key can only be decrypted with the private key
- Messages encrypted with the private key can be decrypted with the public key
- Examples include SSL/TLS and SSH

© Digital Cloud Training | https://digitalcloud.training

Symmetric Encryption

Encryption

Decryption

© Digital Cloud Training | https://digitalcloud.training

AWS Key Management Service (KMS)

© Digital Cloud Training | https://digitalcloud.training

AWS Key Management Service (KMS)

- Create and managed **symmetric** and **asymmetric** encryption keys
- The **KMS keys** are protected by hardware security modules (HSMs)

Developer creates **customer managed** KMS keys in AWS KMS

© Digital Cloud Training | https://digitalcloud.training

KMS Keys

- KMS keys are the primary resources in AWS KMS
- Used to be known as “customer master keys” or CMKs
- The KMS key also contains the key material used to encrypt and decrypt data
- By default, AWS KMS creates the key material for a KMS key
- You can also import your own key material
- A KMS key can encrypt data up to 4KB in size
- A KMS key can generate, encrypt and decrypt Data Encryption Keys (DEKs)

© Digital Cloud Training | https://digitalcloud.training

Alternative Key Stores

External Key Store

- Keys can be stored outside of AWS to meet regulatory requirements
- You can create a KMS key in an AWS KMS external key store (XKS)
- All keys are generated and stored in an external key manager
- When using an XKS, key material never leaves your HSM

Custom Key Store

- You can create KMS keys in an AWS CloudHSM custom key store
- All keys are generated and stored in an AWS CloudHSM cluster that you own and manage
- Cryptographic operations are performed solely in the AWS CloudHSM cluster you own and manage
- Custom key stores are not available for asymmetric KMS keys

© Digital Cloud Training | https://digitalcloud.training

AWS Managed KMS Keys

- Created, managed, and used on your behalf by an AWS service that is integrated with AWS KMS
- You cannot manage these KMS keys, rotate them, or change their key policies
- You also cannot use AWS managed KMS keys in cryptographic operations directly; the service that creates them uses them on your behalf

Alias	Key ID
aws/sqs	025b9386-b1f8-4fa9-84e2-ac3220b1de59
aws/acm	2d604e85-c2d4-42dc-ab0b-0b356f5fe26e
aws/codecommit	41fea9df-e447-4992-8af7-6ddec6d81175
aws/elasticfilesystem	460c4f05-fe98-4a35-b940-3e1992f04314
aws/glue	617516fe-bf19-4da4-a743-2b13c41973e1
aws/lambda	7f513d01-784b-41b9-9c51-51621db7b5e1
aws/ebs	b9baa4f6-3e87-4256-af6a-d181940df286
aws/lightsail	bc7ba666-8e17-444a-800c-d3d4be303a97
aws/fsx	cebc434c-ee2b-4a61-9b5a-f63be9fdb068
aws/kinesis	d99014b5-09d4-480d-9b2a-3a7d7e3e9c5b

© Digital Cloud Training | https://digitalcloud.training

Data Encryption Keys

- Data keys are encryption keys that you can use to encrypt large amounts of data
- You can use AWS KMS keys to generate, encrypt, and decrypt data keys
- AWS KMS does not store, manage, or track your data keys, or perform cryptographic operations with data keys
- You must use and manage data keys outside of AWS KMS

© Digital Cloud Training | https://digitalcloud.training

KMS Keys and Automatic Rotation

Type of KMS Key	Can view	Can manage	Used only for my AWS account	Automatic rotation
Customer managed key	Yes	Yes	Yes	Optional. Every 365 days
AWS managed key	Yes	No	Yes	Required. Every 365 days
AWS owned key	No	No	No	Varies

- You cannot enable or disable key rotation for AWS owned keys
- Automatic key rotation is supported only on symmetric encryption KMS keys with key material that AWS KMS generates (**Origin = AWS_KMS**)

© Digital Cloud Training | https://digitalcloud.training

KMS Keys and Automatic Rotation

- Automatic rotation generates new key material every year (*optional* for **customer managed keys**)

Rotation only changes the **key material** used for encryption, the KMS key remains the same

© Digital Cloud Training | https://digitalcloud.training

KMS Keys and Automatic Rotation

With automatic key rotation:

- The properties of the KMS key, including its key ID, key ARN, region, policies, and permissions, do not change when the key is rotated
- You do not need to change applications or aliases that refer to the key ID or key ARN of the KMS key
- After you enable key rotation, AWS KMS rotates the KMS key automatically every year

Automatic key rotation is not supported on the following types of KMS keys:

- Asymmetric KMS keys
- HMAC KMS keys
- KMS keys in custom key stores
- KMS keys with imported key material

Note: You can rotate these KMS keys **manually**

© Digital Cloud Training | https://digitalcloud.training

Manual Rotation

- Manual rotation is creating a new KMS key with a different key ID
- You must then update your applications with the new key ID
- You can use an **alias** to represent a KMS key so you don't need to modify your application code

The **alias** is associated with the new **KMS key**

© Digital Cloud Training | https://digitalcloud.training

KMS Key Policies

- Key policies define management and usage permissions for KMS keys

```
{
  "Sid": "Allow access for Key Administrators",
  "Effect": "Allow",
  "Principal": {"AWS": "arn:aws:iam::111122223333:user/KMSKeyAdmin"},
  "Action": [
    "kms:Describe*",
    "kms:Put*",
    "kms:Create*",
    "kms:Update*",
    "kms:Enable*",
    "kms:Revoke*",
    "kms:List*",
    "kms:Disable*",
    "kms:Get*",
    "kms:Delete*",
    "kms:ScheduleKeyDeletion",
    "kms:CancelKeyDeletion"
  ],
  "Resource": "*"
}
```

This key policy defines the **administrative actions** that are permitted for a key administrator

© Digital Cloud Training | https://digitalcloud.training

KMS Key Policies

- Multiple policy statements can be combined to specify separate administrative and usage permissions

```
{
    "Sid": "Allow use of the key",
    "Effect": "Allow",
    "Principal": {"AWS": "arn:aws:iam::111122223333:role/EncryptionApp"},
    "Action": [
      "kms:DescribeKey",
      "kms:GenerateDataKey*",
      "kms:Encrypt",
      "kms:ReEncrypt*",
      "kms:Decrypt"
    ],
    "Resource": "*"
  }
```

This key policy defines the **cryptographic** actions for encrypting and decrypting data with the KMS key

© Digital Cloud Training | https://digitalcloud.training

KMS Key Policies

- Permissions can be specified for delegating use of the key to AWS services

```
{
  "Sid": "Allow attachment of persistent resources",
  "Effect": "Allow",
  "Principal": {"AWS": "arn:aws:iam::111122223333:role/EncryptionApp"},
  "Action": [
    "kms:ListGrants",
    "kms:CreateGrant",
    "kms:RevokeGrant"
  ],
  "Resource": "*",
  "Condition": {"Bool": {"kms:GrantIsForAWSResource": true}}
}
```

Grants are useful for **temporary permissions** as they can be used without modifying key policies or IAM policies

© Digital Cloud Training | https://digitalcloud.training

Additional Exam Tips

- To share snapshots with another account you must specify Decrypt and CreateGrant permissions
- The kms:ViaService condition key can be used to limit key usage to specific AWS services
- For example:

```
"Condition": {
  "StringEquals": {
    "kms:ViaService": [
      "ec2.us-west-2.amazonaws.com",
      "rds.us-west-2.amazonaws.com"
```

© Digital Cloud Training | https://digitalcloud.training

Additional Exam Tips

- Cryptographic erasure means removing the ability to decrypt data and can be achieved when using **imported key material** and deleting that key material
- You must use the DeleteImportedKeyMaterial API to remove the key material
- An InvalidKeyId exception when using SSM Parameter Store indicates the KMS key is not enabled
- Make sure you know the differences between AWS managed and customer managed KMS keys and automatic vs manual rotation

© Digital Cloud Training | https://digitalcloud.training

Create KMS Key

© Digital Cloud Training | https://digitalcloud.training

AWS KMS
API and CLI

© Digital Cloud Training | https://digitalcloud.training

AWS KMS API and CLI

Encrypt (aws kms encrypt):

- Encrypts plaintext into ciphertext by using a KMS key
- You can encrypt small amounts of arbitrary data, such as a personal identifier or database password, or other sensitive information
- You can use the Encrypt operation to move encrypted data from one AWS region to another

Decrypt (aws kms decrypt):

- Decrypts ciphertext that was encrypted by an KMS key using any of the following operations:
  - Encrypt
  - GenerateDataKey
  - GenerateDataKeyPair
  - GenerateDataKeyWithoutPlaintext
  - GenerateDataKeyPairWithoutPlaintext

© Digital Cloud Training | https://digitalcloud.training

AWS KMS API and CLI

Re-encrypt (aws kms re-encrypt):

- Decrypts ciphertext and then re-encrypts it entirely within AWS KMS
- You can use this operation to change the KMS key under which data is encrypted, such as when you manually rotate a KMS key or change the KMS key that protects a ciphertext
- You can also use it to re-encrypt ciphertext under the same KMS key, such as to change the encryption context of a ciphertext

Enable-key-rotation:

- Enables automatic rotation of the key material for the specified symmetric KMS key
- You cannot perform this operation on a KMS key in a different AWS account

© Digital Cloud Training | https://digitalcloud.training

AWS KMS API and CLI

GenerateDataKey (aws kms generate-data-key):

- Generates a unique symmetric data key
- This operation returns a plaintext copy of the data key and a copy that is encrypted under a KMS key that you specify
- You can use the plaintext key to encrypt your data outside of AWS KMS and store the encrypted data key with the encrypted data

GenerateDataKeyWithoutPlaintext (generate-data-key-without-plaintext):

- Generates a unique symmetric data key
- This operation returns a data key that is encrypted under a KMS key that you specify
- To request an asymmetric data key pair, use the GenerateDataKeyPair or GenerateDataKeyPairWithoutPlaintext operations

DigitalCloud TRAINING

Throttling and Caching

- AWS KMS has two types of quotas:
  - Resource quotas
  - Request quotas
- If you exceed a resource limit, requests to create an additional resource of that type generate an LimitExceededException error message
- Request quotas apply to API actions such as Encrypt, Decrypt, ReEncrypt, and GenerateDataKey
- If you exceed a request quota you may receive a throttling error:

```
You have exceeded the rate at which you may call KMS. Reduce the frequency of your calls.
(Service: AWSKMS; Status Code: 400; Error Code: ThrottlingException; Request ID: <ID>
```

© Digital Cloud Training | https://digitalcloud.training

Throttling and Caching

- To prevent throttling, you can:
  - Implement a backoff and retry strategy
  - Request a service quota increase
  - Implement data key caching
- Data key caching stores data keys and related cryptographic material in a cache
- Useful if your application:
  - Can reuse data keys
  - Generates numerous data keys
  - Runs cryptographic operations that are unacceptably slow, expensive, limited, or resource-intensive
- You can create a local cache using the AWS Encryption SDK and the LocalCryptoMaterialsCache feature

© Digital Cloud Training | https://digitalcloud.training

AWS CloudHSM

© Digital Cloud Training | https://digitalcloud.training

AWS CloudHSM

- AWS CloudHSM is a cloud-based hardware security module (HSM)
- Generate and use your own encryption keys on the AWS Cloud
- CloudHSM runs in your Amazon VPC
- Uses FIPS 140-2 level 3 validated HSMs
- Managed service and automatically scales
- Retain control of your encryption keys - you control access (and AWS has no visibility of your encryption keys)

© Digital Cloud Training | https://digitalcloud.training

AWS CloudHSM Use Cases

- Offload SSL/TLS processing from web servers
- Protect private keys for an issuing certificate authority (CA)
- Store the master key for Oracle DB Transparent Data Encryption
- Custom key store for AWS KMS – retain control of the HSM that protects the master keys

© Digital Cloud Training | https://digitalcloud.training

AWS CloudHSM vs KMS

	CloudHSM	AWS KMS
Tenancy	Single-tenant HSM	Multi-tenant AWS service
Availability	Customer-managed durability and available	Highly available and durable key storage and management
Root of Trust	Customer managed root of trust	AWS managed root of trust
FIPS 140-2	Level 3	Level 2 / Level 3
3rd Party Support	Broad 3rd Party Support	Broad AWS service support

© Digital Cloud Training | https://digitalcloud.training

AWS Certificate Manager (ACM)

© Digital Cloud Training | https://digitalcloud.training

AWS Certificate Manager (ACM)

- Create, store and renew SSL/TLS X.509 certificates
- Single domains, multiple domain names and wildcards
- Integrates with several AWS services including:
  - **Elastic Load Balancing**
  - **Amazon CloudFront**
  - **AWS Elastic Beanstalk**
  - **AWS Nitro Enclaves**
  - **AWS CloudFormation**

© Digital Cloud Training | https://digitalcloud.training

AWS Certificate Manager (ACM)

- **Public certificates** are signed by the AWS public Certificate Authority
- You can also create a Private CA with ACM
- Can then issue private certificates
- You can also import certificates from third-party issuers

© Digital Cloud Training | https://digitalcloud.training

Managing ACM Certificates

© Digital Cloud Training | https://digitalcloud.training

Secrets Management

© Digital Cloud Training | https://digitalcloud.training

AWS Secrets Manager

- Stores and rotate secrets safely without the need for code deployments
- Secrets Manager offers automatic rotation of credentials (built-in) for:
  - Amazon RDS (MySQL, PostgreSQL, and Amazon Aurora)
  - Amazon Redshift
  - Amazon DocumentDB
- For other services you can write your own AWS Lambda function for automatic rotation

© Digital Cloud Training | https://digitalcloud.training

AWS SSM Parameter Store

- Parameter Store provides secure, hierarchical storage for configuration data and secrets
- Highly scalable, available, and durable
- Store data such as passwords, database strings, and license codes as parameter values
- Store values as plaintext (unencrypted data) or ciphertext (encrypted data)
- Reference values by using the unique name that you specified when you created the parameter
- No native rotation of keys (difference with AWS Secrets Manager which does it automatically)

© Digital Cloud Training | https://digitalcloud.training

AWS Secrets Manager vs SSM Parameter Store

	Secrets Manager	SSM Parameter Store
Automatic Key Rotation	Yes, built-in for some services, use Lambda for others	No native key rotation; can use custom Lambda
Key/Value Type	String or Binary (encrypted)	String, StringList, SecureString (encrypted)
Hierarchical Keys	No	Yes
Price	Charges apply per secret	Free for standard, charges for advanced

© Digital Cloud Training | https://digitalcloud.training

Secrets Manager with RDS and Lambda

© Digital Cloud Training | https://digitalcloud.training

AWS Security Hub and AWS GuardDuty

© Digital Cloud Training | https://digitalcloud.training

AWS Security Hub

- Provides cloud security posture management
- Run security best practice configuration checks such as
  - **AWS Foundational Security Best Practices**
  - **Center for Internet Security (CIS)**
  - **Payment Card Industry (PCI DSS)**
- Aggregates alerts and provides automated remediation
- Integrations include:
  - **Amazon EventBridge – send alerts and trigger targets**
  - **Systems Manager – automation via runbooks**
  - **AWS Organizations – consolidated findings across accounts**
  - **Ticketing, chat, incident management, investigation, GRC, SOAR, and SIEM tools**

DigitalCloud TRAINING

AWS GuardDuty

- Intelligent threat detection service
- Detects account compromise, instance compromise, malicious reconnaissance, and bucket compromise
- Continuous monitoring for events across:
  - **AWS CloudTrail Management Events**
  - **AWS CloudTrail S3 Data Events**
  - **Amazon VPC Flow Logs**
  - **DNS Logs**

© Digital Cloud Training | https://digitalcloud.training

SECTION 8

Networking and Content Delivery

© Digital Cloud Training | https://digitalcloud.training

Private Subnets and NAT Gateways

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 in Public and Private Subnets

Public Subnet Route Table	
Destination	**Target**
10.0.0.0/16	Local
0.0.0.0/0	igw-id

Public subnets require:

- A route to an Internet gateway
- Auto-assign public IPv4 or IPv6 address

Private subnets:

- Do not auto assign public addresses
- Do not have a route to an internet gateway

© Digital Cloud Training | https://digitalcloud.training

NAT Gateways

Private Subnet Route Table	
Destination	**Target**
10.0.0.0/16	Local
0.0.0.0/0	nat-gateway-id

NAT Gateways deployment requirements:

- Deployed in public subnets
- Have an Elastic IP attached
- Route to the NAT gateway added to private subnet

© Digital Cloud Training | https://digitalcloud.training

High Availability for NAT Gateway

NAT gateways have redundancy **within an AZ**

AZ1 Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	nat-gw-az1

AZ2 Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	nat-gw-az2

AZ3 Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	nat-gw-az3

© Digital Cloud Training | https://digitalcloud.training

Launch NAT Gateway

© Digital Cloud Training | https://digitalcloud.training

VPC Connectivity Options

© Digital Cloud Training | https://digitalcloud.training

Hybrid Connectivity Architectures

AWS Direct Connect (DX)

© Digital Cloud Training | https://digitalcloud.training

Direct Connect - Multiple Regions

© Digital Cloud Training | https://digitalcloud.training

AWS Transit Gateway

© Digital Cloud Training | https://digitalcloud.training

AWS TGW + DX Gateway

© Digital Cloud Training | https://digitalcloud.training

Link: VPC connectivity options whitepaper

© Digital Cloud Training | https://digitalcloud.training

VPC to VPC Connectivity Options

- VPC Peering
- AWS Transit Gateway
- Software S2S VPN
- Software VPN to AWS VPN
- AWS Managed VPN
- AWS PrivateLink

Link: VPC connectivity options whitepaper

© Digital Cloud Training | https://digitalcloud.training

VPC Peering Limitations

Many peering connections required for fully meshed routing

Complex routing configuration required

Does not support transitive peering relationships

Cannot have matching or overlapping CIDR Blocks

© Digital Cloud Training | https://digitalcloud.training

Configure VPC Peering

© Digital Cloud Training | https://digitalcloud.training

VPC Peering

Security group

Protocol	Port	Source
ICMP	All	10.1.0.0/16
TCP	22	0.0.0.0/0

Security group

Protocol	Port	Source
ICMP	All	10.0.0.0/16
TCP	22	0.0.0.0/0

Route Table

Destination	Target
10.1.0.0/16	peering-id

Route Table

Destination	Target
10.0.0.0/16	peering-id

© Digital Cloud Training | https://digitalcloud.training

VPC Endpoints

© Digital Cloud Training | https://digitalcloud.training

VPC Interface Endpoints

© Digital Cloud Training | https://digitalcloud.training

VPC Gateway Endpoints

Route Table

Destination	Target
pl-6ca54005 (com.amazonaws.ap-southeast-2.s3, 54.231.248.0/22, 54.231.252.0/24, 52.95.128.0/21)	*vpce-ID*

A **route table** entry is required with the prefix list for S3 and the **gateway ID**

© Digital Cloud Training | https://digitalcloud.training

VPC Endpoints

	Interface Endpoint	Gateway Endpoint
What	Elastic Network Interface with a Private IP	A gateway that is a target for a specific route
How	Uses DNS entries to redirect traffic	Uses prefix lists in the route table to redirect traffic
Which services	API Gateway, CloudFormation, CloudWatch etc.	Amazon S3, DynamoDB
Security	Security Groups	VPC Endpoint Policies

© Digital Cloud Training | https://digitalcloud.training

Service Provider Model

© Digital Cloud Training | https://digitalcloud.training

Some AWS Services that Support PrivateLink

- Amazon API Gateway
- Amazon Athena
- AWS Key Management Service
- AWS CloudHSM
- Amazon CloudWatch
- Amazon CloudWatch Events
- Amazon CloudWatch Logs
- AWS Config
- Amazon EventBridge
- AWS Glue
- AWS Lambda
- Amazon S3
- AWS Secrets Manager
- Amazon SNS
- Amazon SQS
- AWS Step Functions
- AWS Systems Manager
- Full list here

© Digital Cloud Training | https://digitalcloud.training

S3 Gateway / Interface Endpoints

© Digital Cloud Training | https://digitalcloud.training

Interface and Gateway Endpoints for S3

Interface endpoint **not required** if allowing outbound access over port 443 to:

- ssm.region.amazonaws.com
- ssmmessages.region.amazonaws.com
- ec2messages.region.amazonaws.com

Route Table

Destination	Target
pl-6ca54005	*vpce-ID*

© Digital Cloud Training | https://digitalcloud.training

Interface and Gateway Endpoints for S3

Gateway endpoints for Amazon S3	Interface endpoints for Amazon S3
Use Amazon S3 public IP addresses	Use private IP addresses from your VPC to access Amazon S3
Use the same Amazon S3 DNS names	Require endpoint-specific Amazon S3 DNS names
Does not allow access from on premises	Allow access from on premises
Does not allow access from another AWS Region	Allow access from a VPC in another AWS Region using VPC peering or AWS Transit Gateway
Not billed	Billed

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53 Resolver

© Digital Cloud Training | https://digitalcloud.training

Route 53 Resolver – Outbound Endpoints

© Digital Cloud Training | https://digitalcloud.training

Route 53 Resolver – Inbound Endpoints

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudFront Origins and Distributions

© Digital Cloud Training | https://digitalcloud.training

Amazon CloudFront

© Digital Cloud Training | https://digitalcloud.training

CloudFront Origins and Distributions

CloudFront Distribution

Name: **d1schtd9zdwrm1.cloudfront.net**

S3 Origin

S3 static websites can also be origins

Behaviors

Path Pattern
Viewer Protocol Policy
Cache Policy
Origin Request Policy

Custom Origin

RTMP distributions were discontinued so only **web distributions** are currently available

CloudFront Web Distribution:

- Speed up distribution of static and dynamic content, for example, .html, .css, .php, and graphics files
- Distribute media files using HTTP or HTTPS
- Add, update, or delete objects, and submit data from web forms
- Use live streaming to stream an event in real time

© Digital Cloud Training | https://digitalcloud.training

CloudFront Signed URLs and OAI/OAC

© Digital Cloud Training | https://digitalcloud.training

CloudFront Signed URLs

- Signed URLs provide more control over access to content.
- Can specify beginning and expiration date and time, IP addresses/ranges of users.

Signed URLs should be used for **individual files** and clients that don't support **cookies**

Mobile app uses **signed URL** to access distribution

3

Amazon CloudFront

4

Signed URL returned

2

1

Serverless Application

Mobile app **authenticates** to application and requests signed URL

© Digital Cloud Training | https://digitalcloud.training

CloudFront Signed Cookies

- Similar to Signed URLs
- Use signed cookies when you don’t want to change URLs
- Can also be used when you want to provide access to **multiple restricted files** (Signed URLs are for individual files)

© Digital Cloud Training | https://digitalcloud.training

CloudFront Origin Access Identity (OAI)

```
{
    "Version": "2008-10-17",
    "Id": "PolicyForCloudFrontPrivateContent",
    "Statement": [
        {
            "Sid": "1",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::cloudfront:user/CloudFront Origin Access Identity E11A2JL2H306JJ"
            },
            "Action": "s3:GetObject",
            "Resource": "arn:aws:s3:::mybucket/*"
        }
    ]
}
```

Policy restricts access to the **OAI**

HTTP GET **https://d1schtd9zdwrm1.cloudfront.net**

Custom Origin

S3 Bucket configured as static website

Bucket Policy

Origin Access Identity (OAI)

Amazon CloudFront

Users

Blocked by **bucket policy**

GET **https://mybucket.s3.amazonaws.com/beach.jpg**

© Digital Cloud Training | https://digitalcloud.training

CloudFront Origin Access Control (OAC)

- Like a Signed OAI but supports additional use cases
- AWS recommend the OAC for the following use cases:
  - All Amazon S3 buckets in all AWS Regions
  - Amazon S3 server-side encryption with AWS KMS (SSE-KMS)
  - Dynamic requests (PUT and DELETE) to Amazon S3
- Requires an S3 bucket policy that allows the CloudFront service principal

```
{
    "Version": "2012-10-17",
    "Statement": {
        "Sid": "AllowCloudFrontServicePrincipalReadOnly",
        "Effect": "Allow",
        "Principal": {
            "Service": "cloudfront.amazonaws.com"
        },
        "Action": "s3:GetObject",
        "Resource": "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*",
        "Condition": {
            "StringEquals": {
                "AWS:SourceArn": "arn:aws:cloudfront::111122223333:distribution/EDFDVBD6EXAMPLE"
            }
        }
    }
}
```

© Digital Cloud Training | https://digitalcloud.training

S3 Static Website Hosting

© Digital Cloud Training | https://digitalcloud.training

Configure Protected Origin

© Digital Cloud Training | https://digitalcloud.training

Collect and Interpret Logs

© Digital Cloud Training | https://digitalcloud.training

S3 Server Access Logging

- Provides detailed records for the requests that are made to a bucket
- Details include the requester, bucket name, request time, request action, response status, and error code (if applicable)
- Disabled by default
- Only pay for the storage space used
- Must configure a separate bucket as the destination (can specify a prefix)
- Must grant write permissions to the Amazon S3 Log Delivery group on destination bucket

VPC Flow Logs

- 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 or S3
- Flow logs can be created at the following levels:
  - VPC
  - Subnet
  - Network interface

© Digital Cloud Training | https://digitalcloud.training

ELB Access Logs

- Disabled by default
- Includes information about the clients (not included in CloudWatch metrics):
  - Time
  - Client IP address
  - Latencies
  - Request paths
  - Server response
  - Trace ID
- Can be optionally stored and retained in S3

© Digital Cloud Training | https://digitalcloud.training

CloudFront Access Logs

- Contain detailed information about every user request that CloudFront receives at Edge Locations
- Known as *standard logs* and stored in S3
- Can log separately for different distributions
- Can also enable real-time logs which are recorded in real time (within seconds)
- Real-time logs can be used to monitor, analyze, and take action based on content delivery performance
- Edge function logs record requests processed by Lambda@Edge and CloudFront Functions

© Digital Cloud Training | https://digitalcloud.training

VPC Flow Logs vs ELB Access Logs

VPC Flow Log

version	account-id	interface-id	srcaddr	dstaddr	srcport	dstport	protocol	packets	bytes	start	end	action	log-status
2	55112233445	eni-0f5...	11.200.185.200	10.0.1.15	52933	22	6	1	40	1599...	1599...	ACCEPT	OK
2	55112233445	eni-0f5...	10.0.1.15	11.200.185.200	22	52933	6	1	40	1599...	1599...	ACCEPT	OK
2	55112233445	eni-0f5...	11.200.185.200	10.0.1.15	3624	80	6	1	44	1599...	1599...	REJECT	OK
2	55112233445	eni-0f5...	11.200.185.200	10.0.1.15	3624	80	6	1	44	1599...	1599...	REJECT	OK

ELB Access Log

```
http 2018-07-02T22:23:00.186641Z app/my-loadbalancer/50dc6c495c0c9188
192.168.131.39:2817 10.0.0.1:80 0.000 0.001 0.000 200 200 34 366
"GET http://www.example.com:80/ HTTP/1.1" "curl/7.46.0" - -
arn:aws:elasticloadbalancing:us-east-2:123456789012:targetgroup/my-targets/73e2d6bc24d8a067
"Root=1-58337262-36d228ad5d99923122bbe354" "-" "-"
0 2018-07-02T22:22:48.364000Z "forward" "-" "-" 10.0.0.1:80 200 "-" "-"
```

© Digital Cloud Training | https://digitalcloud.training

AWS Web Application Firewall (WAF)

© Digital Cloud Training | https://digitalcloud.training

AWS WAF

- AWS WAF is a web application firewall
- WAF lets you create rules to filter web traffic based on conditions that include IP addresses, HTTP headers and body, or custom URIs
- WAF makes it easy to create rules that block common web exploits like SQL injection and cross site scripting

© Digital Cloud Training | https://digitalcloud.training

AWS WAF

© Digital Cloud Training | https://digitalcloud.training

AWS WAF

- **Web ACLs** – You use a web access control list (ACL) to protect a set of AWS resources
- **Rules** – Each rule contains a statement that defines the inspection criteria, and an action to take if a web request meets the criteria
- **Rule groups** – You can use rules individually or in reusable rule groups

© Digital Cloud Training | https://digitalcloud.training

AWS WAF

- **IP Sets** - An IP set provides a collection of IP addresses and IP address ranges that you want to use together in a rule statement
- **Regex pattern set** - A regex pattern set provides a collection of regular expressions that you want to use together in a rule statement

© Digital Cloud Training | https://digitalcloud.training

AWS WAF

A **rule action** tells AWS WAF what to do with a web request when it ***matches*** the criteria defined in the rule:

- **Count** – AWS WAF counts the request but doesn't determine whether to allow it or block it. With this action, AWS WAF continues processing the remaining rules in the web ACL
- **Allow** – AWS WAF allows the request to be forwarded to the AWS resource for processing and response
- **Block** – AWS WAF blocks the request and the AWS resource responds with an HTTP 403 (Forbidden) status code

© Digital Cloud Training | https://digitalcloud.training

AWS WAF

Match statements compare the web request or its origin against conditions that you provide

Match Statement	Description
Geographic match	Inspects the request's country of origin
IP set match	Inspects the request against a set of IP addresses and address ranges
Regex pattern set	Compares regex patterns against a specified request component
Size constraint	Checks size constraints against a specified request component
SQLi attack	Inspects for malicious SQL code in a specified request component
String match	Compares a string to a specified request component
XSS scripting attack	Inspects for cross-site scripting attacks in a specified request component

DigitalCloud TRAINING

AWS Shield

© Digital Cloud Training | https://digitalcloud.training

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** – no cost
  - **Advanced** - $3k USD per month and 1 year commitment
- Integrated with Amazon CloudFront (standard included by default)

© Digital Cloud Training | https://digitalcloud.training

WAF Web ACL For CloudFront

© Digital Cloud Training | https://digitalcloud.training

SECTION 9

Cost and Performance Optimization

© Digital Cloud Training | https://digitalcloud.training

AWS Compute Optimizer

© Digital Cloud Training | https://digitalcloud.training

AWS Compute Optimizer

- Recommends optimal AWS resources for your workloads to reduce costs and improve performance
- Uses machine learning to analyze historical utilization metrics
- Offers optimization guidance for:
  - Amazon EC2 instances
  - Amazon EBS volumes
  - AWS Lambda functions
- Results can be viewed in the console or via the CLI

© Digital Cloud Training | https://digitalcloud.training

AWS Compute Optimizer

AWS Compute Optimizer > Dashboard

Dashboard Info

Findings per AWS resource

090765505187

Filter by one or more Regions

Region: US East (N. Virginia) | Clear filters

EC2 instances (13) Info

30.77% Under-provisi... | 7.... | 61.54% Over-provisioned

Findings

- Under-provisioned: 4 instances
- Optimized: 1 instance
- Over-provisioned: 8 instances

View recommendations for EC2 instances

Auto Scaling groups (1) Info

100% Not optimized

Findings

- Not optimized: 1 group
- Optimized: 0 groups

View recommendations for Auto Scaling groups

© Digital Cloud Training | https://digitalcloud.training

AWS Compute Optimizer

AWS Compute Optimizer > Dashboard > Recommendations for EC2 instances

Recommendations for EC2 instances (8) Info

Recommendations for modifying current resources for better cost and performance.

Action ▼ | View detail

Filter by one or more Regions | 090765505187 ▼ | Over-provisioned ▼ | 1

Region: US East (N. Virginia) ✕ | Clear filters

Instance ID	Instance name	Finding	Current instance type	Current On-Demand price	Recommended instance type	Recommended On-Demand price
i-0fb9323080785de1e	-	Over-provisioned	c5.xlarge	$0.17 per hour	t3.large	$0.0832 per hour
i-0f4f4c06ad8afe81a	-	Over-provisioned	m5.2xlarge	$0.384 per hour	r5.xlarge	$0.252 per hour
i-0f277818dfef522e9	-	Over-provisioned	c5.xlarge	$0.17 per hour	t3.large	$0.0832 per hour
i-0ceb95ed248026d24	-	Over-provisioned	m5.xlarge	$0.192 per hour	r5.large	$0.126 per hour
i-0af9322ff627d7e8f	-	Over-provisioned	m5.xlarge	$0.192 per hour	r5.large	$0.126 per hour
i-07084b94d1bcf391b	-	Over-provisioned	c5.xlarge	$0.17 per hour	t3.large	$0.0832 per hour
i-069f6e837890db127	-	Over-provisioned	c5.xlarge	$0.17 per hour	t3.large	$0.0832 per hour
i-0218a45abd8b53658	-	Over-provisioned	m5.xlarge	$0.192 per hour	r5.large	$0.126 per hour

DigitalCloud TRAINING

AWS Cost Allocation Tags

© Digital Cloud Training | https://digitalcloud.training

AWS Cost Management Tools

© Digital Cloud Training | https://digitalcloud.training

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

© Digital Cloud Training | https://digitalcloud.training

AWS Cost & Usage Report

- Publish AWS billing reports to an Amazon S3 bucket
- Reports break down costs by:
  - Hour, day, month, product, product resource, tags
- Can update the report up to three times a day
- Create, retrieve, and delete your reports using the AWS CUR API Reference

© Digital Cloud Training | https://digitalcloud.training

AWS Price List API

- Query the prices of AWS services
- **Price List Service API** (AKA the Query API) – query with JSON
- **AWS Price List API** (AKA the Bulk API) – query with HTML
- Alerts via Amazon SNS when prices change

© Digital Cloud Training | https://digitalcloud.training

AWS Cost Management Tools

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Pricing Options

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Pricing Options

On-Demand

Standard rate - no discount; no commitments; dev/test, short-term, or unpredictable workloads

Reserved

1 or 3-year commitment; up to 75% discount; steady-state, predictable workloads and reserved capacity

Spot Instances

Get discounts of up to 90% for unused capacity. Can be terminated at any time

Dedicated Instances

Physical isolation at the host hardware level from instances belonging to other customers; pay per instance

Dedicated Hosts

Physical server dedicated for your use; Socket/core visibility, host affinity; pay per host; workloads with server-bound software licenses

Savings Plans

Commitment to a consistent amount of usage (EC2 + Fargate + Lambda); Pay by $/hour; 1 or 3-year commitment

DigitalCloud TRAINING

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Billing

Commercial Linux distros such as **Red Hat EL** and **SUSE ES** use **hourly** pricing

Billed per hour;
Minimum of 1 hour

Billed per second;
Minimum of 1 minute

Volumes billed per second;
Minimum of 1 minute

Per-second billing is for **Amazon Linux, Windows** and **Ubuntu** in **On-Demand**, **Reserved**, and **Spot** forms

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Reserved Instances (RIs)

Standard RI

Change **AZ**,
instance size (Linux),
networking type

Use **ModifyReservedInstances** API

Term is 1 or 3 years

Can pay **All Upfront**, **Partial Upfront**, **No Upfront**

Convertible RI

Change **AZ**,
instance size (Linux),
networking type

+

Change **family**, **OS**,
tenancy, **payment option**

Use **ExchangeReservedInstances** API

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Reserved Instances (RIs)

Tenancy: **Default** or **Dedicated**

When the **attributes** of a used instance match the **attributes** of an RI the discount is applied

Reserves capacity in specified AZ

Does **not** reserve capacity; discount applies to all AZs

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 On-Demand Capacity Reservations

- Reserve compute capacity for your Amazon EC2 instances in a specific Availability Zone
- Any duration can be specified
- Mitigates against the risk of being unable to get On-Demand capacity
- Does not require any term commitments and can be cancelled at any time
- When you create a Capacity Reservation, you specify:
  - The **Availability Zone** in which to reserve the capacity
  - The **number of instances** for which to reserve capacity
  - The **instance attributes**, including the instance type, tenancy, and platform/OS

© Digital Cloud Training | https://digitalcloud.training

AWS Savings Plans

Compute Savings Plan

1 or 3-year; hourly commitment to usage of **Fargate**, **Lambda**, and **EC2**; Any Region, family, size, tenancy, and OS

EC2 Savings Plan

1 or 3-year; hourly commitment to usage of **EC2** within a **selected Region** and **Instance Family**; Any size, tenancy and OS

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 Spot Instances

Spot Instance: One or more EC2 instances

Spot Fleet: launches and maintains the number of Spot / On-Demand instances to meet specified target capacity

EC2 Fleet: launches and maintains specified number of Spot / On-Demand / Reserved instances in a **single API call**

2-minute warning if AWS need to reclaim capacity – available via **instance metadata** and **CloudWatch Events**

Can define separate OD/Spot **capacity targets**, **Spot price**, **instance types**, and **AZs**

© Digital Cloud Training | https://digitalcloud.training

Spot Block

Requirement:
Uninterrupted for
1-6 hours

Solution: **Spot Block**

```
$ aws ec2 request-spot-instances \
  --block-duration-minutes 360 \
  --instance-count 5 \
  --spot-price "0.25" ...
```

Pricing is **30% - 45%** less than On-Demand

© Digital Cloud Training | https://digitalcloud.training

Dedicated Instances and Dedicated Hosts

Characteristic	Dedicated Instances	Dedicated Hosts
Enables the use of dedicated physical servers	X	X
Per instance billing (subject to a $2 per region fee)	X	
Per host billing		X
Visibility of sockets, cores, host ID		X
Affinity between a host and instance		X
Targeted instance placement		X
Automatic instance placement	X	X
Add capacity using an allocation request		X

© Digital Cloud Training | https://digitalcloud.training

EC2 Placement Groups

© Digital Cloud Training | https://digitalcloud.training

EC2 Placement Groups

Cluster Placement Group

- Packs instances close together inside an Availability Zone
- Enables low-latency network performance
- Useful for tightly-coupled node-to-node communications
- Higher Performance Computing (HPC) is a common use case

Partition Placement Group

- Spreads instances across logical partitions to avoid sharing underlying hardware with instances in different partitions
- Typically used by large distributed and replicated workloads such as Hadoop, Cassandra, and Kafka

Spread Placement Group

- Instances are placed across distinct underlying hardware to reduce correlated failures

DigitalCloud TRAINING

Cluster Placement Group

Uses **enhanced networking**, low network latency and high throughput for **inter-instance** traffic

© Digital Cloud Training | https://digitalcloud.training

Partition Placement Group

Each partition is located on a **separate AWS rack**

© Digital Cloud Training | https://digitalcloud.training

Spread Placement Group

Each instance is located on a **separate AWS rack**

© Digital Cloud Training | https://digitalcloud.training

EC2 Placement Group Use Cases

Cluster	Tightly-coupled application that requires low-latency, high throughput network traffic between instances
Partition	Distributed and replicated NoSQL database; requires separate hardware for node groups
Spread	Small number of critical instances that should be kept separate from each other

© Digital Cloud Training | https://digitalcloud.training

EC2 Placement Groups

Rules and Limitations

- An instance can be launched in one placement group at a time; it cannot span multiple placement groups
- You can't merge placement groups
- A cluster placement group can't span multiple Availability Zones
- A partition placement group supports a maximum of seven partitions per Availability Zone
- A rack spread placement group supports a maximum of seven running instances per Availability Zone

https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html

© Digital Cloud Training | https://digitalcloud.training

Multipart Upload & Transfer Acceleration

© Digital Cloud Training | https://digitalcloud.training

S3 Multipart Upload

- Multipart upload uploads objects in parts independently, in parallel and in any order
- Performed using the S3 Multipart upload API
- It is recommended for objects of 100 MB or larger
- Can be used for objects from 5 MB up to 5 TB
- Must be used for objects larger than 5 GB

© Digital Cloud Training | https://digitalcloud.training

S3 Multipart Upload with the AWS CLI

You can use the AWS CLI to perform a multipart upload to S3:

```
aws s3api create-multipart-upload --bucket my-bucket --key large-file.txt --content-type text/plain
```

Then, upload each part of the file using the following command:

```
aws s3api upload-part --bucket my-bucket --key large-file.txt --part-number 1 --body /path/to/part1 --upload-id
<upload-id>
```

Upload additional parts using the same command with different part numbers, and then complete the multipart upload using the following command:

```
aws s3api complete-multipart-upload --bucket my-bucket --key large-file.txt --upload-id <upload-id>
--multipart-upload '{"Parts":[{"ETag":"<etag1>","PartNumber":1},{"ETag":"<etag2>","PartNumber":2},...]}'
```

Use the list-parts command to list the parts uploaded for a specific multipart upload:

```
aws s3api complete-multipart-upload --bucket my-bucket --key large-file.txt --upload-id <upload-id>
--multipart-upload '{"Parts":[{"ETag":"<etag1>","PartNumber":1},{"ETag":"<etag2>","PartNumber":2},...]}'
```

DigitalCloud TRAINING

S3 Multipart Upload with Boto3

```python
# Start multipart upload
response = s3_client.create_multipart_upload(Bucket=bucket_name, Key=object_key)
upload_id = response['UploadId']

# Upload parts
with open(file_path, 'rb') as file:
    file_size = os.fstat(file.fileno()).st_size
    part_count = math.ceil(file_size / part_size)

    for part_number in range(part_count):
        start = part_size * part_number
        end = min(start + part_size, file_size)
        part_data = file.read(end - start)

        response = s3_client.upload_part(
            Body=part_data,
            Bucket=bucket_name,
            Key=object_key,
            PartNumber=part_number + 1,
            UploadId=upload_id
        )

        print(f'Uploaded part {part_number + 1}: {response}')
```

This example uses the boto3 library for Python

© Digital Cloud Training | https://digitalcloud.training

S3 Transfer Acceleration (S3TA)

- Uses CloudFront edge locations to improve performance of transfers from client to S3 bucket

```
http://bucketname.s3-accelerate.amazonaws.com

http://bucketname.s3-accelerate.dualstack.amazonaws.com
```

Transfer Acceleration is enabled at the bucket level

AWS only charges if there's a performance improvement

S3 Transfer Acceleration (S3TA)

- Speeds transfers by 50-500%
- Moves data faster over long distances
- Must be enabled in the bucket properties

© Digital Cloud Training | https://digitalcloud.training

S3 Transfer Acceleration (S3TA)

- Use the speed comparison tool to check if it's worth the cost:

© Digital Cloud Training | https://digitalcloud.training

SECTION 10

Exam Scenarios

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Amazon EC2 and AWS Lambda

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 and AWS Lambda

Exam Scenario	Solution
Administrator needs to check if any EC2 instances will be affected by scheduled hardware maintenance	Check the AWS Personal Health Dashboard
Scheduled hardware maintenance will affect a critical EC2 instance	Stop and start the instance to move it to different underlying hardware
When launching an EC2 instance the InsufficientInstanceCapacity error is experienced	This means AWS does not currently have enough capacity to service the request for that instance type. Try a different AZ or instance type
The error InstanceLimitExceeded is experienced when launching EC2 instances	EC2 instance limits have been reached, need to contact support to request an increased limit

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 and AWS Lambda

Exam Scenario	Solution
System status checks are failing for an EC2 instance	Stop and start again to move to a new host
For security and compliance reasons EC2 instances must not be able to access the internet	Launch them in a private subnet without a NAT gateway or NAT instance
EC2 instances must communicate with an internet-based service which whitelists a single source IP address	Place the instances behind a NAT gateway as the device will have a single elastic IP address that can be whitelisted
A distributed app is running on EC2 and can handle processing interruptions. Determine the best pricing model to use	Use Spot instances as the application can handle it if the instances are terminated

DigitalCloud TRAINING

Amazon EC2 and AWS Lambda

Exam Scenario	Solution
Define AWS' responsibilities for EC2 hardware according to the AWS Shared Responsibility Model	AWS are responsible for managing the health of the underlying hosts
A nightly job runs on EC2 and stores results in S3. Takes 2 hours using multiple on-demand instances. If it fails, it must start again. Determine the best pricing model to use	Request a Spot block for time period required
An asynchronous process runs on EC2 and feeds data to a data warehouse for weekly/monthly reporting. Determine the best pricing model to use	Use Spot instances as the asynchronous nature of the reporting means the app can handle interruption if AWS need the capacity back
Need to track EC2 and on-premise computer memory utilization	Install the unified CloudWatch agent on both EC2 and on-premises servers

© Digital Cloud Training | https://digitalcloud.training

Amazon EC2 and AWS Lambda

Exam Scenario	Solution
Amazon EC2 Auto Scaling automatically terminates unhealthy instances but Administrator needs to keep the logs for subsequent analysis	Install the CloudWatch agent to stream logs to CloudWatch Logs
There is a suspected memory leak on an Amazon EC2 instance	Install the CloudWatch agent to monitor memory utilization
An AWS Lambda function is expected to see a large increase in traffic and must scale	Ensure the concurrency limit is higher than the expected simultaneous executions
Need to invoke an AWS Lambda function every 15 minutes	Create an event rule in Amazon CloudWatch events to execute the function periodically

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Elastic Load Balancing and Auto Scaling

© Digital Cloud Training | https://digitalcloud.training

Elastic Load Balancing and Auto Scaling

Exam Scenario	Solution
Design required for highly available and secure website on EC2 with ALB, and DB on EC2	Launch ALB in public subnets, web servers in private subnets and DB layer in private subnets – all layers across AZs
HealthyHostCount metrics for an ALB have dropped from 6 to 2. Need to determine the cause	The health checks on target EC2 instances are failing
An instance attached to an ALB exceeded the UnhealthyThresholdCount for consecutive health check failures. What will happen?	Health checks will continue and the ALB will take the instance out of service

© Digital Cloud Training | https://digitalcloud.training

Elastic Load Balancing and Auto Scaling

Exam Scenario	Solution
Requirement to track the source IP of clients and the instance that processes the request	Check the ALB access logs for this information
Requirement to trigger an alarm when all instances are unhealthy	Use Amazon CloudWatch with the condition: "AWS/ApplicationELB HealthyHostCount <= 0"
Need to check why users cannot connect to web server public IP and port (behind ALB)	Check the VPC Flow Logs

© Digital Cloud Training | https://digitalcloud.training

Elastic Load Balancing and Auto Scaling

Exam Scenario	Solution
HTTPCode_ELB_5XX_Count Amazon CloudWatch metrics are noticed for an ALB	The target group may not contain any healthy instances
CloudWatch shows 4XX errors for app with ALB but the Instances have already been terminated and need to analyze the root cause	Use ELB access logs to retrieve info from S3 bucket to find the originators of the requests
Need a load balancer where specific static public IP addresses can be whitelisted by clients	Use a Network Load Balancer (NLB)

© Digital Cloud Training | https://digitalcloud.training

Elastic Load Balancing and Auto Scaling

Exam Scenario	Solution
Poor performance has been experienced for an application running on Amazon EC2	Use EC2 Auto Scaling to dynamically scale
503 and 504 errors experienced and instances have high CPU utilization	Use EC2 Auto Scaling to dynamically scale
ASG does not launch instances during busy periods despite max capacity not being reached	Could be due to service limits (check Trusted Advisor) or check for RunInstances requests in CloudTrail in case they are failing
Need to analyze instances before they are terminated	Use Auto Scaling lifecycle hooks to pause termination

DigitalCloud TRAINING

Elastic Load Balancing and Auto Scaling

Exam Scenario	Solution
Auto Scaling scales based on queue depth but at beginning of day app slows down	Create a scheduled scaling policy
Create highly available EC2 Auto Scaling group for a single instance app	Use at least 3 AZs, min size of 2, desired capacity of 2, and max of 2
Elastic Beanstalk worker node reads messages from SQS queue. Auto Scaling scales instances. App slows down when number of messages in queue increases	Update ASG to scale on queue depth
ALB is expecting a large spike in traffic and the application is memory heavy	Use the RequestCountPerTarget metric to control scaling

© Digital Cloud Training | https://digitalcloud.training

Elastic Load Balancing and Auto Scaling

Exam Scenario	Solution
New instances in an Auto Scaling group are not showing up in the aggregated metrics. Step scaling is used	Likely due to the warm-up period having not yet expired

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Amazon EBS, EFS, and AWS Storage Gateway

© Digital Cloud Training | https://digitalcloud.training

Amazon EBS, EFS, and AWS Storage Gateway

Exam Scenario	Solution
User deleted some data in an Amazon EBS volume and there's a recent snapshot	Can create a new EBS volume from the snapshot and attach it to an instance and copy the delete file across
EBS volume runs out of space and need to prevent it happening again	Use CloudWatch agent on EC2 and monitor disk metrics with CloudWatch alarm
Most cost-effective option for big data app that stores sequentially and infrequent access	Cold HDD (sc1)
EBS volume capacity is increased but cannot see the space	Need to extend the volume's file system to gain access to extra space

© Digital Cloud Training | https://digitalcloud.training

Amazon EBS, EFS, and AWS Storage Gateway

Exam Scenario	Solution
Need to replace user-shared drives. Must support POSIX permissions and NFS protocols and be accessible from on-premise servers and EC2	Use Amazon EFS
Low latency access required for image files in an office location with synchronized backup to offsite location. Local access required and disaster recovery	Use an AWS Storage Gateway volume gateway configured as a stored volume
Performance issues with iSCSI drives in volume gateway. CacheHitPercent metric is below 55% and CachePerecentUsed is above 95%	Create a larger disk for cached volume and select it in management console
Tape archival system needs replacement	Use an AWS Storage Gateway tape gateway

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
AWS Systems Manager and OpsWorks

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager and OpsWorks

Exam Scenario	Solution
Application running on EC2 needs login credentials for a DB that are stored as secure strings in SSM Parameter Store	Create an IAM role for the instance and grant permission to read the parameters
Linux instances are patched with Systems Manager Patch Manager. Application slows down whilst updates are happening	Change maintenance window to patch 10% of instances in the patch group at a time
Custom Linux AMI used with AWS Systems Manager. Can't find instances in Session Manager console	Need to add permissions to instance profile and install the SSM agent on the instances

© Digital Cloud Training | https://digitalcloud.training

AWS Systems Manager and OpsWorks

Exam Scenario	Solution
Multiple environments require authentication credentials for external service. Deployed using CloudFormation	Store credentials in SSM Parameter Store and pass an environment tag as a parameter in CloudFormation template
IAM access keys used to manage EC2 instances using the CLI. Company policy mandates that access keys are automatically disabled after 60 days	Use an AWS Config rule to identify noncompliant keys. Create a custom AWS Systems Manager Automation document for remediation

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
AWS CloudFormation

© Digital Cloud Training | https://digitalcloud.training

AWS CloudFormation

Exam Scenario	Solution
Need to review updates to a CloudFormation stack before deploying them in production	Use change sets
Stack deployed and manual changes were made. Need to capture changes and update template	Use drift detection and use output to update template and redeploy the stack
Need to update new version of app on EC2 and ALB. Must avoid DNS changes and be able to rollback	Update template with AutoScalingReplacingUpdate policy and perform an update
Need to write a single template that can be deployed across several environments / Region	Use parameters to enter custom values and use Ref intrinsic function to reference the parameter
Tried to launch instance in a different region from a working template and it fails	Probably due to incorrect AMI ID

© Digital Cloud Training | https://digitalcloud.training

AWS CloudFormation

Exam Scenario	Solution
CloudFormation stack created for first time and fails with ROLLBACK_COMPLETE status	To continue administrator must relaunch the template to create a new stack
Template for infrastructure in one region used to deploy in another and fails	Template likely referenced an AMI that doesn't exist in the new region and/or services that don't exist
CloudFormation stack fails and returns UPDATE_ROLLBACK_FAILED	Fix the error that caused the rollback to fail and then select "Continue update rollback" in the console
Need to deploy a single CloudFormation template across multiple accounts	Use StackSets
CloudFormation deploys stack with separate VPC for each app. Fails to deploy	May have reached the default limit for VPCs in the account

DigitalCloud TRAINING

AWS CloudFormation

Exam Scenario	Solution
Would like to manually address any issues with CloudFormation stack creation	Set the OnFailure parameter to "DO_NOTHING”
CloudFormation fails with "The image id ‘[ami-2a69aa47]’ does not exist"	Most likely the template is being run in a different region where the AMI does not exist
When creating Stack a wait condition error is experienced: ""received 0 signals out of the 1 expected from the EC2 instance"."	Check instance has a route through NAT device and in the cfn logs confirm that the cfn-signal command ran successfully

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Amazon VPC

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

Exam Scenario	Solution
Need to identify the instances that are generating the most traffic using a NAT gateway	Use VPC flow logs on the NAT gateway ENI and use CloudWatch insights to filter based on source IP address
NAT gateway is NOT highly available across AZs, only within an AZ	Use multiple NAT gateways for HA across AZs
Need to enable access to S3 without the instances using public IP addresses	Use a NAT gateway or VPC endpoint

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

Exam Scenario	Solution
EC2 instance in private subnet cannot reach the Internet. Route table has a route to a NAT gateway with a status of "Blackhole"	Indicates the NAT gateway has been deleted
Need to connect to S3 from EC2 using private network only. Must also ensure that only the instances can access the bucket	Create a VPC endpoint and a bucket policy with a Condition that limits S3 actions to the VPC endpoint as the source
VPC endpoint setup to allow private IP address connectivity to S3 bucket, permissions configured, but instances still can't connect	Make sure the subnet has a target in the route table for the VPC endpoint

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

Exam Scenario	Solution
Need to manage EC2 instances in a private subnet from an office using SSH but instances cannot have internet access	Add a VGW and configure routing in the VPC and establish a VPN to the office
Need encryption in-transit and at-rest for hybrid environment	Use an AWS VPN and use KMS keys for data encryption
Network change was made that resulted in application to DB connection issues	Analyze using VPC Flow Logs
Inbound and outbound internet connectivity required for EC2 instances	Need to attach an internet gateway to the VPC and add an entry in the route table for the subnet that points to the internet gateway

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

Exam Scenario	Solution
Web application has EC2 with public IPs behind an ALB. EC2 instances cannot connect to external service	Need to create an attach an IGW to the VPC and update the route table
VPC peering connection setup between two different VPCs. Instances in private subnets still can't communicate	Make sure the route tables are updated
A company has configured a VPC peering connection between two VPCs and needs to set up connectivity between instances in private subnets	Configure the VPC route tables with routes pointing to the address range of the other VPC
Company backing up one VPC to another in different region. All data must be private and encrypted	Use inter-region VPC peering which encrypts across the AWS global network

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

Exam Scenario	Solution
Malicious IP identified and must be blocked from all ingress and egress connectivity	Add a rule to a network ACL for all affected subnets
VPC connected to data center by VPN. User pings private subnet instance from on-prem computer and fails. VPC Flow Logs show accept for inbound but reject for outbound traffic	Modify the network ACL to allow outbound traffic
Malicious traffic coming from a single IP address	Use a NACL for the web server subnet to deny IP address
Admin has setup instance for remote access and can SSH from internet but cannot ping	Most likely reason is that the instance's security group does not have a rule allowing ICMP

© Digital Cloud Training | https://digitalcloud.training

Amazon VPC

Exam Scenario	Solution
Admin connecting to EC2 instance using SSH from office but gets connection timeout from home	Most likely doesn't have the home network IP range in the security group allow rule for SSH

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Amazon Route 53

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53

Exam Scenario	Solution
Use Route 53 to direct based on health checks with (2xx) traffic to primary and other responses to secondary	Need to create an A record for each server and a HTTP (not TCP) health check
Route 53 health check uses string matching for "/html". Alert shows health check fails	The search string must appear entirely within the first 5,120 bytes of the response body
Need to make a website promotion visible to users from a specific country only	Use Route 53 geolocation routing policy

© Digital Cloud Training | https://digitalcloud.training

Amazon Route 53

Exam Scenario	Solution
New website runs on EC2 behind ALB. Need to create record in Route 53 to point to the domain apex (e.g. example.com)	Use an alias record
Hosted zone in Account A and ALB in Account B. Need the most cost-effective and efficient solution for pointing to the ALB	Create an Alias record in Account A that points to ALB in Account B

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Amazon S3 and CloudFront

© Digital Cloud Training | https://digitalcloud.training

Amazon S3 and CloudFront

Exam Scenario	Solution
Static website on Amazon S3 with custom domain name	Requires that the bucket name matches the DNS name / record set name in Route 53
503 errors experienced with new site and thousands of user	Request rate is too high
Discrepancy with number of objects in bucket console vs CloudWatch	Use Amazon S3 Inventory to properly determine the number of objects in a bucket

© Digital Cloud Training | https://digitalcloud.training

Amazon S3 and CloudFront

Exam Scenario	Solution
Unauthorized users tried to connect to S3 buckets. Need to know which buckets are targeted and who is trying to get access	Use S3 server access logs and Athena to query for HTTP 403 errors and look for IAM user or role making requests
Need to provide access to third-party to S3 bucket and must limit amount of access. List of users changes a lot	Use a pre-signed URL allowing access to the specific files
Need to protect S3 data from ransomware attacks that encrypt data	Enable S3 versioning
After enabling MFA on a bucket, what operations will require MFA authentication?	Permanently removing object versions and suspending versioning on the bucket

© Digital Cloud Training | https://digitalcloud.training

Amazon S3 and CloudFront

Exam Scenario	Solution
Files are downloaded from S3, edited and uploaded with same file name. Sometimes they are accidentally modified or deleted	To allow recovery enable versioning on the bucket
Existing application uses RDS, EFS and S3. Need to enable encryption	Encryption is enabled by default on S3. You cannot change encryption status after deployment for RDS and EFS
Static website deployed but "HTTP 403 Forbidden" message received	Add bucket policy granting everyone read access to objects
Application on EC2 must save files to Amazon S3 and needs access	Create an IAM role for S3 access and attach to EC2 instance

© Digital Cloud Training | https://digitalcloud.training

Amazon S3 and CloudFront

Exam Scenario	Solution
History of revisions to files stored in an S3 bucket must be maintained	Implement S3 versioning
Large volume of log files stored in S3 bucket and processed daily	Most cost-effective option is S3 standard
Need to restrict S3 bucket access to same account after previously shared with other account	Change ACL to restrict only to bucket owner
Static content is served from Amazon S3 with long loading times	Use CloudFront to cache for better performance
Need to use custom domain name with CloudFront	Create an alias record in Route 53 pointing to the distribution URL

© Digital Cloud Training | https://digitalcloud.training

Amazon S3 and CloudFront

Exam Scenario	Solution
CloudFront in front of ALB and EC2 and logging enabled. Need to view logs for HTTP layer 7 status codes	Check ALB access logs and CloudFront access logs
App running on EC2 with RDS multi-AZ has static content on S3. Need to improve performance as load testing slowed it down	Use CloudFront to cache the content
Need to secure S3 bucket that is used with CloudFront	Use an OAI and grant permissions to read objects in the bucket
Website with dynamic content and need to restrict access from certain countries and regions	Use Amazon CloudFront geo-restriction and Amazon Route 53 geolocation routing

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Amazon RDS and ElastiCache

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS and ElastiCache

Exam Scenario	Solution
Automated failover of a multi-AZ DB occurred	This may be due to storage failure on primary DB or the instance type could have been changed
Need to encrypt unencrypted RDS database	Take a snapshot, encrypt it, then restore a new encrypted instance from the snapshot
RDS DB query latency is high and CPU utilization is at 100%	Scale up with larger instance type
Need to share RDS DB snapshots across different accounts. Data must be encrypted	Use an AWS KMS key for encryption and update key policy to grant accounts with access then share snapshot

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS and ElastiCache

Exam Scenario	Solution
DB needs to be made HA to protect against failure and updates cannot impact users in business hours	Change to Multi-AZ outside of business hours
Need to protect RDS databases against table corruption within a 30 day window of protection	Enable automated backups and set the appropriate retention period
Shared Responsibility Model	AWS is responsible for maintenance, patches and other updates for Aurora DB
AuroraReplicaLagMaximum is high for DB on eCommerce site. What affect could this have?	may result in cart not updating correctly (inconsistency)

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS and ElastiCache

Exam Scenario	Solution
EC2 connects to RDS instance and fails with: "Error Establishing a Database Connection”	Web server may be using certificate validation and RDS does not trust the certificate. Or, the DB security group does not have the correct ingress rule
Aurora DB is hitting 100% CPU. Read-heavy app with many lookups	Add Aurora Replicas and use a Reader Endpoint for product table lookups
Database is running MySQL on Amazon EC2. Need to increase availability and durability without changing application	Use Aurora MySQL and configure an Aurora Replica in another AZ
Reporting job runs against RDS instance and is causing performance issues	Create a read replica and point the reporting job to the read replica endpoint

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS and ElastiCache

Exam Scenario	Solution
Backup of RDS instance must be copied regularly to another account for testing	Create a snapshot with create-db-snapshot CLI, share with other account, then create a copy in that account
MySQL database on RDS must be patched due to a security vulnerability. Who is responsible?	AWS is responsible for patching Amazon RDS database instances
Reporting job runs against RDS instance and is causing performance issues	Create a read replica and point the reporting job to the read replica endpoint

© Digital Cloud Training | https://digitalcloud.training

Amazon RDS and ElastiCache

Exam Scenario	Solution
How can a Redis cluster be scaled to improve read times	Scale horizontally by adding shards
High CPU on a Memcached cluster	Options are to add additional nodes to cluster or vertically scale the node types
ElastiCache Memcached storing session state. Performance poor, eviction count metrics are high	Scale the cluster by adding additional nodes
A Memcached cluster is experiencing increased traffic, need to change to larger node type	Create a new cache cluster with the new node type using the CreateCacheCluster API

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Management, Governance and Billing

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Audit requests to AWS Organizations for creating new accounts by federated users	Use CloudTrail and look for the federated identity user name
Employees have created individual AWS accounts not under control. Security team need them in AWS Organizations	Send each account an invitation from the central organization
Need to restrict ability to launch specific instance types for a specific team/account	Use an organizations SCP to deny launches unless the instance type is T2, create an IAM group in the account granting access to T2 instances to the relevant users

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Need to ensure that S3 buckets are NEVER deleted in a production account	Use an SCP to deny the s3:DeleteBucket API action
Need to create user-defined cost allocation tags for new account	Use Tag Editor in new account to create user-defined tags and then use the billing and cost management console in the payer account to mark them as cost allocation tags
Separate departments must operate in isolation and only use pre-approved services	Use AWS Organizations to create accounts (Organizations API) and SCPs to control the services available for use

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Developers can manipulate IAM policies/roles and need to block them from some services	Use an SCP to block those services
AWS bill is increasing and unauthorized services are being used across accounts	Use AWS Organizations with an SCP to restrict the unauthorized services
Configuring IAM Identity Center for an Organizations main account. Directory created and full access enabled	Next step is to create a permission set and associate with directory users and groups
Process to create a custom dashboard in CloudWatch for custom metrics after installing agent on EC2	Create metric filters and select custom metrics

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Need to test notification settings for CloudWatch alarm with SNS	Use the set-alarm-state CLI command to test
App with EC2 and RDS is running slowly and suspected high CPU	Use CloudWatch metrics to examine resource usage
Site uses CloudFront and S3. Users accessing content that does not exist or they don't have access to	Check the 4XXErrorRate metric in CloudWatch to understand the extent of the issue
Script generates custom CloudWatch metrics from EC2 instance and clock is configured incorrectly by 30 mins	CloudWatch will accept the custom metric data and record it

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Need to collect logs from many EC2 instances	Use the unified CloudWatch Agent
External auditor needs to check for unauthorized changes to AWS account	Create an IAM user, assign an IAM policy with read access to CloudTrail logs on Amazon S3
Need to identify who is creating EIPs and not using them	Use CloudTrail and query logs using Athena to search for EIP address events
S3 bucket holds sensitive data. Must monitor object upload / download activity including AWS account and IAM user account of caller and time of API call	Use AWS CloudTrail and enable data event logging

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Need to record any modifications or deletions of CloudTrail logs in an S3 bucket	Enable CloudTrail log file integrity validation and enabled MFA delete on the bucket
Large increase in requests to SQS. Need to determine the source of the calls	Use CloudTrail to audit API calls
Need to ensure that S3 buckets have logging enabled without stopping users creating them	Auto remediate with AWS Config managed rule S3_BUCKET_LOGGING_ENABLE
Need to provide real-time compliance reporting for security groups to check that port 80 is not being used	Use the AWS Config restricted-common-ports rule and add port 80

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Company wants to limit the AMIs that are used. Need to review compliance with the policy	Create an AWS Config rule to check that only approved AMIs are used
Need to automatically disable access keys that are greater than 90 days old	Use Config rule to identify noncompliant keys and use Systems Manager Automation to remediate
Need to address concerns about exposing sensitive data in buckets without restricting ability to create them	Use AWS Config rules to identify public buckets and send SNS notification to security team
Need to ensure CloudFormation deployment changes are tracked for governance	Use AWS Config

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Company needs to verify that specific KMS CMK is used to encrypted EBS volumes	Use AWS Config with the encrypted-volumes managed rule and specify the key ID of the CMK
Need to create replica of existing infrastructure in new account. AWS Service Catalog is used	Most efficient option is to share the portfolio with the new accounts and import into those other accounts
Users have a specialized EC2 instance config and don't want to configure EC2 settings but need to launch/terminate instances. Special instance must only be available to them	Use CloudFormation template with AWS Service Catalog portfolio and grant permissions to users
Shared portfolio is imported into a second AWS account controlled by a different administrator	Admin can add products from the imported portfolio to a local portfolio

DigitalCloud TRAINING

Management, Governance and Billing

Exam Scenario	Solution
Need to monitor costs per user in an account	Activate the createdBy tag and analyze with AWS Cost Explorer
How to check for underutilized EC2 instances?	Use AWS Cost Explorer to generate resource optimization recommendations
Bill is increasing over time, need to determine the cause of increased cost	Use AWS Cost Explorer
Need breakdown of costs per project in a single account using Cost Explorer	Do this by activating cost allocation tags and creating and applying resource tags

© Digital Cloud Training | https://digitalcloud.training

Management, Governance and Billing

Exam Scenario	Solution
Need to check that security best practices are being followed for the AWS account root user	Use AWS Trusted Advisor security checks to review configuration of root user
Costs rising and need to be alerted when a specific spending limit is forecast to be exceeded	Use AWS Budgets
Company needs to track the allocation of reserved instances in consolidated bill	Use the AWS Cost and Usage report
Company needs to integrate AWS maintenance events that may affect their resources into an operations dashboard	Use the AWS Health API

© Digital Cloud Training | https://digitalcloud.training

Exam Scenarios
Security and Compliance

© Digital Cloud Training | https://digitalcloud.training

Security and Compliance

Exam Scenario	Solution
Company wishes to force users to change their passwords regularly	Create an IAM password policy and enabled password expiration
Need to restrict access to a bucket based on source IP range	Use bucket policy with "Condition": "NotIpAddress": statement`
Need to control access to group of EC2 instances with specific tags	Use an IAM policy with a condition element granting access based on the tag and attach an IAM policy to the user or groups that require access
IAM policy for SQS queue allows too much access. Who is responsible for correcting the issue?	According the AWS shared responsibility mode, this is a customer responsibility

© Digital Cloud Training | https://digitalcloud.training

Security and Compliance

Exam Scenario	Solution
Data is encrypted with AWS KMS customer-managed CMKs. Need to enable rotation ensuring the data remains readable	Just enable key rotation in AWS KMS for the KMS key (backing key is rotated, data key is not changed)
Company must rotate encryption keys once a year with least effort	Use customer-managed KMS key and enabled automatic key rotation
App uses KMS key with imported key material and references the KMS key by alias in the application. Must be rotated every 6 months	To rotate, create a new KMS key with new imported material and update the key alias to point to new KMS key

© Digital Cloud Training | https://digitalcloud.training

Security and Compliance

Exam Scenario	Solution
Certificate request rejected by ACM	Submit a request for a certificate using the correct domain name NOT the ALB FQDN
Security findings are missing in Amazon Inspector	Verify agent installed on affected instances and restart agent
Security team need to verify vulnerabilities and exposures are addressed for EC2 instances regularly	Use Amazon Inspector and perform regular assessments
There may be a vulnerable version of software installed on EC2 instances and need to check	Create and run an Amazon Inspector assessment template

© Digital Cloud Training | https://digitalcloud.training

Security and Compliance

Exam Scenario	Solution
Need to use information in request header to count requests from each front-end server	Use a string match statement
Large amount of suspicious HTTP requests hitting an ALB from various source IPs	Block the traffic using AWS WAF with a rate-based rule and a defined threshold
Many 404 errors being sent to one IP address every minute. Bot may be collecting info	Use AWS WAF to block the activity
Website has been deployed and penetration testing shows its vulnerable to cross-site scripting	Use AWS WAF to mitigate cross-site scripting attacks

© Digital Cloud Training | https://digitalcloud.training

Security and Compliance

Exam Scenario	Solution
Application is under repeated DDoS attacks. Need to minimize downtime and require 24/7 support	Setup AWS Shield Advanced
Company needs to understand the PCI status of the AWS infrastructure	Use AWS Artifact to locate this information

© Digital Cloud Training | https://digitalcloud.training

Security and Compliance

Exam Scenario	Solution
Company uses LDAP and needs to implement access control in AWS as part of an integration between internal and cloud	Need to configure SAM federation of IAM users and groups with the LDAP DB and map LDAP user and groups to IAM roles
Permissions policy for cross-account access must be created and attached. Who is responsible for doing this?	According to the AWS shared responsibility model, this is a customer responsibility
Company wishes to move from IAM user accounts to using on-premises Active Directory accounts for AWS management console access	Configure a VPN tunnel and use Active Directory Connector

© Digital Cloud Training | https://digitalcloud.training