

# AWS Foundational Services



# Compute: Amazon Elastic Compute Cloud (EC2)



# Amazon Elastic Compute Cloud (EC2)



Amazon  
EC2

- **Resizable compute** capacity
- Complete control of your resources:
  - CPU
  - Memory
  - Networking
  - Storage
- Reduced time required to obtain and boot new server instances

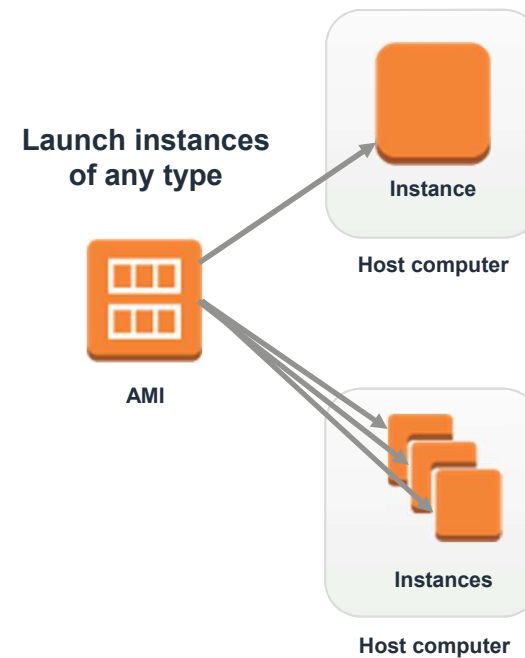


# Instances and AMIs



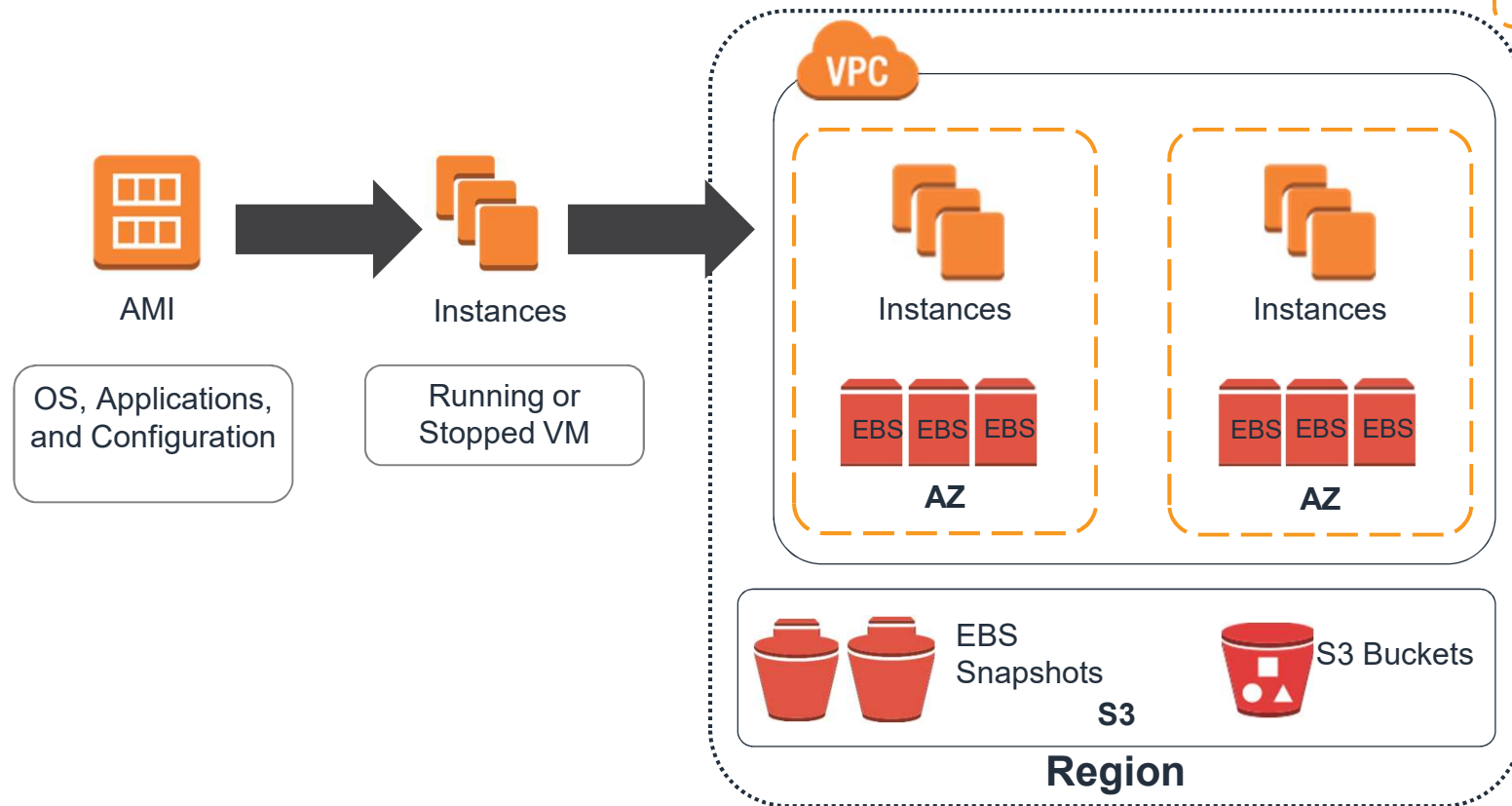
Select an AMI based on:

- Region
- Operating system (Windows or Linux)
- Architecture (32-bit or 64-bit)
- Launch permissions
- Storage for the root device





# Amazon EC2 Instances





# AWS Marketplace

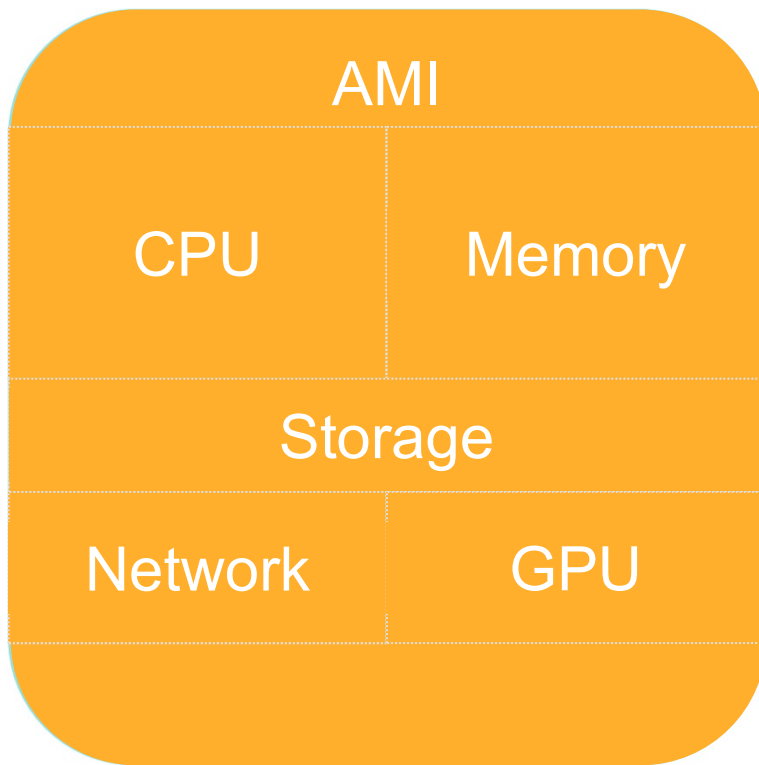


- Online store to discover, purchase, and deploy IT software on top of the AWS infrastructure.
- Catalog of 4000+ IT software solutions including Paid, BYOL, Open Source, SaaS, and free-to-try options.
- Pre-configured to operate on AWS.
- Software checked by AWS for security and operability.

The screenshot shows the AWS Marketplace homepage. At the top, there's a navigation bar with the AWS Marketplace logo, a search bar, and links for 'Sign in or Create a new account', 'Your Account', 'Help', and 'Sell on AWS Marketplace'. Below the navigation bar, there's a 'Shop All Categories' sidebar on the left, listing various software categories like Desktop Apps, Software Infrastructure, Application Development, etc. The main content area features a large banner for 'Production-ready cluster deployments in minutes with AWS Marketplace and AWS CloudFormation'. Below the banner, there are sections for 'Featured Products' and 'Operating Systems'. The 'Featured Products' section includes items like WebSphere Application Server, Matillion ETL for Redshift, and TIBCO Clarity. The 'Operating Systems' section lists various Linux distributions like Amazon Linux, CentOS, Oracle Linux, Ubuntu, and Red Hat Enterprise Linux. Each product listing includes a logo, a brief description, and pricing information.



# The inside of an Instance



CPU	Sockets, number of cores, clock frequency, bursting capability
Memory	Memory capacity
Disks	Input / output operations per second, throughput
Network Interface	Max bandwidth, packet rate



## Type and size

Instance generation

c4.large



Instance family

Instance size

The diagram illustrates the components of the AWS instance type 'c4.large'. The text 'c4.large' is centered. Above it, 'Instance generation' is written with a bracket pointing to the 'c4' part. Below 'c4', 'Instance family' is written with a bracket pointing to 'c4'. Below 'large', 'Instance size' is written with a bracket pointing to 'large'.



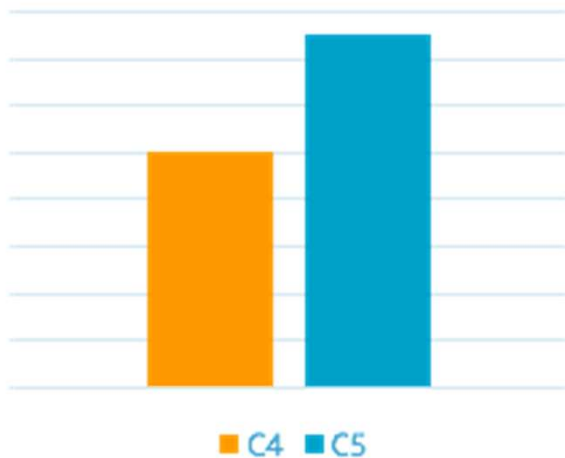
# Amazon EC2 Instance Types

General Purpose	Compute Optimized	Memory Optimized	Accelerated Computing	Storage Optimized
t2	c5 / c5d	x1 / x1e	g3	h1
m5 / m5d	 c4	r4	p3	i3
m4 		r5 / r5d	p2	d2
		z1d	f1	



# C5: Compute Optimized Instances

25% price/performance  
improvement over C4



- Based on 3.0 GHz Intel Xeon Scalable Processors (Skylake)

- Up to 72 vCPUs and 144 GiB of memory (2:1 Memory:vCPU ratio)

- 25 Gbps NW bandwidth. •Support for Intel AVX-512

**NETFLIX**

*"We saw significant performance improvement on Amazon EC2*

*C5, with up to a 140% performance improvement in industry standard CPU benchmarks over C4."*

**GRAIL**

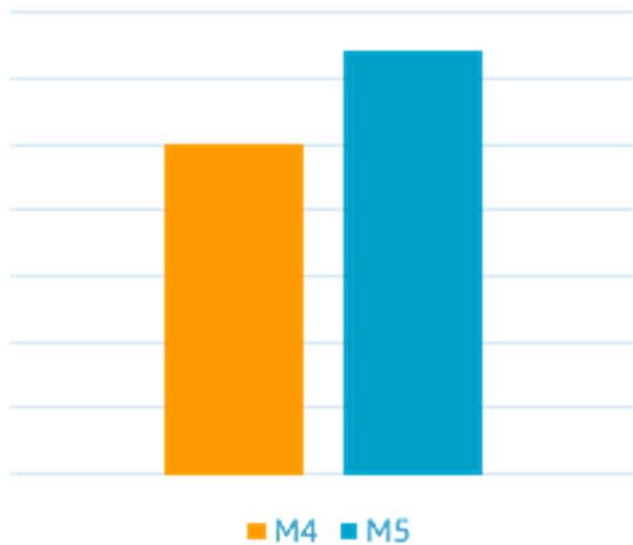
*"We are eager to migrate onto the AVX-512 enabled c5.18xlarge*

*instance size... We expect to decrease the processing time of some of our key workloads by more than 30%."*



## M5: Next-Gen General Purpose instance


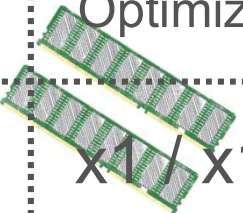



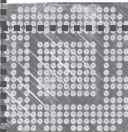
**14% price/performance improvement With M5**



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



# Amazon EC2 Instance Types

General Purpose	Compute Optimized	Memory Optimized	Accelerated Computing	Storage Optimized
 t2	c5 / c5d	 x1 / x1e	g3	h1
m5 / m5d	 c4	r4	p3	 i3 → ← d2
m4 		r5 / r5d	p2	
		z1d	 f1	



# Instance User Data



- Can be passed to the instance at launch.
- Can be used to perform common automated configuration tasks.
- Runs scripts after the instance starts.
- You can specify user data when launching an instance.
- User data can be:
  - Linux script – executed by cloud-init
  - Windows batch or PowerShell scripts – executed by EC2Config service
- User data scripts run once per instance ID by default.



# User Data Example Linux



```
#!/bin/sh
```

```
yum -y install httpd  
chkconfig httpd on  
/etc/init.d/httpd start
```

User data shell scripts must start with the `#!` characters and the path to the interpreter you want to read the script.

Install Apache web server  
Enable the web server  
Start the web server



# Amazon EC2 Purchasing Options



## On-Demand Instances

Pay by the ~~hour~~  
\*Second  
!

## Reserved Instances

Purchase, at a significant discount, instances that are always available

1-year to 3-year terms.

## Spot Instances

Spare compute capacity in the AWS cloud at **steep discounts**. But...

## Dedicated Hosts

Pay for a physical host that is fully dedicated to running your instances.

- Per-second billing - Amazon Linux and Ubuntu only
- Per-hour billing - All other OSs



# Networking: Virtual Private Cloud (VPC)



# Amazon Virtual Private Cloud (VPC)



Amazon  
VPC

- Provision a **private, isolated** virtual network on the AWS cloud.
- Have complete control over your virtual networking environment.



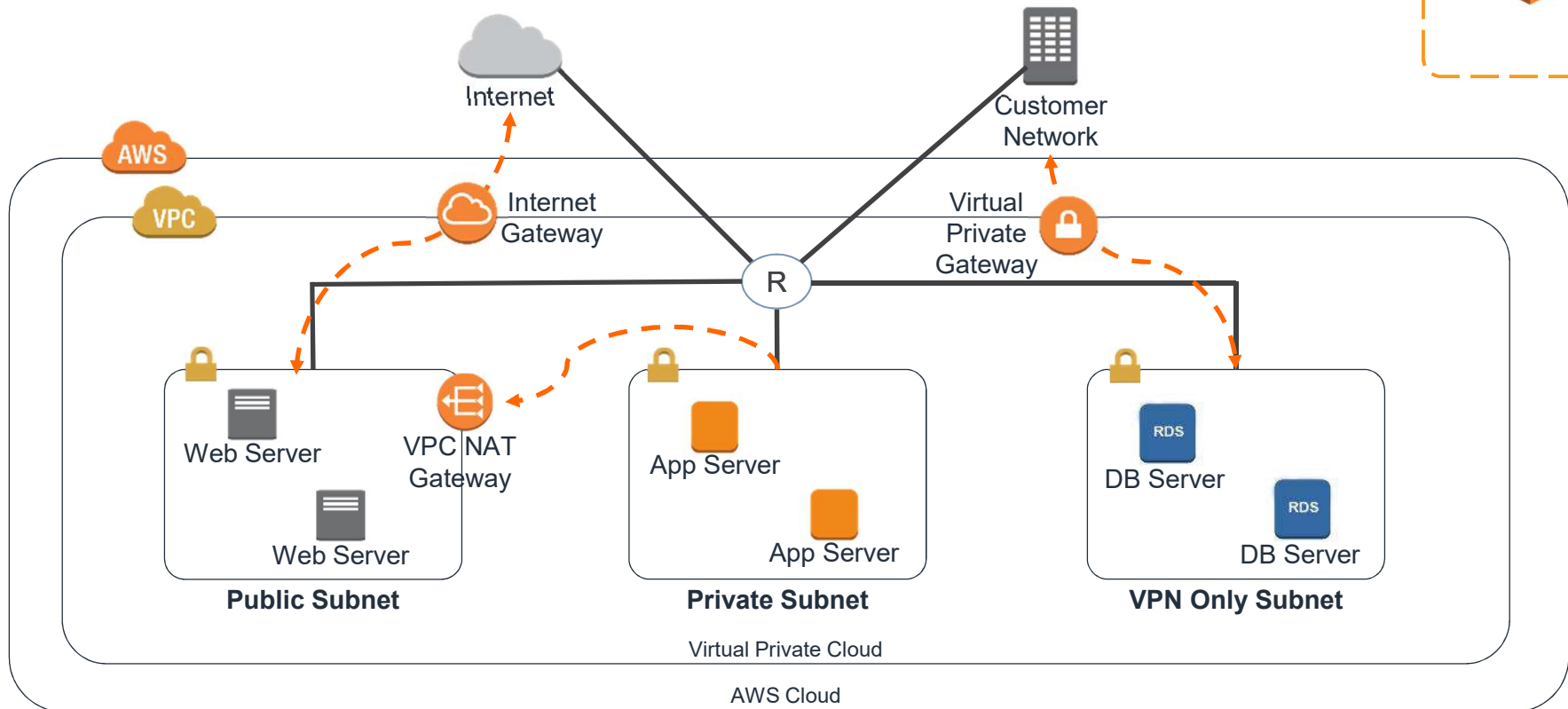
# VPCs and Subnets



- A subnet defines a range of IP addresses in your VPC.
  - You can launch AWS resources into a subnet that you select.
  - A **private subnet** should be used for resources that won't be accessible over the Internet.
  - A **public subnet** should be used for resources that will be accessed over the Internet.
  - Each subnet must reside entirely within one Availability Zone and cannot span zones.
-



# Amazon VPC Example

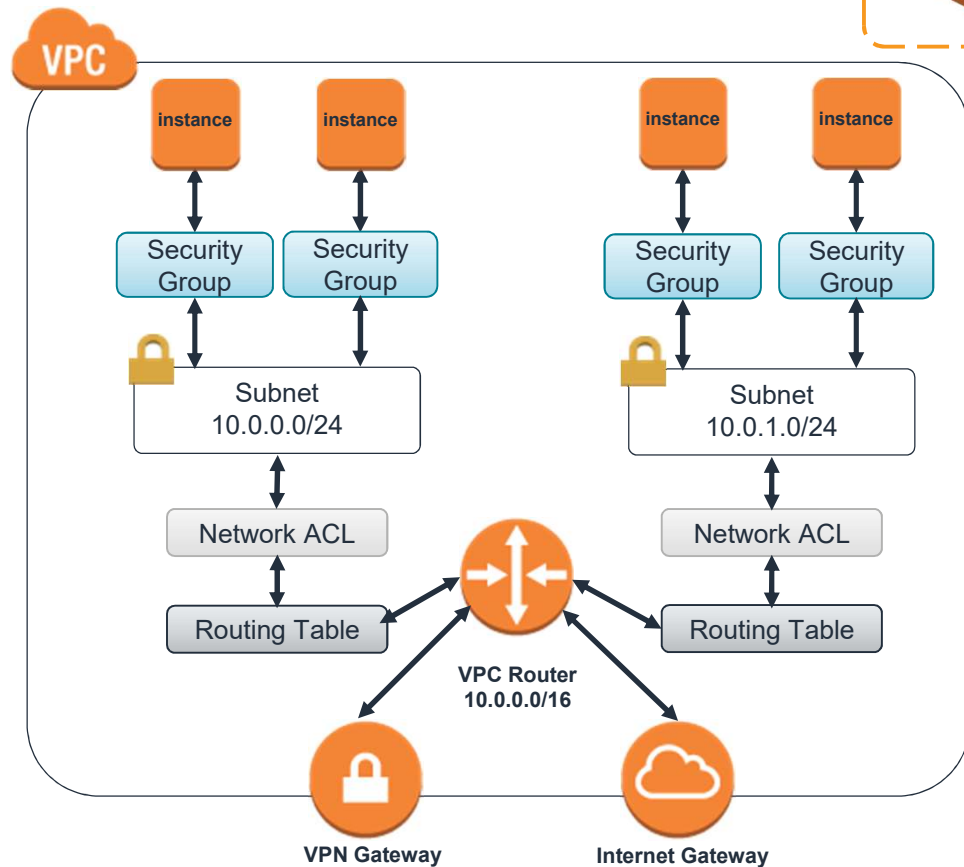




# Security in Your VPC



- OS controls
- Security groups
- IPs
- Network access control lists (ACLs)
- Routing tables
- Gateways





# VPN Connections



VPN Connectivity option	Description
AWS Hardware VPN	You can create an <b>IPsec</b> hardware VPN connection between your VPC and your remote network.
AWS Direct Connect	AWS Direct Connect provides a <b>dedicated private</b> connection from a remote network to your VPC.
Software VPN	You can create a VPN connection to your remote network by using an Amazon EC2 instance in your VPC that's running a <b>software VPN appliance</b> .



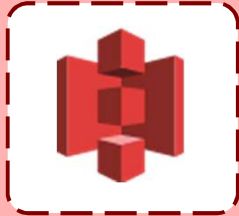


## Storage:

- Amazon Simple Storage Service (S3)
- Amazon Elastic Block Store (EBS)



# Amazon Simple Storage Service (S3)



Amazon S3

- Storage for the Internet
- Natively online, HTTP access
- Storage that allows you to store and retrieve **any amount** of data, **any time**, from **anywhere** on the web
- Highly scalable, reliable, fast and durable



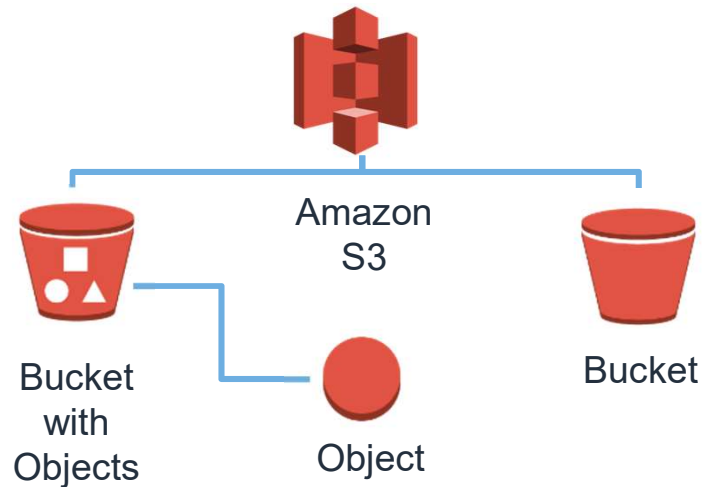
# Amazon S3 Facts



- Can store an unlimited number of objects in a bucket
  - Objects can be up to 5 TB; no bucket size limit
  - Designed for 99.999999999% durability and 99.99% availability of objects over a given year
  - Can use HTTP/S endpoints to store and retrieve any amount of data
  - Provides standards-based REST and SOAP interfaces
-



# Amazon S3 Concepts



- Amazon S3 stores data as objects within buckets
- An object is composed of a file and optionally any metadata that describes that file
- You can have up to 100 buckets in each account
- You can control access to the bucket and its objects



# Object Keys



An object key is the unique identifier for an object in a bucket.

http://**not-taken**.s3.amazonaws.com/**mysecret.gif**

Bucket      Object/Key

A diagram illustrating the components of an S3 URL. The URL 'http://not-taken.s3.amazonaws.com/mysecret.gif' is shown at the top. Below it, the word 'Bucket' is positioned under 'not-taken' and 'Object/Key' is positioned under 'mysecret.gif'. Two curved arrows point from these labels to their respective parts in the URL: one from 'Bucket' to 'not-taken' and another from 'Object/Key' to 'mysecret.gif'.



# Amazon S3 Security



- You can control access to buckets and objects with:
    - Access Control Lists (ACLs)
    - Bucket policies
    - Identity and Access Management (IAM) policies
  - You can upload or download data to Amazon S3 via SSL encrypted endpoints.
  - You can encrypt data using AWS SDKs.
-



# Amazon S3 Versioning



- Protects from accidental overwrites and deletes with no performance penalty.
- Generates a new version with every upload.
- Allows easily retrieval of deleted objects or roll back to previous versions.
- Three states of an Amazon S3 bucket
  - Un-versioned (default)
  - Versioning-enabled
  - Versioning-suspended



Versioning Enabled



# Amazon S3 Object Lifecycle



Lifecycle management defines how Amazon S3 manages objects during their lifetime. Some objects that you store in an Amazon S3 bucket might have a well-defined lifecycle:

- Log files
  - Archive documents
  - Digital media archives
  - Financial and healthcare records
  - Data that must be retained for regulatory compliance
-



# Amazon S3 Pricing



- Pay only for what you use
- No minimum fee
- Prices based on location of your Amazon S3 bucket
- Pricing is available as:
  - Storage Pricing
  - Request Pricing
  - Data Transfer Pricing: data transferred out of Amazon S3





# Amazon Glacier



- Long term low-cost archiving service
- Optimal for infrequently accessed data
- Designed for 99.999999999% durability
- Less than \$0.01 per GB/month (depending on region)

Expedited

-

1–5 minutes

Standard

-

3–5 hours

Bulk

-

5–12 hours



# Amazon S3 Storage Classes

Storage Class	Durability	Availability	Other Considerations
Amazon S3 Standard	99.999999999%	99.99%	
Amazon S3 Standard - Infrequent Access (IA)	99.999999999%	99.9%	<ul style="list-style-type: none"> <li>Retrieval fee associated with objects</li> <li>Most suitable for infrequently accessed data</li> </ul>
Amazon S3 One Zone-Infrequent Access	99.999999999%	99.5%	<ul style="list-style-type: none"> <li>stores data in a single AZ</li> <li>costs 20% less than Standard - IA</li> </ul>
Glacier	99.999999999%	99.99% <small>[1 SEP]</small> (once restored)	<ul style="list-style-type: none"> <li>Must restore objects before you can access them</li> <li>Restoring objects can take 3-5 hours</li> </ul>



# Amazon Elastic Block Store (EBS)

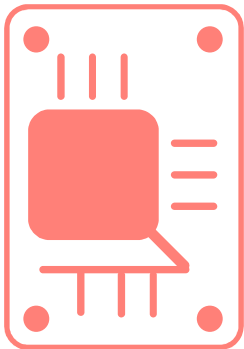


Amazon  
EBS

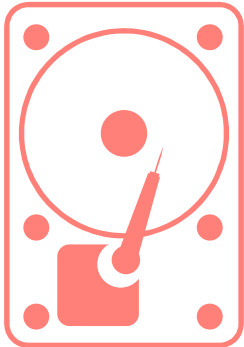
- **Persistent block level** storage volumes offer consistent and low-latency performance.
- Stored data is **automatically replicated** within its Availability Zone.
- Snapshots are stored durably in Amazon S3.



# Amazon EBS Volume Types



- SSD-backed volumes are
  - Optimized for transactional workloads that involve frequent read/write operations with small I/O size.
  - Dominant in IOPS performance.



- HDD-backed volumes are
    - Optimized for large streaming workloads.
    - Dominant in throughput (measured in MiB/s).
-



# Amazon EBS Volume Types



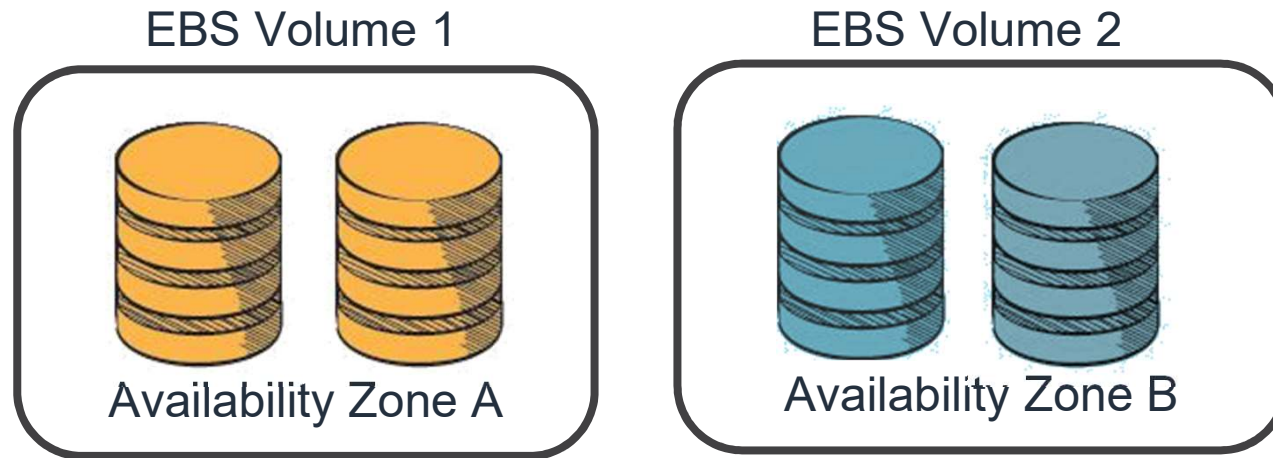
	SSD		HDD	
Volume Type	General Purpose SSD (gp2)	Provisioned IOPS SSD (io1)	Throughput Optimized HDD (st1)	Cold HDD (sc1)
Description	Balances price and performance for a wide variety of <b>transactional</b> loads.	<b>Highest-performance</b> SSD volume designed for <b>mission-critical</b> applications.	<b>Low-cost</b> HDD designed for <b>frequently accessed, throughput-intensive</b> workloads.	<b>Lowest cost</b> HDD designed for less frequently accessed workloads.
Volume Sizes	1 GiB – 16 TiB	4 GiB – 16 TiB	500 GiB – 16 TiB	500 GiB – 16 TiB
Dominant Performance Attribute	IOPS	IOPS	MiB/s	MiB/s



# Amazon EBS Scope



Amazon EBS volumes are in a single Availability Zone



Volume data is replicated across multiple servers in an Availability Zone.

---



# Amazon EBS Pricing





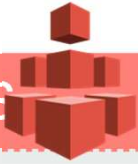
Pay for what you provision:

- Pricing based on region
- Review Pricing Calculator online
- Pricing is available as:
  - Storage
  - IOPS

\* Check Amazon EBS Pricing page for current pricing for all regions.



# Amazon EBS and Amazon S3

	Amazon EBS 	Amazon S3 	Amazon EFS 
Paradigm	Block storage with file system	Object store	File System
Redundancy	Across multiple servers in an Availability Zone	Across multiple facilities in a Region	Across multiple facilities in a Region
Security	EBS Encryption – Data volumes and Snapshots	Encryption at rest and in transit	Encryption at rest and in transit
Access from the Internet?	No (1)	Yes (2)	No
Typical use case	It is a disk drive	Online storage	Shared file system

(1) Accessible from the Internet if mounted to server and set up as FTP, etc.

(2) Only with proper credentials, unless ACLs are world-readable



# Amazon EC2 Instance Store



- Is local, **complimentary** direct attached block storage.
  - Includes availability, number of disks, and size based on EC2 instance type.
  - Is optimized for up to **365,000** Read IOPS and 315,000 First Write IOPS.
  - Is SSD, NVMe or magnetic.
  - Has **no persistence**. Automatically deletes data when an EC2 instance stops, fails or is terminated.
-



# Amazon EBS vs. Amazon EC2 Instance Store



## **Amazon EBS**

- Data stored on an Amazon EBS volume can persist independently of the life of the instance.
- Storage is persistent.

## **Amazon EC2 Instance Store**

- Data stored on a local instance store persists only as long as the instance is alive.
  - Storage is ephemeral.
-



## Knowledge Check

Q: What AWS service would help support your web application to offload serving static assets and store user uploaded images and video off-instance?

Amazon S3

Q: What acts as an additional layer of security at the subnet level in a VPC?

Network ACLs

True or False: S3 limits the amount you can store.

False

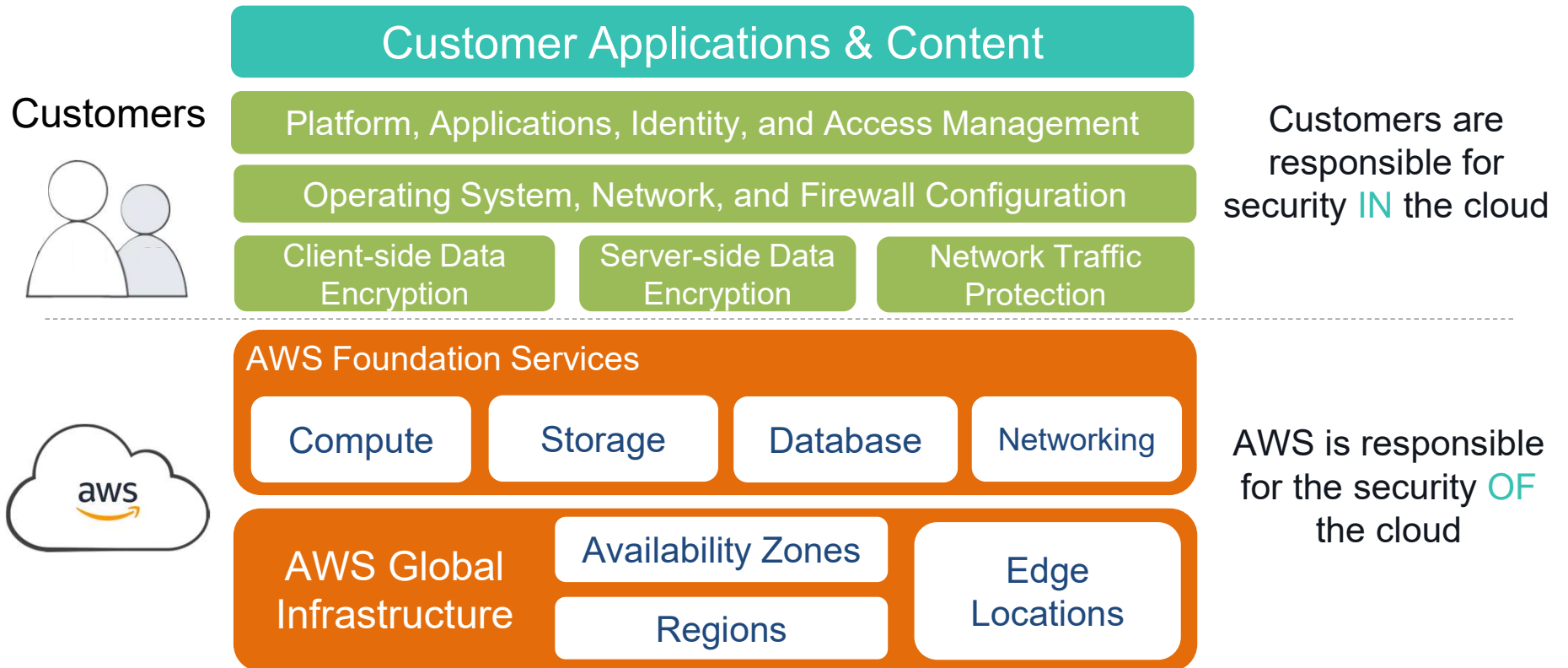
---



# Security, Identity, and Access Management



# AWS Shared Responsibility Model





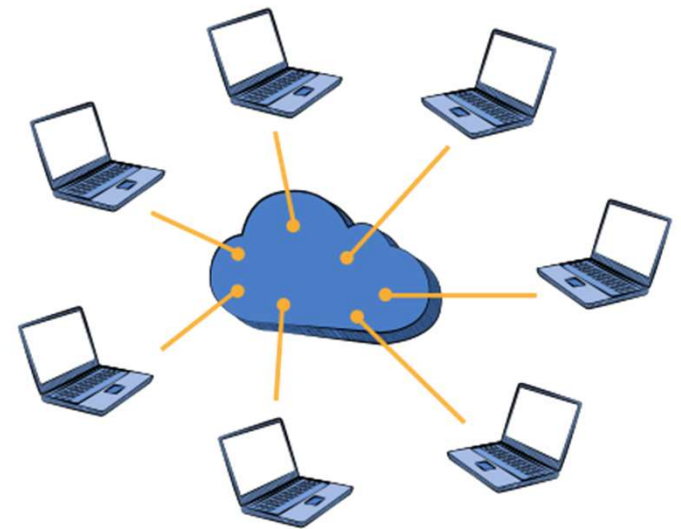
# Physical Security

- 24/7 trained security staff
- AWS data centers in nondescript and undisclosed facilities
- Two-factor authentication for authorized staff
- Authorization for data center access
- Automated change-control process
- Bastion servers that record all access attempts
- Firewall and other boundary devices
- AWS monitoring tools





# Hardware, Software, and Network





# Certifications and Accreditations



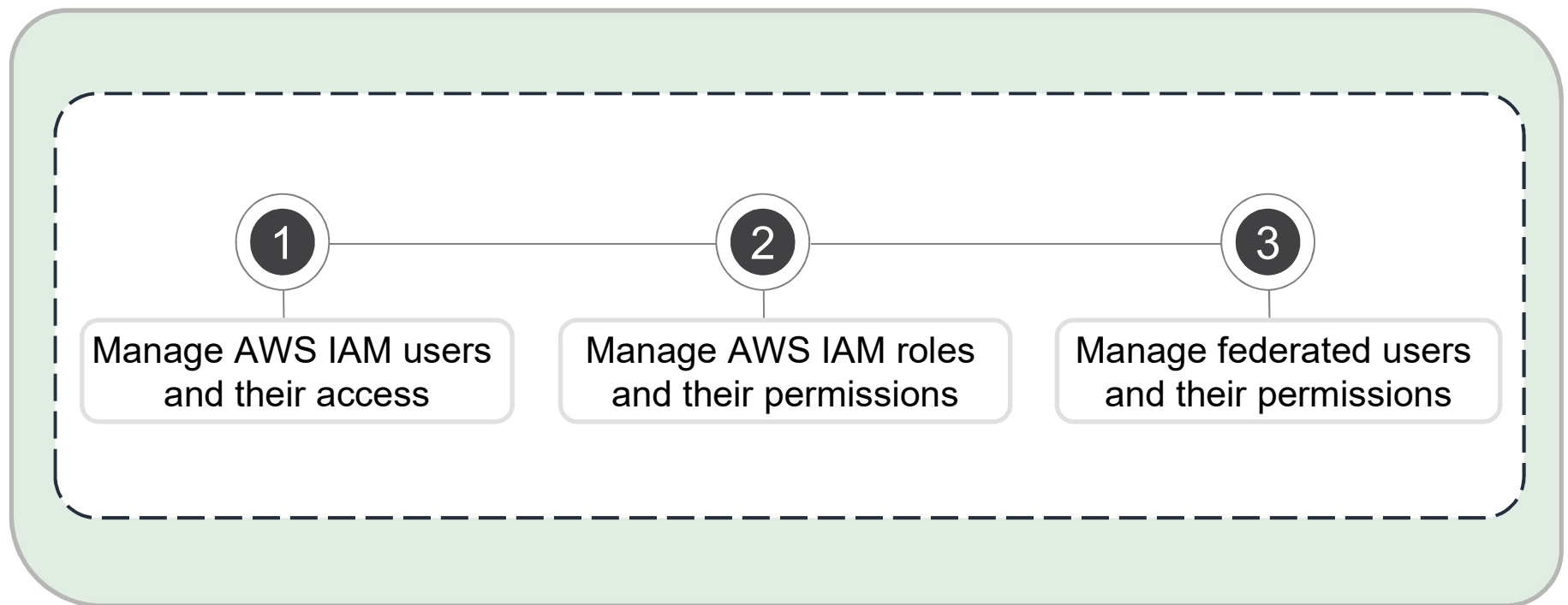
NIST



ISO 9001, ISO 27001, ISO 27017, ISO 27018, IRAP (Australia), MLPS Level 3 (China), MTCS Tier 3 Certification (Singapore) and more ...



# AWS Identity and Access Management (IAM)





# AWS IAM Authentication



- Authentication
- AWS Management Console
  - User Name and Password



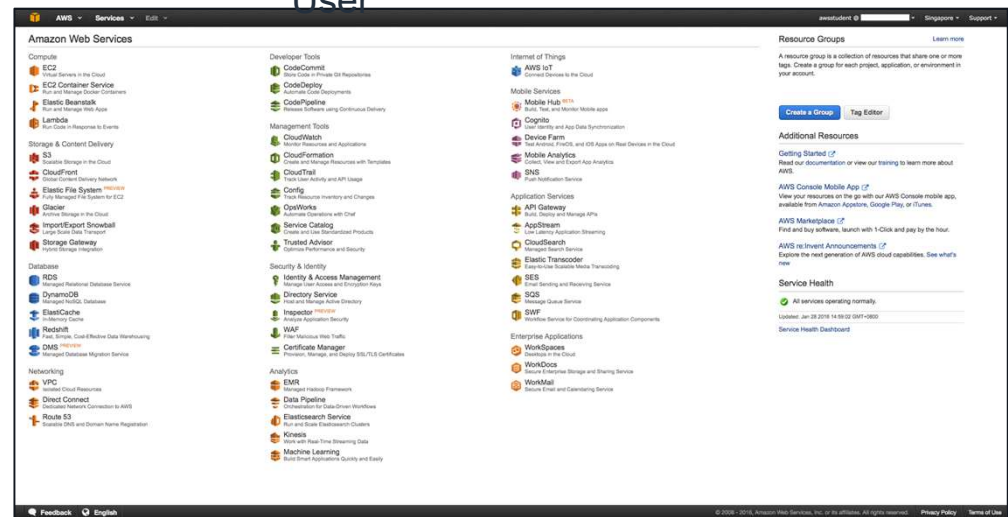
IAM User

**Account:**

**User Name:**

**Password:**

MFA users, enter your code on the next screen.





# AWS IAM Authentication



- Authentication
- AWS CLI or SDK API
  - Access Key and Secret Key



IAM User

Access Key ID: AKIAIOSXXXXXXXXXDORIK  
Secret Access Key: wJalrXUtnFEXXXXXXXXXXXXXXXXXXXXXNIFLEKEY

## AWS CLI

```
:~$ aws configure
AWS Access Key ID [*****O22A]:
AWS Secret Access Key [*****4m8i]:
Default region name [ap-southeast-1]:
Default output format [json]:
```

## AWS SDK & API



Java



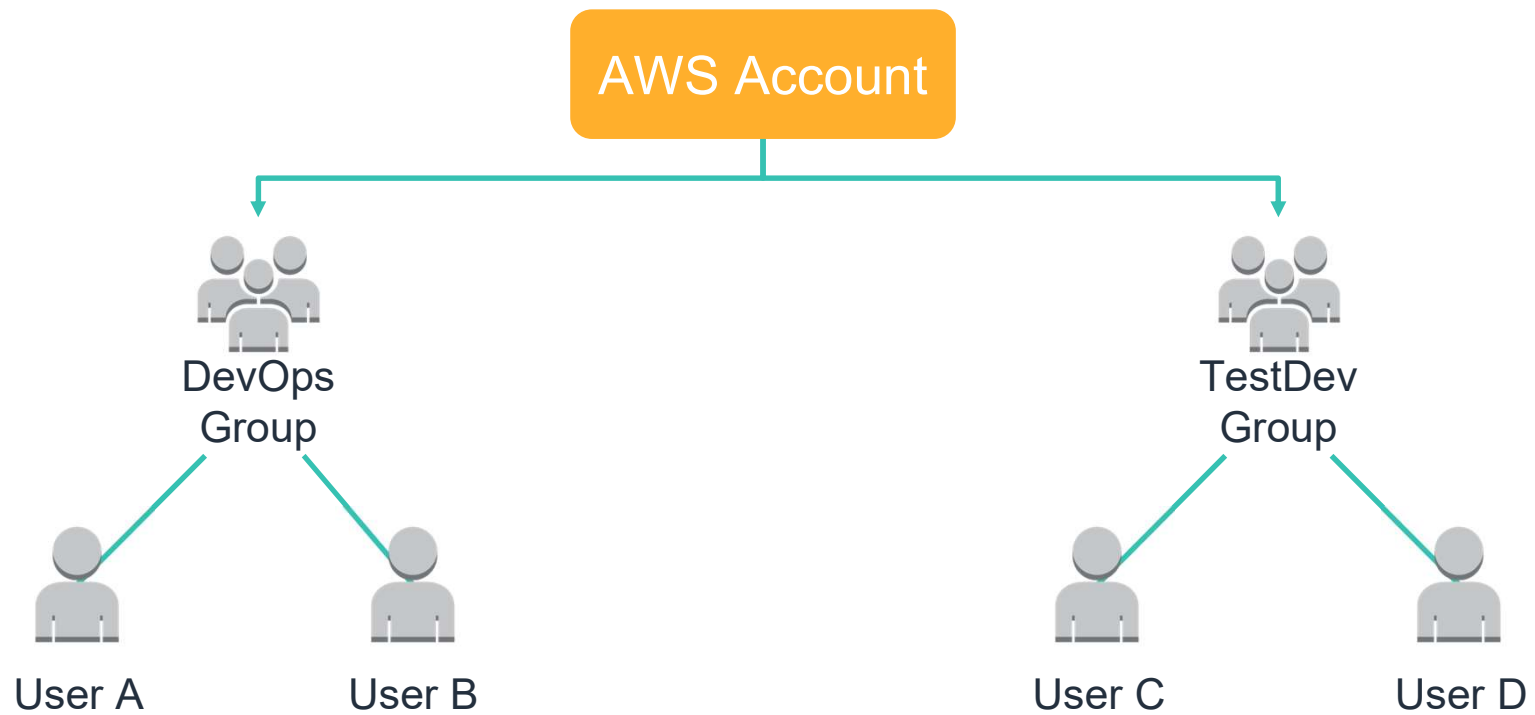
Python



.NET



# AWS IAM User Management - Groups





# AWS IAM Authorization



## Authorization

- Policies:
  - Are JSON documents to describe permissions.
  - Are assigned to users, groups or roles.



IAM User



IAM Group



IAM Roles



# AWS IAM Policy Elements



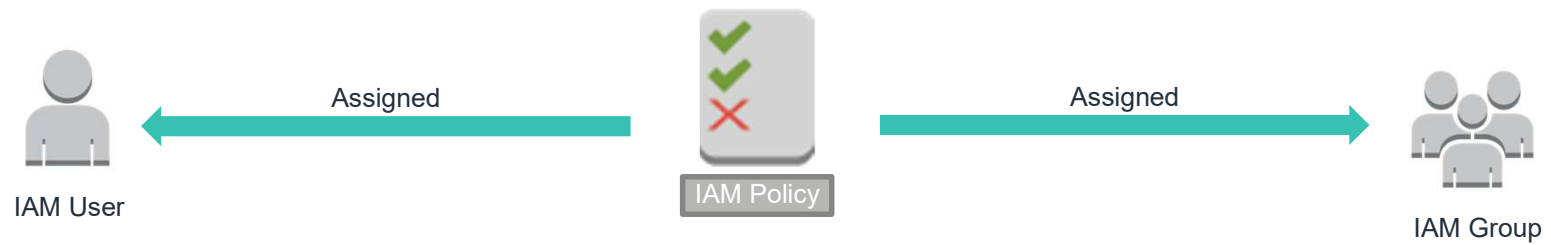
```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt1453690971587",
      "Effect": "Allow",
      "Action": [
        "ec2:Describe*",
        "ec2:StartInstances",
        "ec2:StopInstances"
      ],
      "Resource": "*",
      "Condition": {
        "IpAddress": {
          "aws:SourceIp": "54.64.34.65/32"
        }
      }
    },
    {
      "Sid": "Stmt1453690998327",
      "Effect": "Allow",
      "Action": [
        "s3:GetObject*"
      ],
      "Resource": "arn:aws:s3:::example_bucket/*"
    }
  ]
}
```



IAM Policy

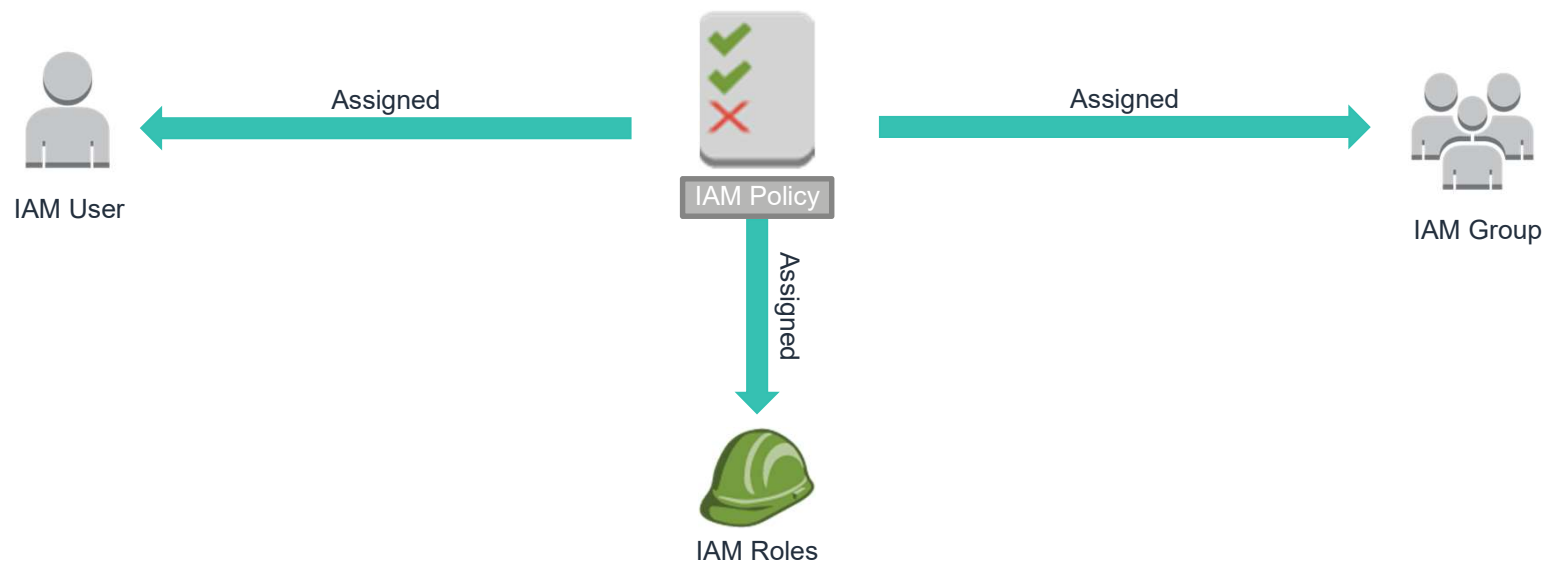


# AWS IAM Policy Assignment





# AWS IAM Policy Assignment





# AWS IAM Roles



- An IAM role uses a policy.
- An IAM role has no associated credentials.
- IAM users, applications, and services may assume IAM roles.

A role is intended to be assumable by anyone who needs it.

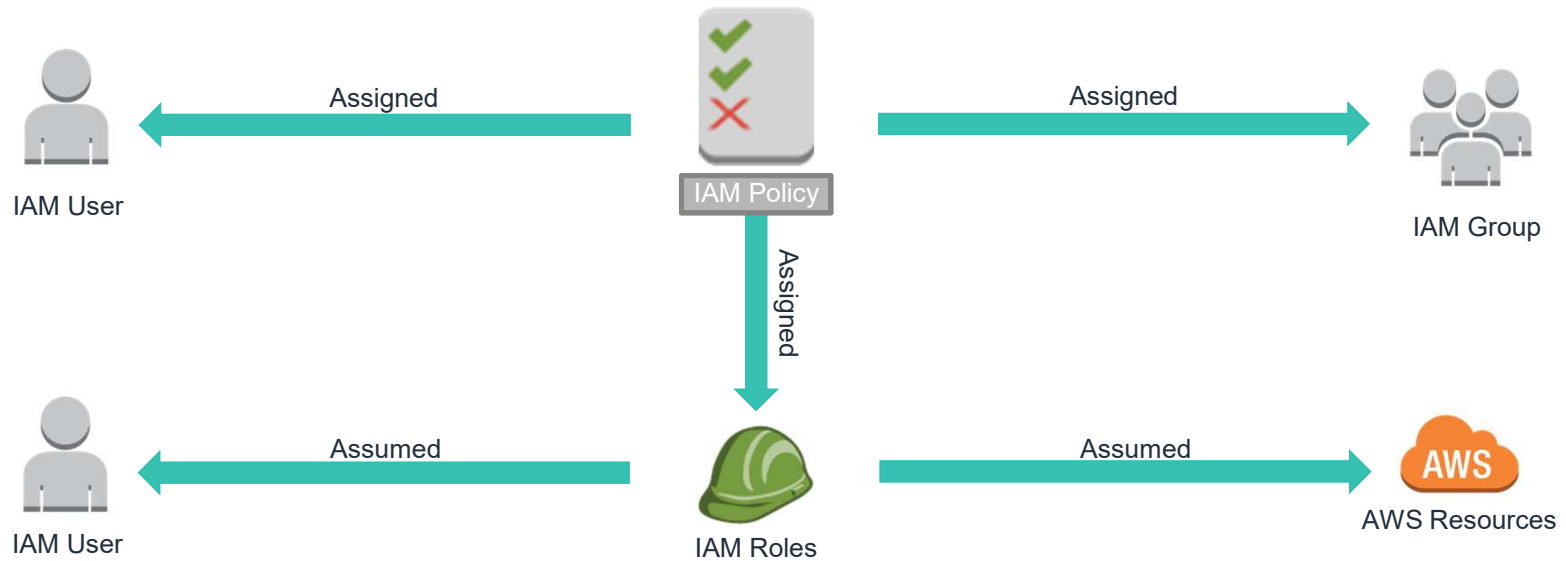
IAM Role



“Hi, my name is...”

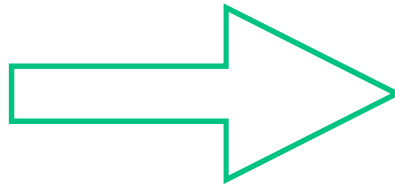


# AWS IAM Policy Assignment





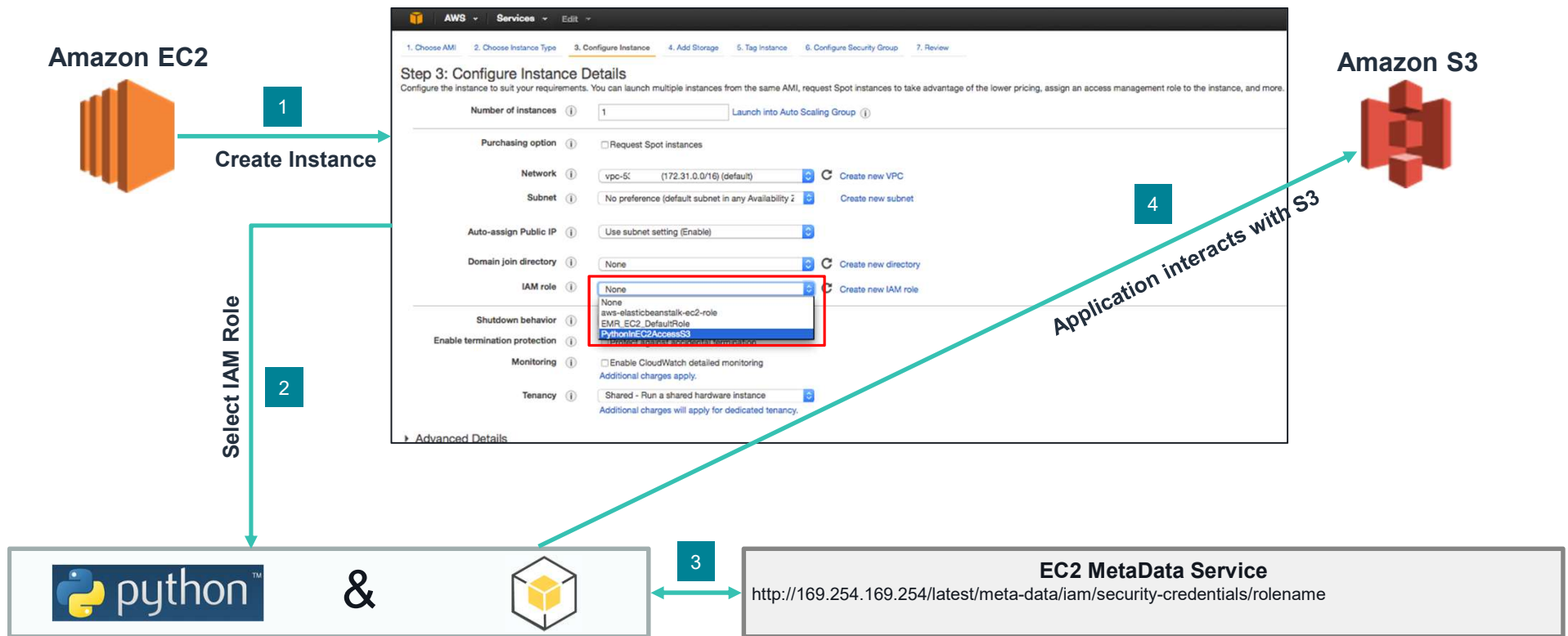
## Example: Application Access to AWS Resources



- Python application hosted on an Amazon EC2 Instance needs to interact with Amazon S3.
- AWS credentials are required:
  - ~~Option 1: Store AWS Credentials on the Amazon EC2 instance.~~
  - Option 2: Securely distribute AWS credentials to AWS Services and Applications.

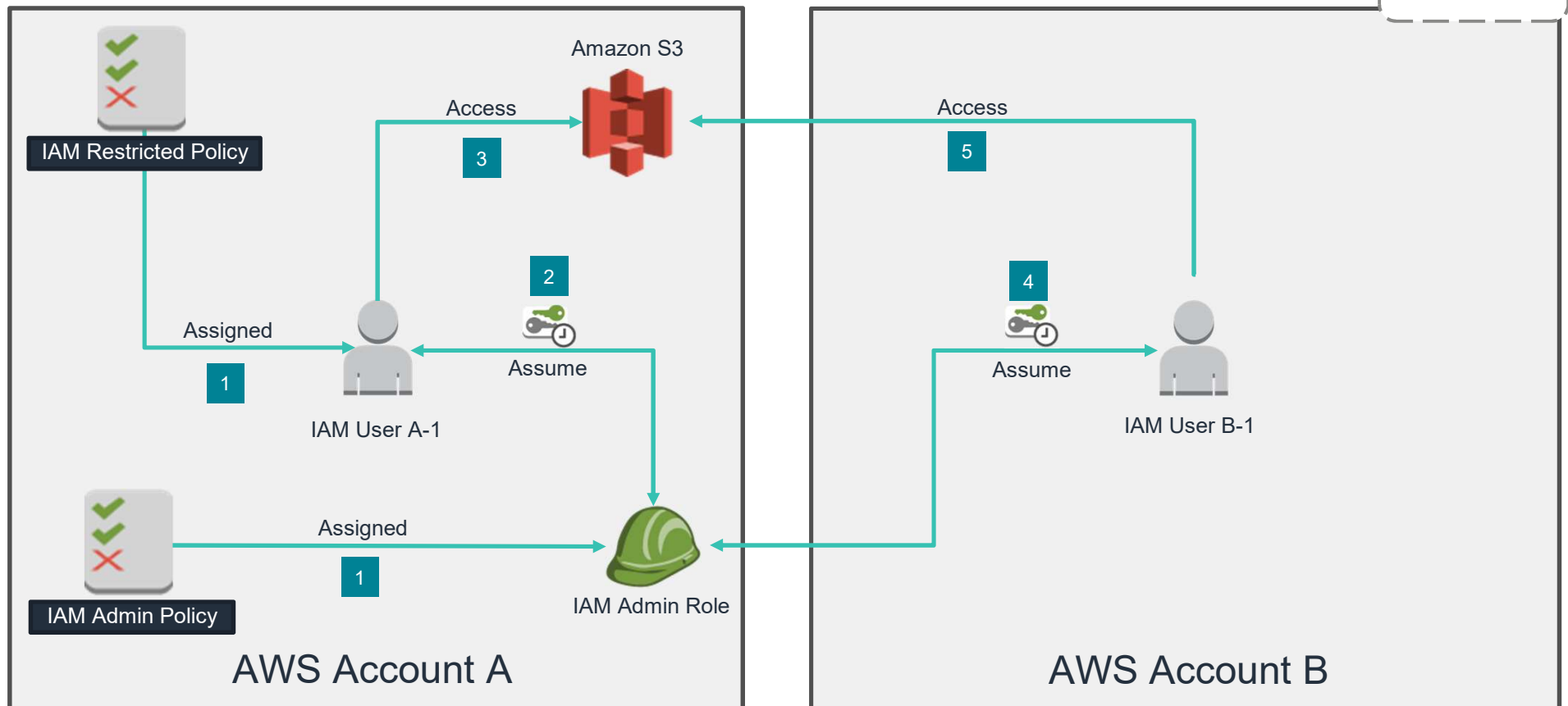


# AWS IAM Roles - Instance Profiles



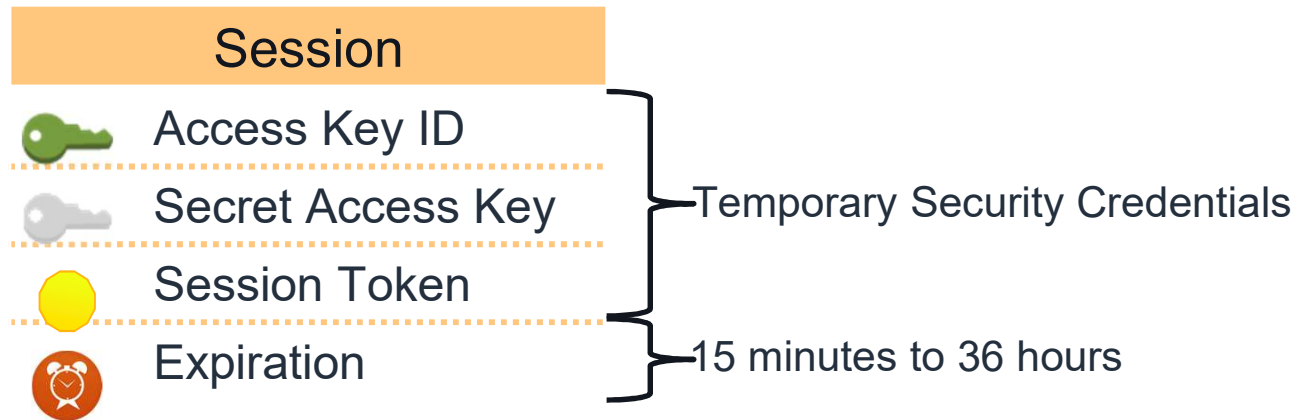


# AWS IAM Roles – Assume Role





# Temporary Security Credentials (AWS STS)



## Use Cases

- Cross account access
- Federation
- Mobile Users
- Key rotation for Amazon EC2-based apps



# Application Authentication





# AWS IAM Authentication and Authorization



## Authentication

- AWS Management Console
  - User Name and Password
- AWS CLI or SDK API
  - Access Key and Secret Key

## Authorization

- Policies



IAM User



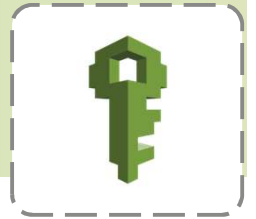
IAM Group



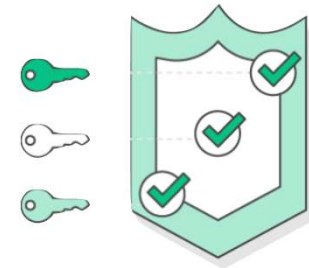
IAM Roles



# AWS IAM Best Practices



- Create individual IAM users.
- Grant least privilege.
- Enable MFA.
- Use roles for applications (instead of sharing credentials).
- Rotate credentials regularly.
- Use policy conditions for extra security.
- Monitor activity in your AWS account.





## AWS IAM Best Practices (cont.)



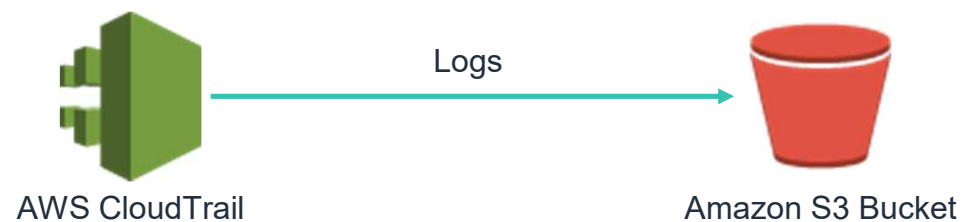
- Use roles for applications that run on Amazon EC2 instances.
  - Delegate by using roles instead of by sharing credentials.
  - Rotate credentials regularly.
  - Remove unnecessary users and credentials.
  - Use policy conditions for extra security.
  - Monitor activity in your AWS account.
-



# AWS CloudTrail



- Records AWS API calls for accounts.
- Delivers log files with information to an Amazon S3 bucket.
- Makes calls using the AWS Management Console, AWS SDKs, AWS CLI and higher-level AWS services.





## Knowledge Check

Q: Your web application needs to read from an Amazon DynamoDB table and get objects from an Amazon S3 bucket. This operation requires AWS credentials and authorization from AWS services. What IAM entity will be used?

Group

Role

Policy

Bucket

User

---





## Databases:

### Amazon Relational Database Service (RDS)

### Amazon DynamoDB



# SQL and NoSQL Databases

	SQL	NoSQL
Data Storage	Rows and Columns	Key-Value
Schemas	Fixed	Dynamic
Querying	Using SQL	Focused on collection of documents
Scalability	Vertical	Horizontal

## SQL

ISBN	Title	Author	Format
9182932465265	Cloud Computing Concepts	Wilson, Joe	Paperback
3142536475869	The Database Guru	Gomez, Maria	eBook

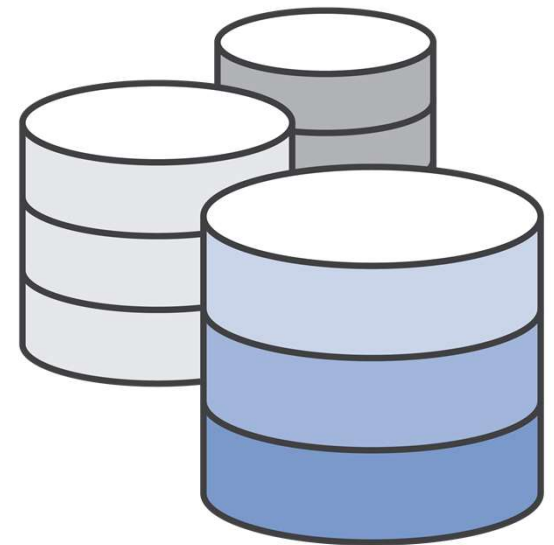
## NoSQL

```
{  
  ISBN: 9182932465265,  
  Title: "Cloud Computing Concepts",  
  Author: "Wilson, Joe",  
  Format: "Paperback"  
}
```



# Data Storage Considerations

- No one size fits all.
- Analyze your data requirements by considering:
  - Data formats
  - Data size
  - Query frequency
  - Data access speed
  - Data retention period





# AWS Managed Database Services

## Relational Databases



Amazon RDS

Aurora

Commercial

Community

MySQL

ORACLE

MySQL

PostgreSQL

Microsoft SQL Server

PostgreSQL

MariaDB



Amazon Redshift

Data Warehouse

## Non-Relational Databases



Amazon  
DynamoDB

Key Value

Document



Amazon  
ElastiCache

In-Memory  
Data Store

redis

MEMCACHED



Amazon  
Neptune

Graph



AWS Database Migration Service



# Amazon Relational Database Service (RDS)



Amazon  
RDS

- Cost-efficient and resizable capacity
- Manages time-consuming database administration tasks
- Access to the full capabilities of Amazon Aurora, MySQL, MariaDB, Microsoft SQL Server, Oracle, and PostgreSQL databases



ORACLE®



Amazon  
Aurora



# How Amazon RDS Backups Work

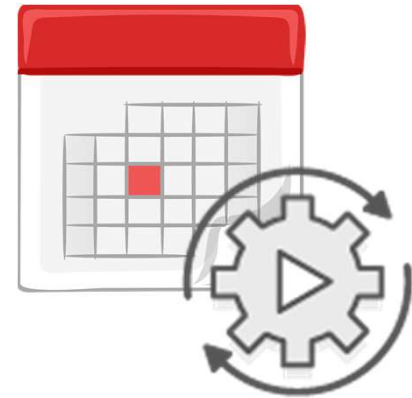


## Automatic Backups:

- Restore your database to a point in time.
- Are enabled by default.
- Let you choose a retention period up to 35 days.

## Manual Snapshots:

- Let you build a new database instance from a snapshot.
- Are initiated by the user.
- Persist until the user deletes them.
- Are stored in Amazon S3.





# Cross-Region Snapshots

- Are a copy of a database snapshot stored in a different AWS Region.
- Provide a backup for disaster recovery.
- Can be used as a base for migration to a different region.





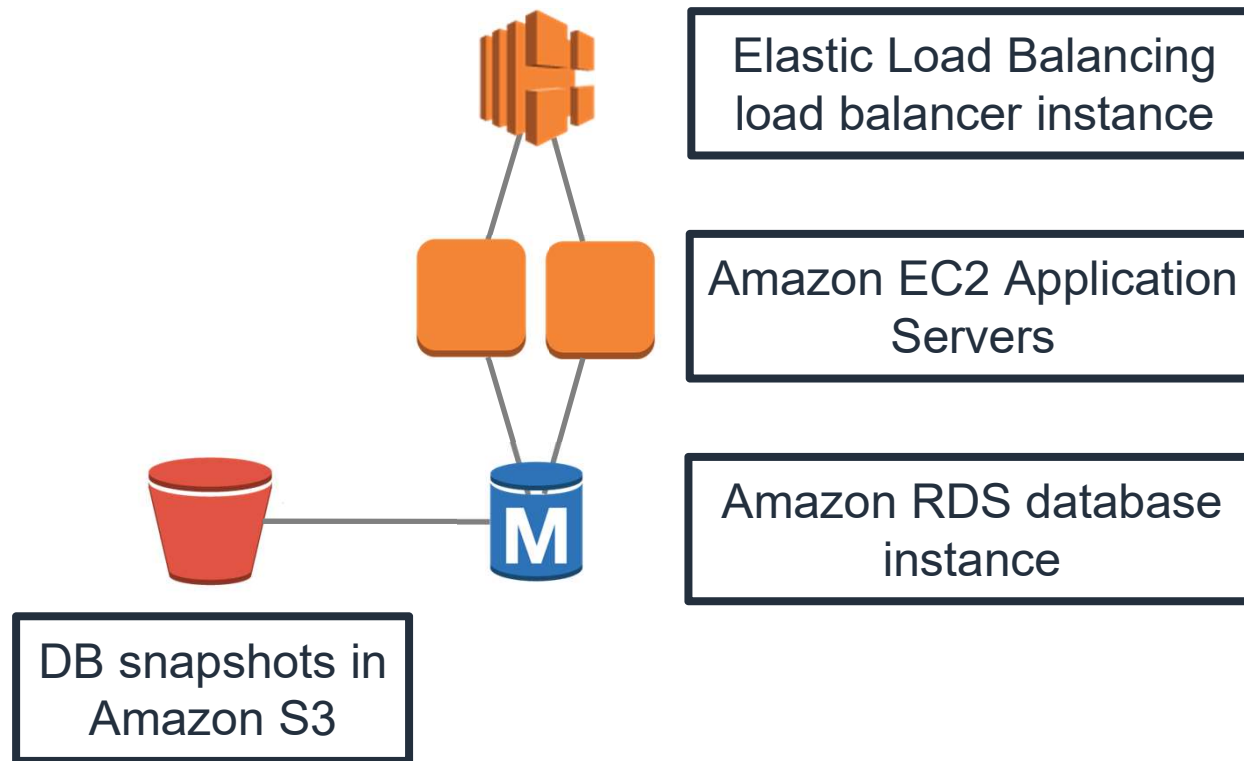
# Amazon RDS Security



- Run your DB instance in an Amazon VPC.
  - Use IAM policies to grant access to Amazon RDS resources.
  - Use security groups.
  - Use Secure Socket Layer (SSL) connections with DB instances (Amazon Aurora, Oracle, MySQL, MariaDB, PostgreSQL, Microsoft SQL Server).
  - Use Amazon RDS encryption to secure your RDS instances and snapshots at rest.
  - Use network encryption and transparent data encryption (TDE) with Oracle DB and Microsoft SQL Server instances.
  - Use the security features of your DB engine to control access to your DB instance.
-



# A Simple Application Architecture

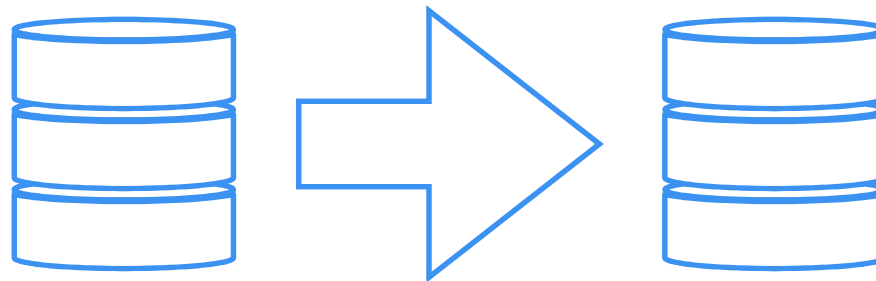




# Multi-AZ RDS Deployment

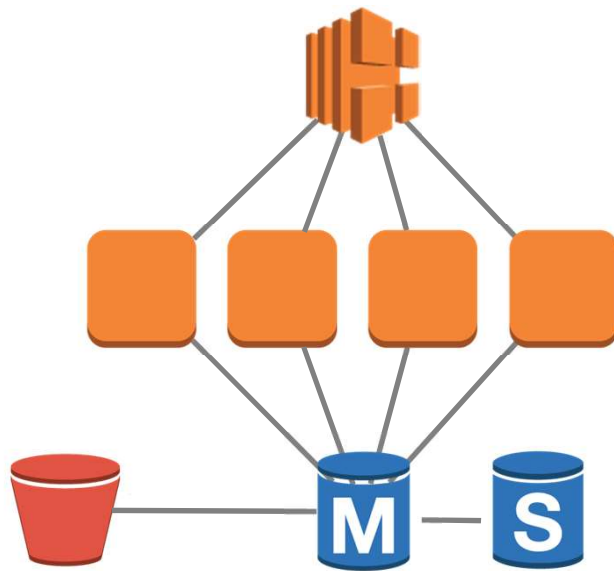


- With Multi-AZ operation, your database is synchronously replicated to another Availability Zone in the same AWS Region.
- Failover to the standby automatically occurs in case of master database failure.
- Planned maintenance is applied first to standby databases.





# A Resilient, Durable Application Architecture



Elastic Load Balancing  
load balancer instance

Application, in Amazon  
EC2 instances

Amazon RDS database instances:  
Master and Multi-AZ standby

DB snapshots in  
Amazon S3



# Amazon RDS Best Practices



- Monitor your memory, CPU, and storage usage.
  - Use Multi-AZ deployments to automatically provision and maintain a synchronous standby in a different Availability Zone.
  - Enable automatic backups.
  - Set the backup window to occur during the daily low in WriteIOPS.
  - To increase the I/O capacity of a DB instance:
    - Migrate to a DB instance class with high I/O capacity.
    - Convert from standard storage to provisioned IOPS storage and use a DB instance class optimized for provisioned IOPS.
    - Provision additional throughput capacity (if using provisioned IOPS storage).
  - If your client application is caching the DNS data of your DB instances, set a TTL of less than 30 seconds.
  - Test failover for your DB instance.
-



# Amazon DynamoDB

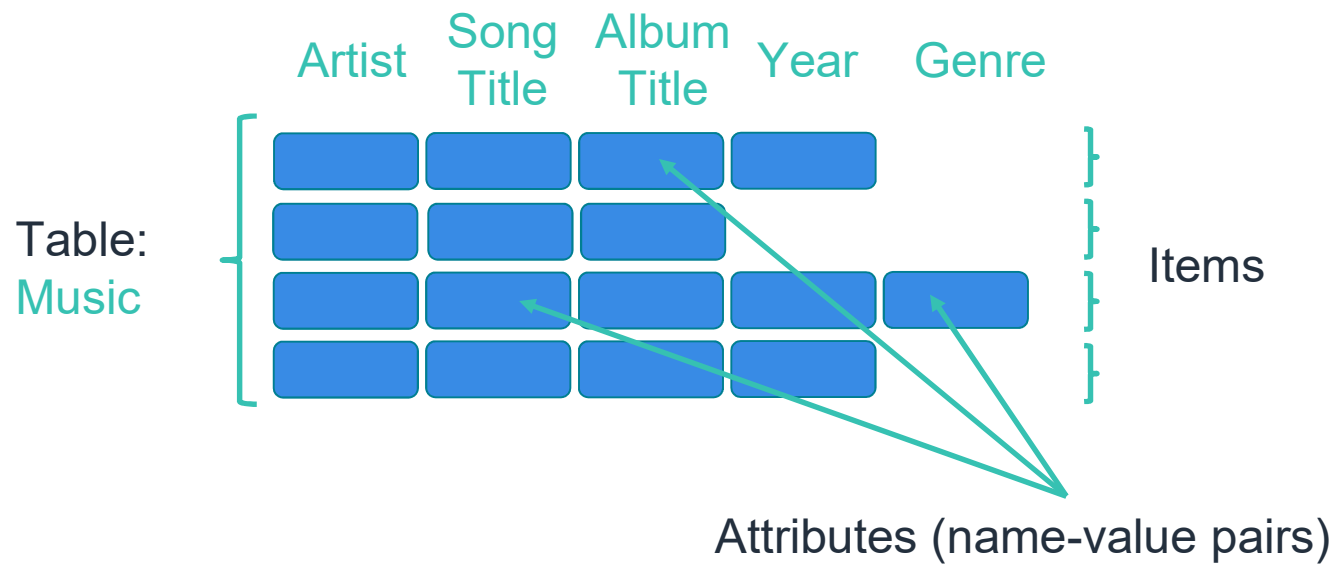


Amazon  
DynamoDB

- Allows you to store any amount of data with **no limits**.
- Provides fast, predictable performance using SSDs.
- Provision and change the request capacity needed for each table.
- Is a **fully managed, NoSQL** database service.



# DynamoDB Data Model





# Primary Keys



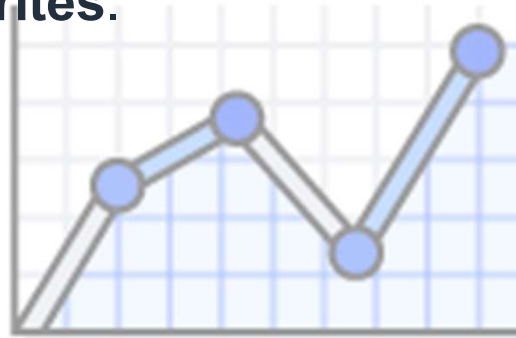
(DynamoDB maintains a sorted index for both keys)



# Provisioned Throughput



You specify **how much** provisioned throughput capacity you need for **reads and writes**.



Amazon DynamoDB allocates the necessary machine resources to **meet your needs**.



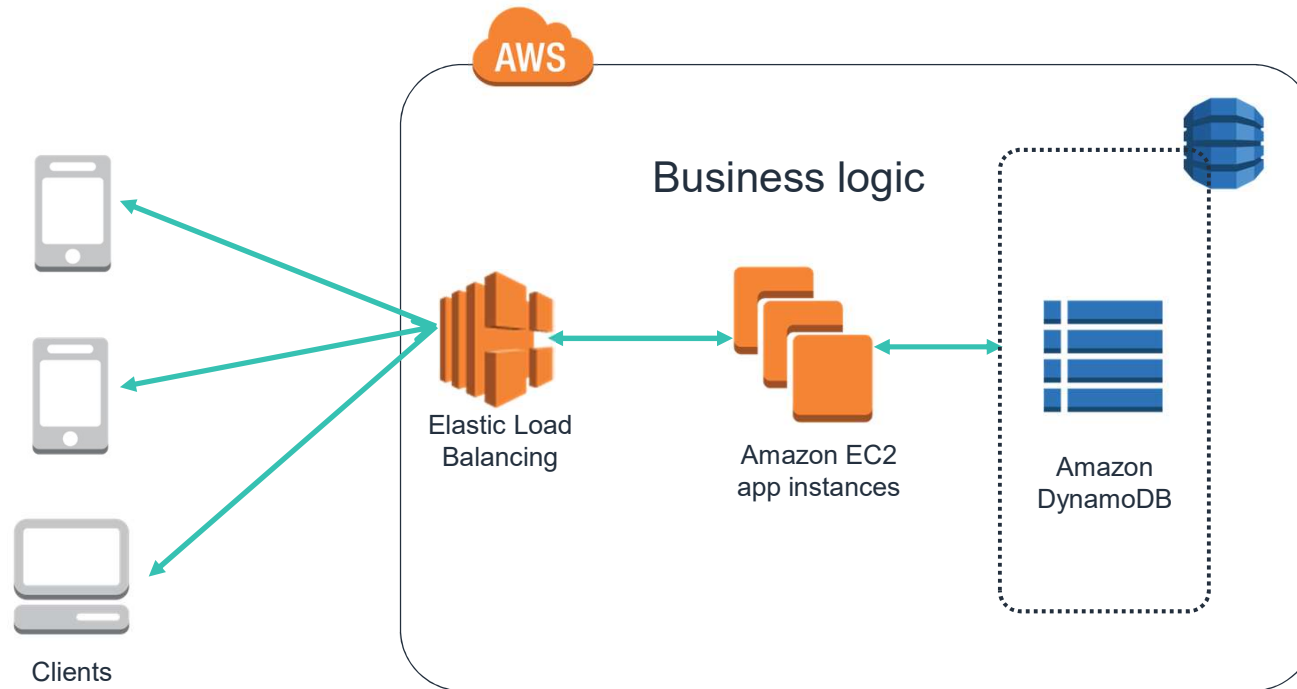
# Supported Operations



- Query:
    - Query a table using the partition key and an optional sort key filter.
    - If the table has a secondary index, query using its key.
    - It is the most efficient way to retrieve items from a table or secondary index.
  - Scan:
    - You can scan a table or secondary index.
    - Scan reads every item – slower than querying.
  - You can use conditional expressions in both Query and Scan operations.
-



# Simple Application Architecture





## Knowledge Check

Q: What are the basic building blocks of Amazon Relational Database Service (RDS)?

**DB Instances**

True or False: Amazon DynamoDB allows you to store any amount of data with no limits.

**True**

True or False: Scan is the most efficient way to retrieve items from a DynamoDB table.

**False**

Q: You are creating a resilient, durable application using Amazon RDS. In addition to Amazon RDS's automatic backups, what feature should you use to ensure that your backups are durable retained?

**Manual Snapshots**

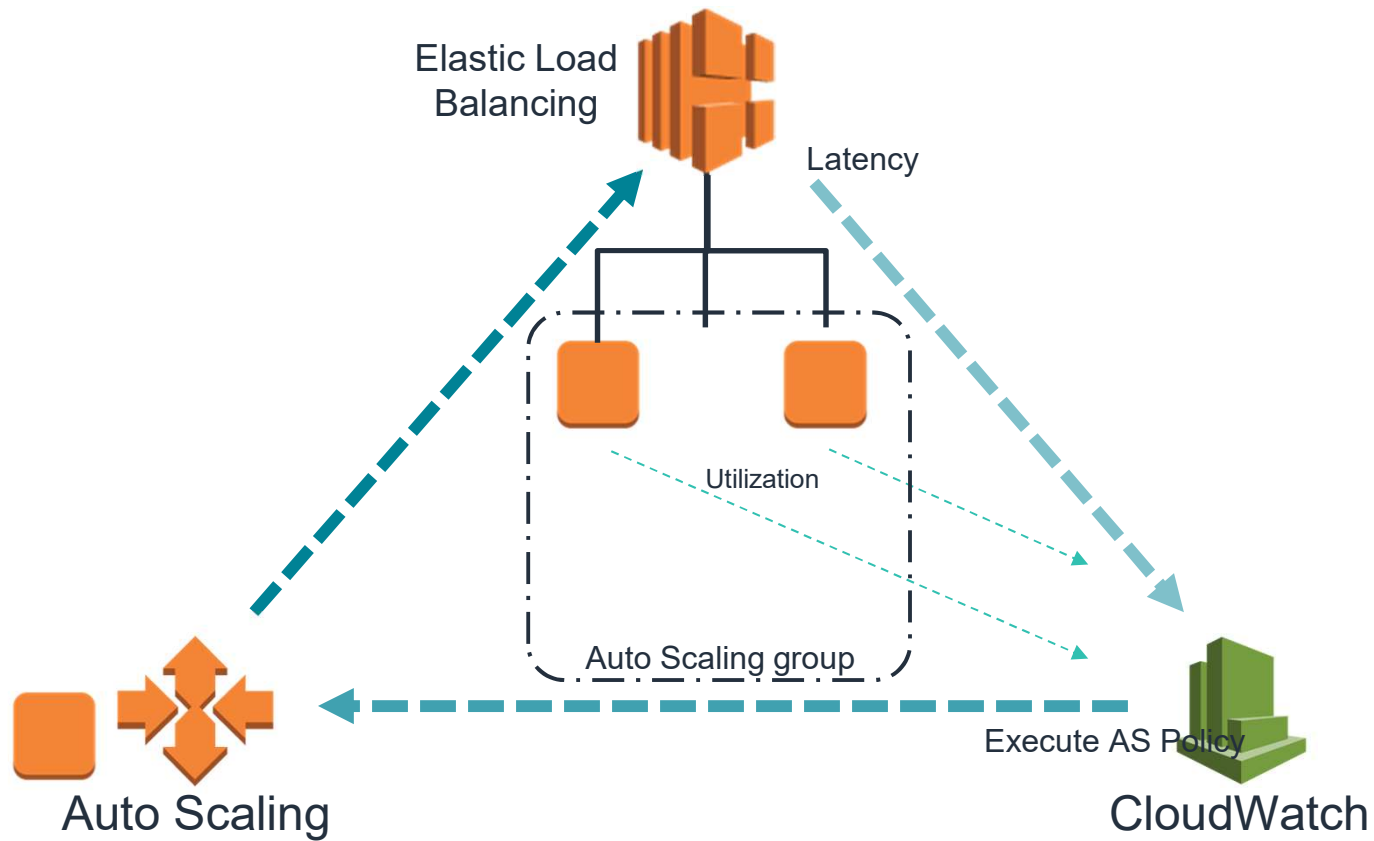
---



# AWS Elasticity and Management Tools



# Triad of Services





# Elastic Load Balancing

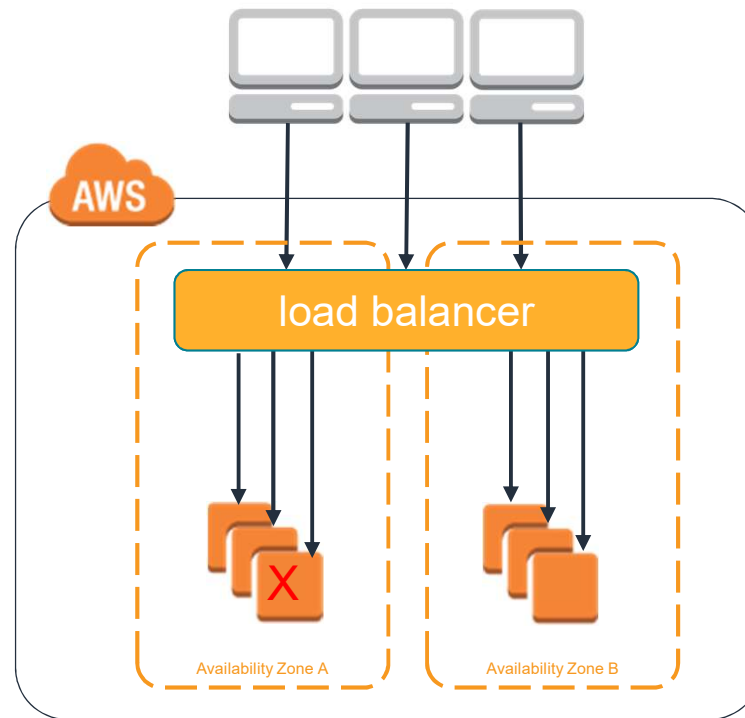


Elastic Load  
Balancing

- **Distributes traffic** across multiple EC2 instances, in multiple Availability Zones
- Supports **health checks** to detect unhealthy Amazon EC2 instances
- Supports the routing and load balancing of **HTTP**, **HTTPS**, **SSL**, and **TCP** traffic to Amazon EC2 instances



# Classic Load Balancer - How It Works

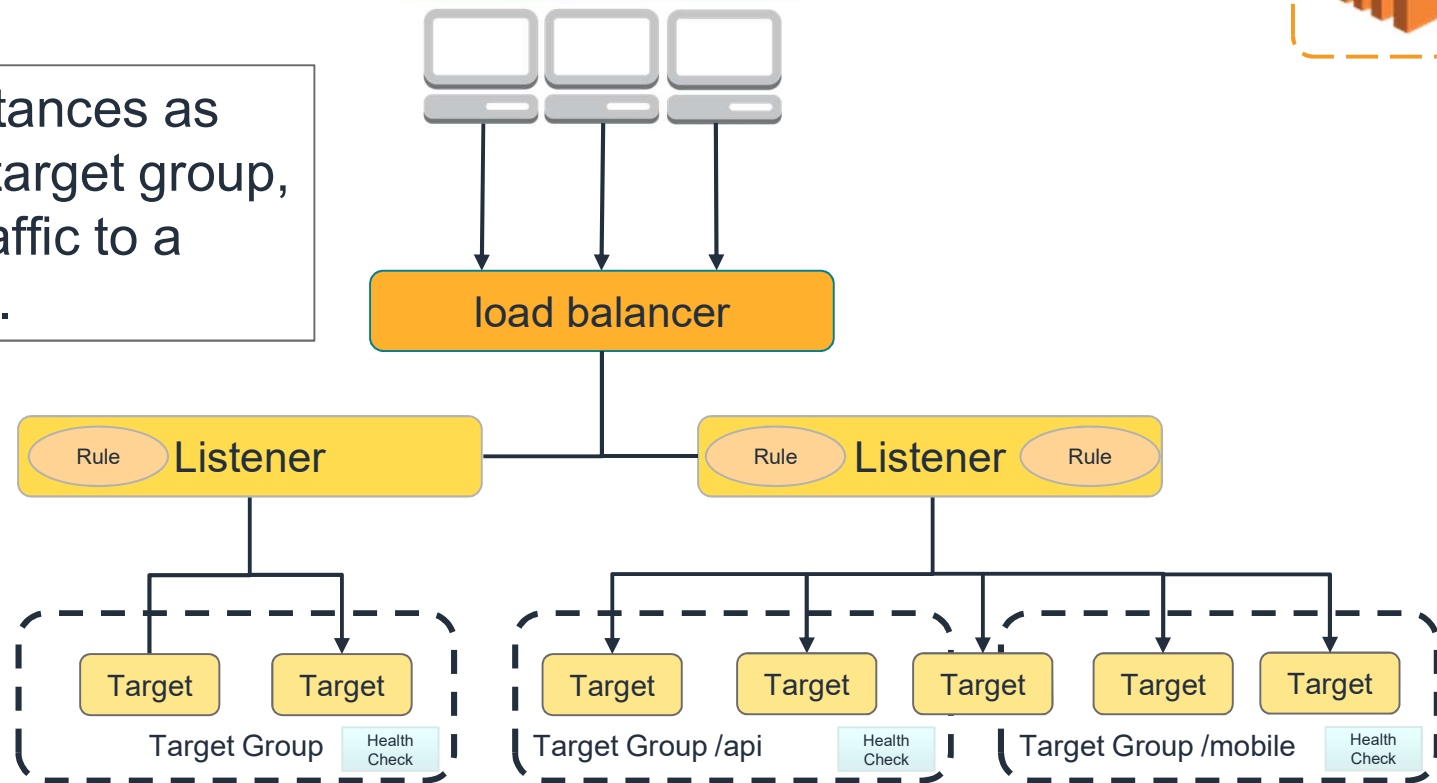




# Application Load Balancer – How It Works

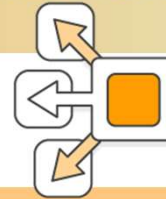


Register instances as targets in a target group, and route traffic to a target group.





# Comparison of Elastic Load Balancing



Feature	Application Load Balancer	Network Load Balancer	Classic Load Balancer
Protocols	HTTP, HTTPS	TCP	TCP, SSL, HTTP, HTTPS
Platforms	VPC	VPC	EC2-Classic, VPC
Load Balancing to multiple ports on the same instance	✓	✓	
Path-Based Routing	✓		
Static IP		✓	



# Load Balancer Comparison



Classic Load Balancer benefits include support for:

- EC2-Classic.
- VPC.
- TCP and SSL listeners.
- Sticky sessions.

ALB benefits include support for:

- Path-based routing.
  - Routing requests to multiple services on a single EC2 instance.
  - Containerized applications.
  - Monitoring the health of each service independently.
-



# Amazon CloudWatch



Amazon  
CloudWatch

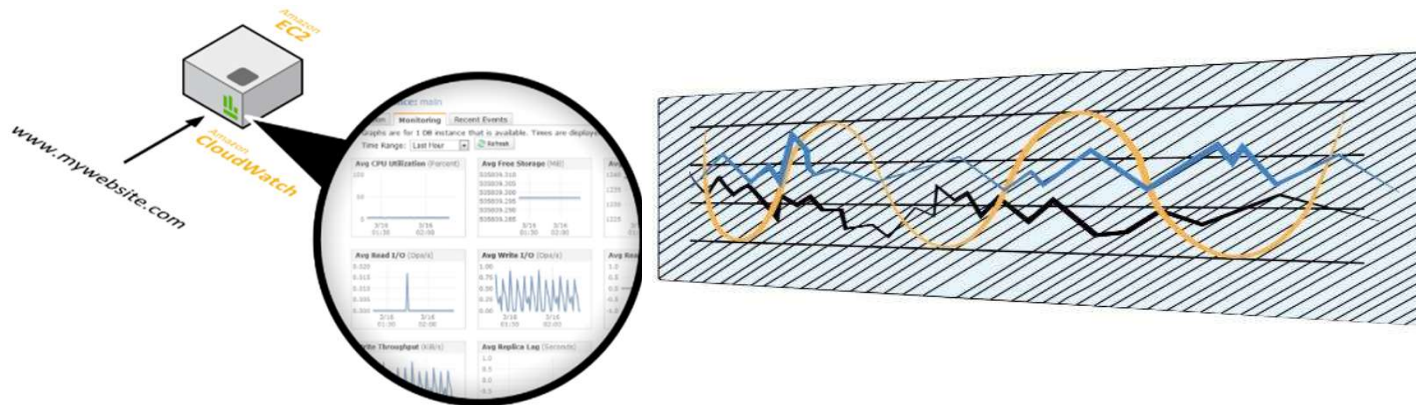
- A monitoring service for AWS cloud resources and the applications you run on AWS
- Visibility into resource utilization, operational performance, and overall demand patterns
- Custom application - specific metrics of your own
- Accessible via AWS Management Console, APIs, SDK, or CLI



# Amazon CloudWatch Facts

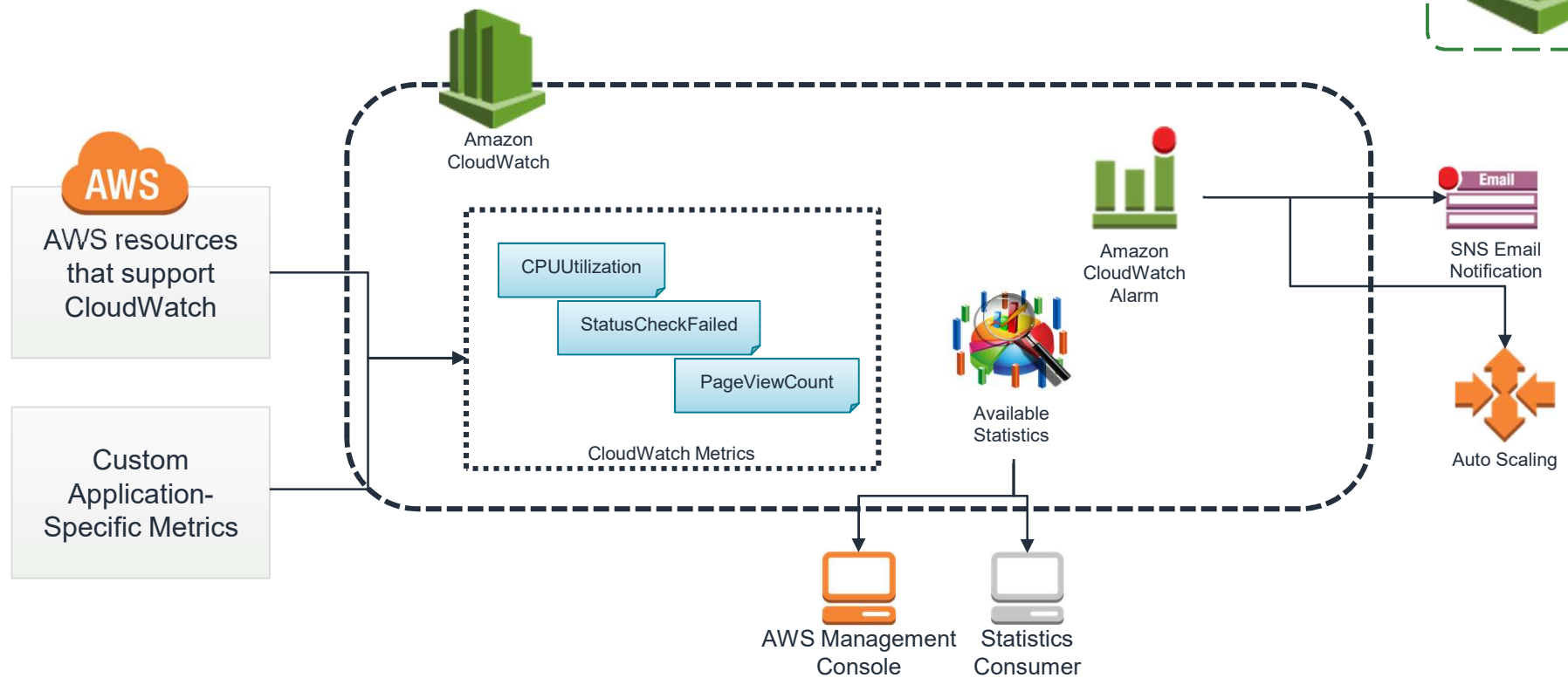


- Monitor other AWS resources
  - View graphics and statistics
- Set Alarms





# Amazon CloudWatch Architecture





# CloudWatch Metrics Examples



## CloudWatch Metrics by Category

Your CloudWatch metric summary has loaded. Total metrics: 97

EBS Metrics: 24

EC2 Metrics: 38

S3 Metrics: 18

SNS Metrics: 3

SQS Metrics: 14

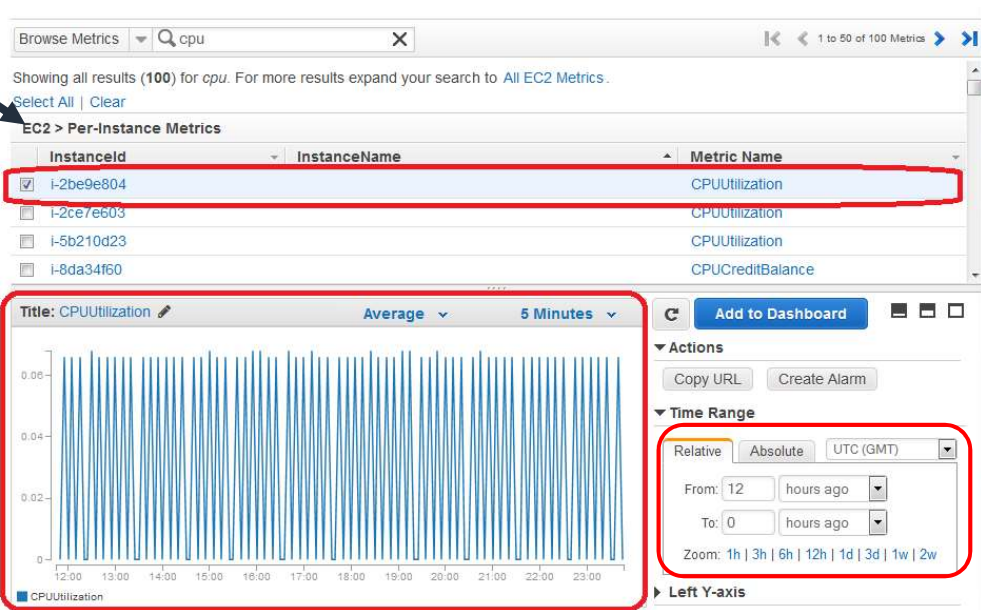
Per-Volume Metrics: 24

Per-Instance Metrics: 38

Storage Metrics: 18

Topic Metrics: 3

Queue Metrics: 14





# Auto Scaling



Auto  
Scaling

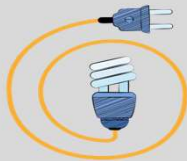
- **Scale** your Amazon EC2 **capacity** automatically
- Well-suited for applications that experience **variability** in **usage**
- Available at **no additional charge**



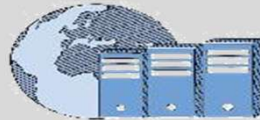
# Auto Scaling Benefits



Better Fault  
Tolerance



Better  
Availability



Better Cost  
Management





# Launch Configurations



- A launch configuration is a template that an Auto Scaling group uses to launch EC2 instances.
- When you create a launch configuration, you can specify:
  - AMI ID
  - Instance type
  - Key pair
  - Security groups
  - Block device mapping
  - User data

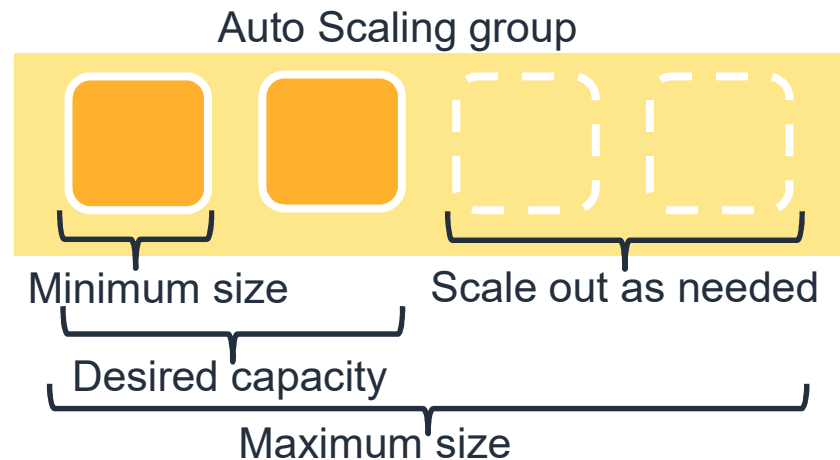




# Auto Scaling Groups



- Contain a collection of EC2 instances that share similar characteristics.
- Instances in an Auto Scaling group are treated as a logical grouping for the purpose of instance scaling and management.





## Knowledge Check

True or False: Auto Scaling helps you ensure that you have the correct number of EC2 instances available to handle the load for your application.

True

Q: What feature would you use with an auto scaling policy to determine when your auto scaling group should scale out/in?

Amazon CloudWatch alarms

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

Application Load Balancer

---

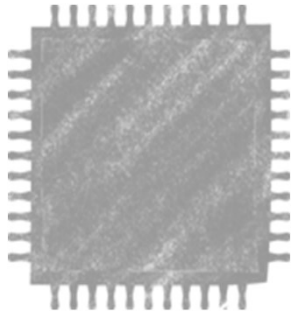


# Serverless AWS Lambda & Amazon API Gateway



# AWS Lambda - Serverless compute

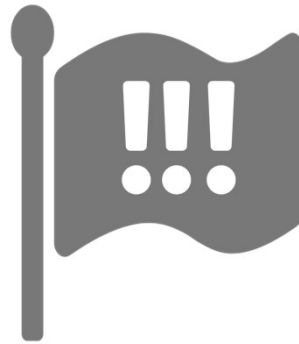
Compute service



Run arbitrary code  
without managing  
servers.



Event driven



Code only runs  
when it needs to  
run.



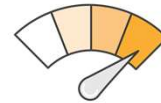


# Using AWS Lambda

## Bring your own code



- Java, Node.js, C#, Python and Go
- Bring your own libraries (even native ones)



## Simple resource model

- Select power rating from 128MB to 3.0 GB
- CPU and network allocated proportionately

## Flexible use



- Synchronous or asynchronous
- Integrated with other AWS services



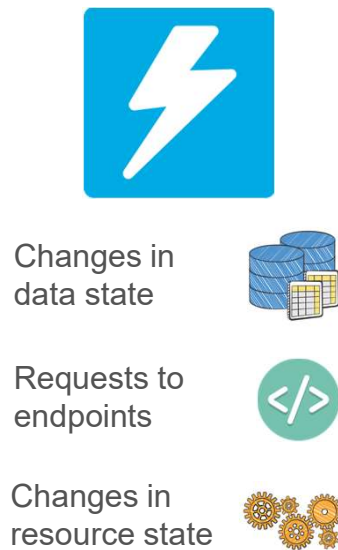
## Flexible authorization

- Securely grant access to resources and VPCs
- Fine-grained control for invoking your functions



# Application components for serverless apps

## EVENT SOURCE



## FUNCTION



## SERVICES (ANYTHING)





# Event sources that integrate with AWS Lambda

## DATA STORES



Amazon  
S3



Amazon  
DynamoDB

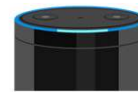


Amazon  
Kinesis



Amazon  
Cognito

## ENDPOINTS



Amazon  
Alexa



Amazon API  
Gateway



AWS IoT

## REPOSITORIES



AWS  
CloudFormation



AWS  
CloudTrail



Amazon  
CloudWatch

## EVENT/MESSAGE SERVICES



Amazon  
SES



Amazon  
SNS

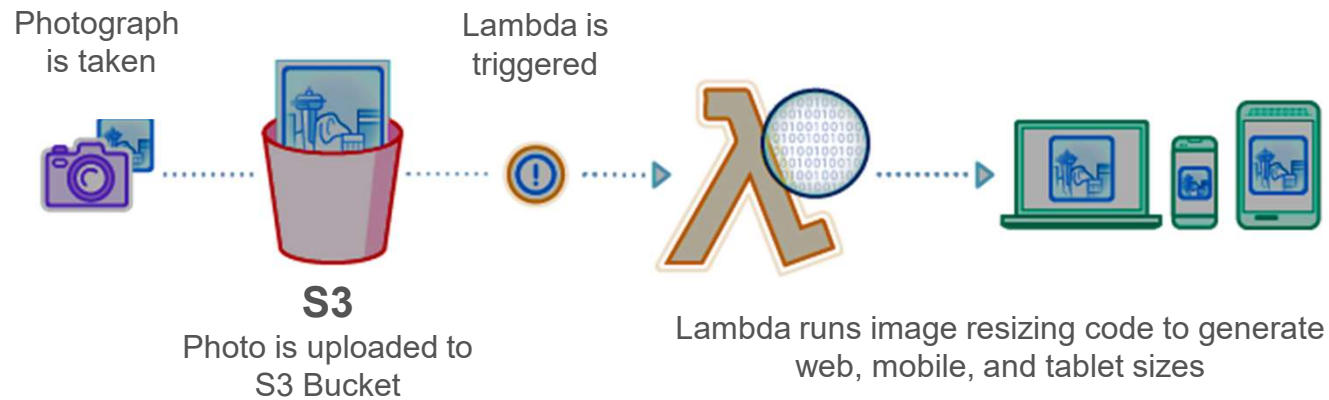


Cron events

*... and the list will continue to grow!*



# Use case: Image Thumbnail creation





# Amazon API Gateway

Allows you to **create APIs** that act as "front doors" for your applications to access data, business logic, or functionality from your **back-end services**.

**Fully managed** and handles all tasks involved in accepting and processing up to **hundreds of thousands of concurrent API calls**.



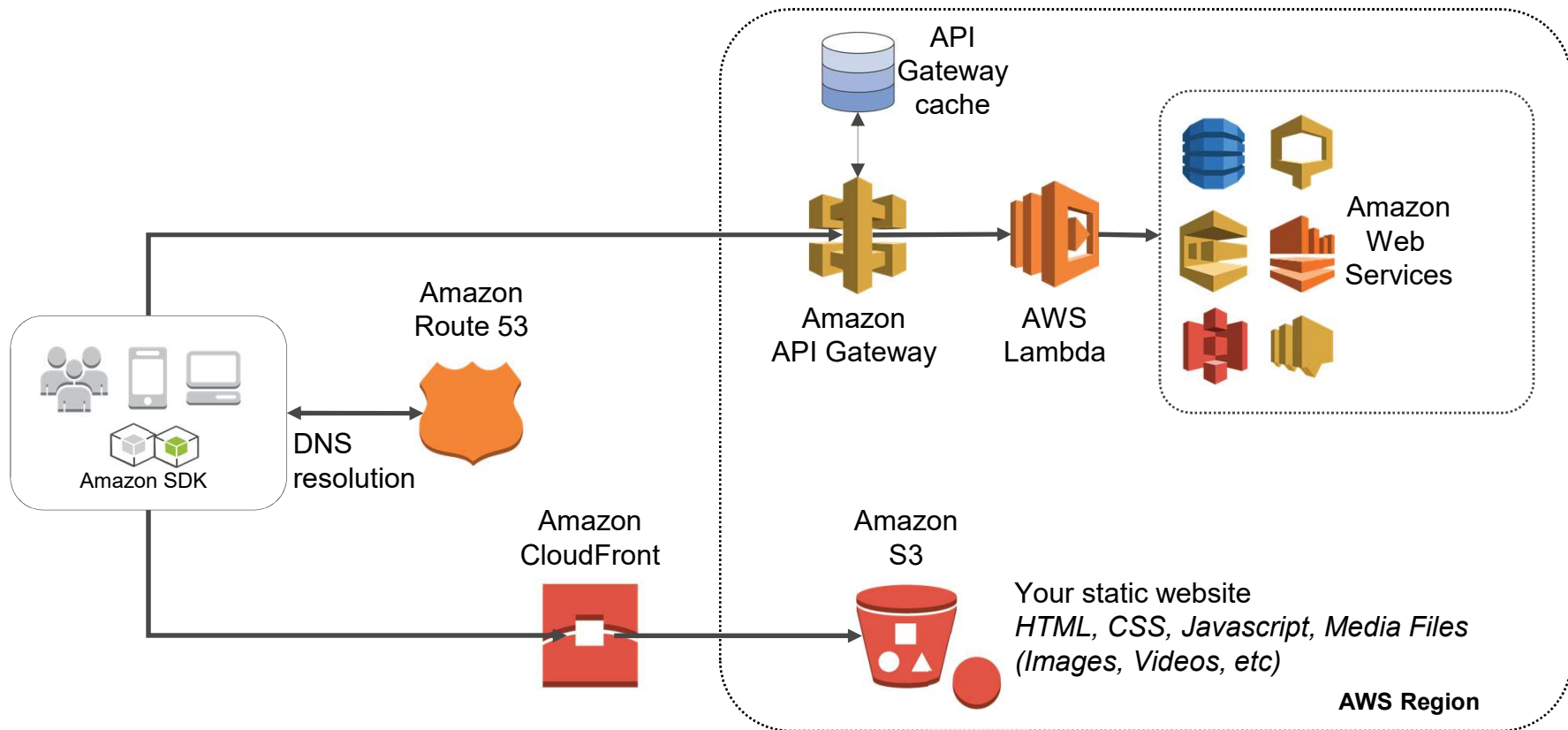


# Benefits of Amazon API Gateway

- Managed cache to store API responses.
- Reduced latency and Distributed Denial of Service (DDoS) protection through Amazon CloudFront.
- SDK generation for iOS, Android, and JavaScript.
- OpenAPI Specification (Swagger) support.
- Request/response data transformation.

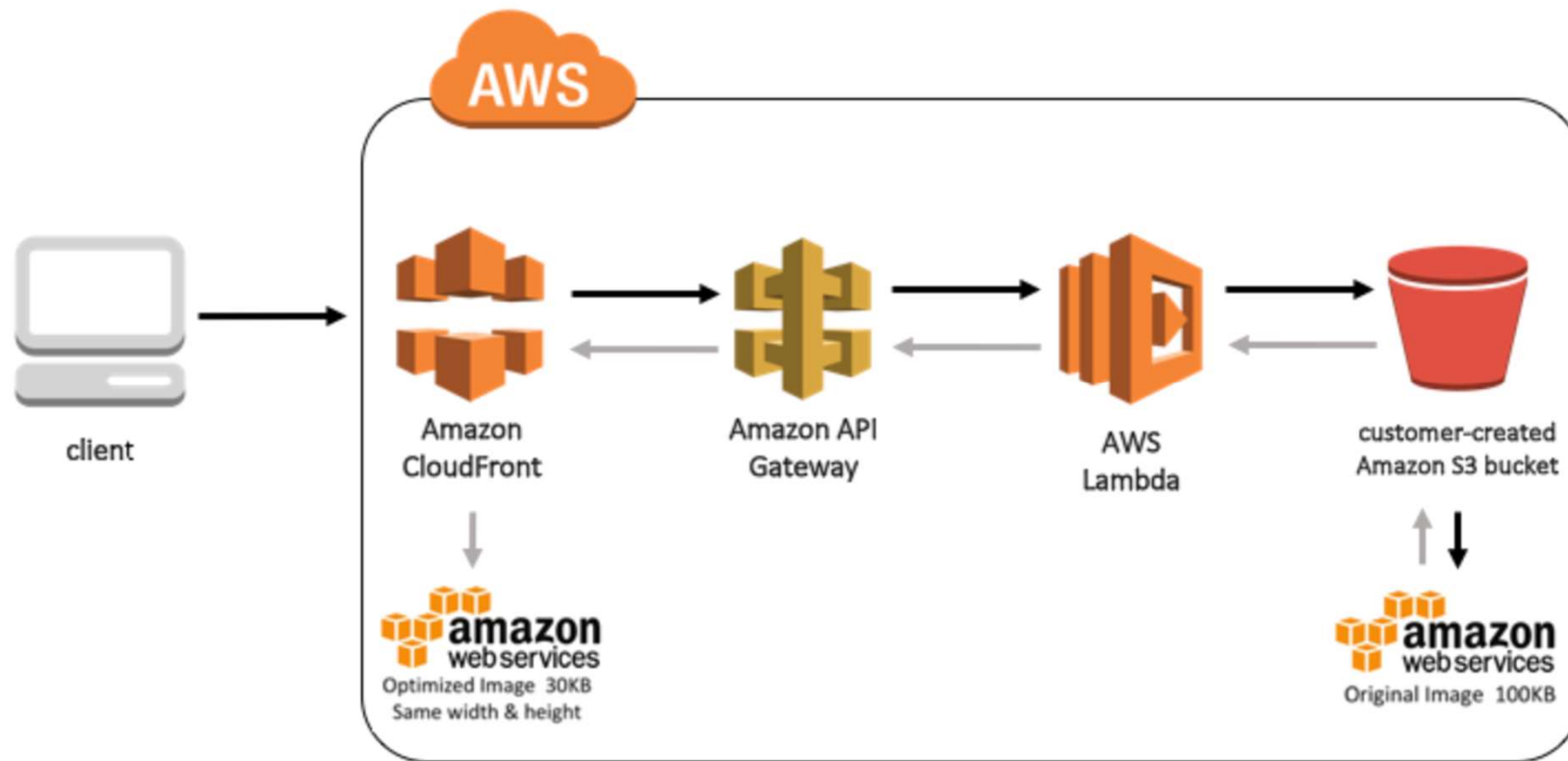






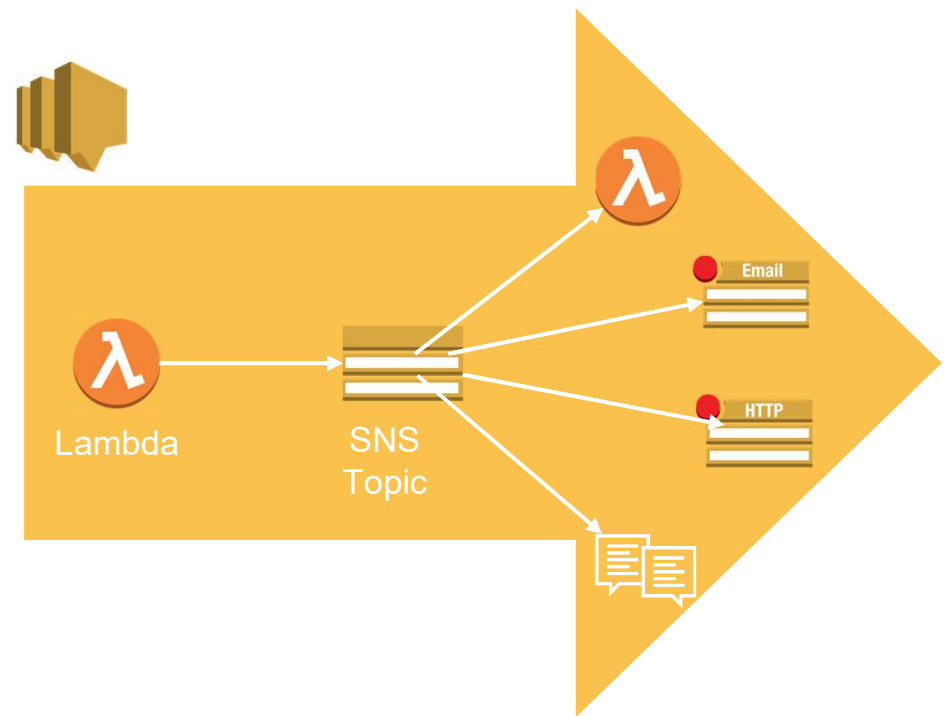
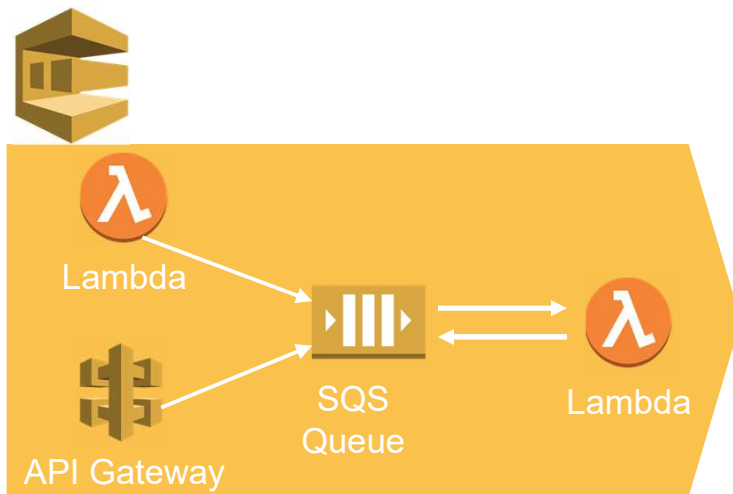


# Serverless Architecture Using API Gateway





# Decoupled Communication





## Knowledge Check

Q: I want to start processing an object as soon as available (for example, when an object is uploaded to S3). Which service can I use?

**AWS Lambda**

Q: I want to create backends for my IoT devices that use REST APIs, which service can I use?

**AWS Lambda**

Q: I have messages sent to a queue processed by an EC2 Instance. A message arrives every 10 minutes. Which service can replace my EC2 Instance?

**AWS Lambda**

Q: I want to....





# Wrap-Up



# What we've talked about today...



Amazon  
EC2



Amazon  
VPC



AWS  
Lambda



Elastic Load  
Balancing



Amazon  
Glacier



Amazon  
S3



IAM



Amazon  
RDS



Amazon  
DynamoDB



Amazon  
CloudWatch



Amazon  
SNS



Amazon  
SQS

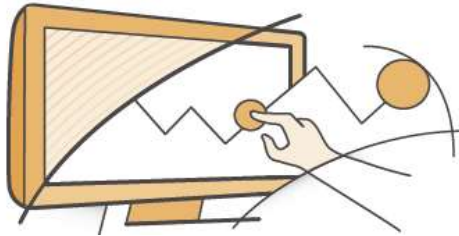


Amazon API  
Gateway



# Expand Your Cloud Skills with AWS

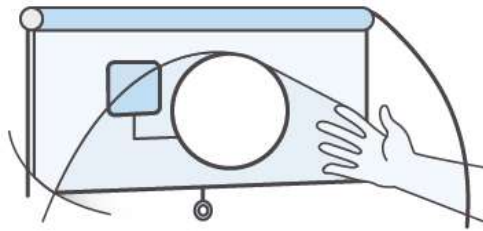
## Online videos and labs



Start working with an AWS service in minutes with free online instructional videos and labs

[aws.amazon.com/training/self-paced-labs](https://aws.amazon.com/training/self-paced-labs)

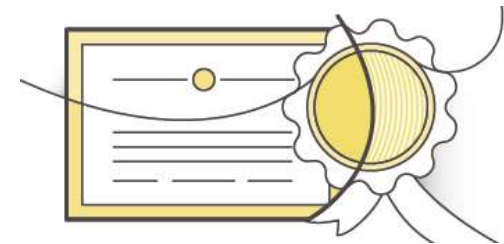
## Instructor-led courses



Learn how to design, deploy, and operate highly available, cost-effective, and secure applications on AWS

[aws.amazon.com/training](https://aws.amazon.com/training)

## Certification



Validate your proven technical expertise with the AWS platform and gain recognition for your skills

[aws.amazon.com/certification](https://aws.amazon.com/certification)



# AWS Courses

## Courses for beginners: build foundational knowledge

### AWS Business Essentials

- Instructor-led
- One day

### AWS Technical Essentials

- Instructor-led
- One day

### Job Roles in the Cloud

- Free, Online Course
- 30 minutes

## Technical courses: advance your skills

### Architecting on AWS

- Instructor-led
- Three days

### Advanced Architecting on AWS

- Instructor-led
- Three days

### Developing on AWS

- Instructor-led
- Three days

### DevOps Engineering on AWS

- Instructor-led
- Three days

### Systems Operations on AWS

- Instructor-led
- Three days

## Specialty topics: learn best practices

### AWS Security Fundamentals

- Free, Online Course
- Three Hours

### Security Operations on AWS

- Instructor-led
- Three Days

### Big Data Technology Fundamentals

- Free, Online Course
- 90 Minutes

### Big Data on AWS

- Instructor-led
- Three Days

### Introduction to Machine Learning

- Free, Online Course
- 40 Minutes

### Deep Learning on AWS

- Instructor-led
- One Day

### The Secret to Successful Cloud Transformations

- Free, Online Course
- 40 Minutes

### Migrating to AWS

- Instructor-led
- Two Days

### Data Warehousing on AWS

- Instructor-led
- Three Days

### Building a Serverless Data Lake

- Instructor-led
- One Day

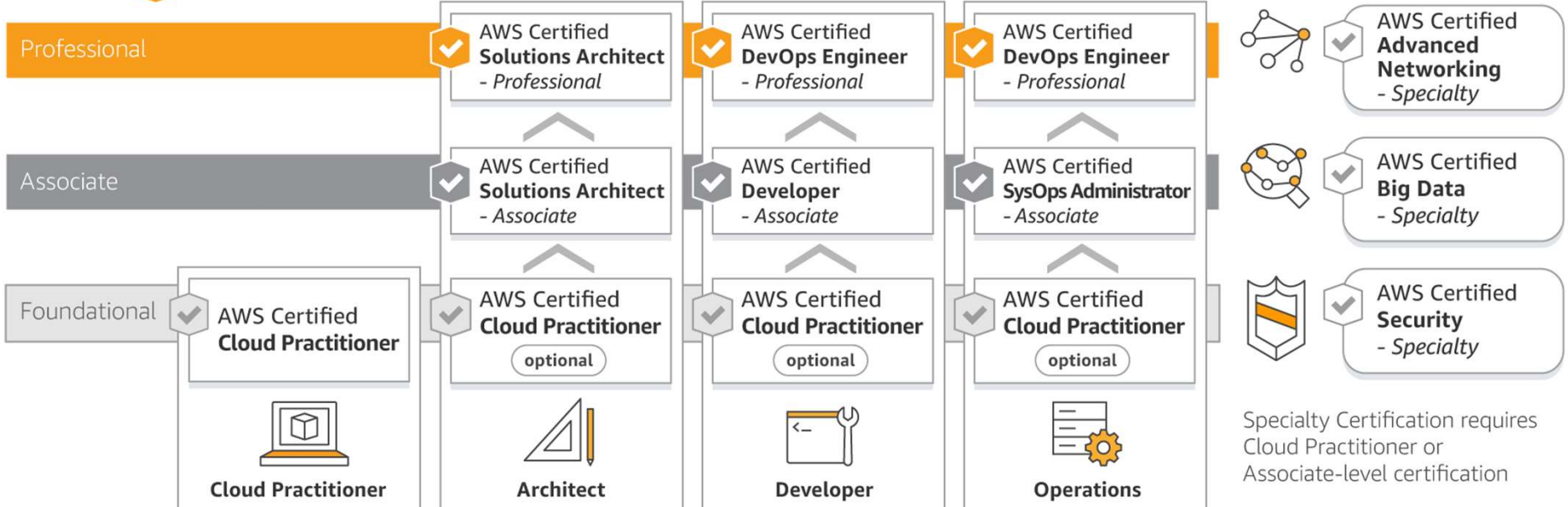


# Get Certified

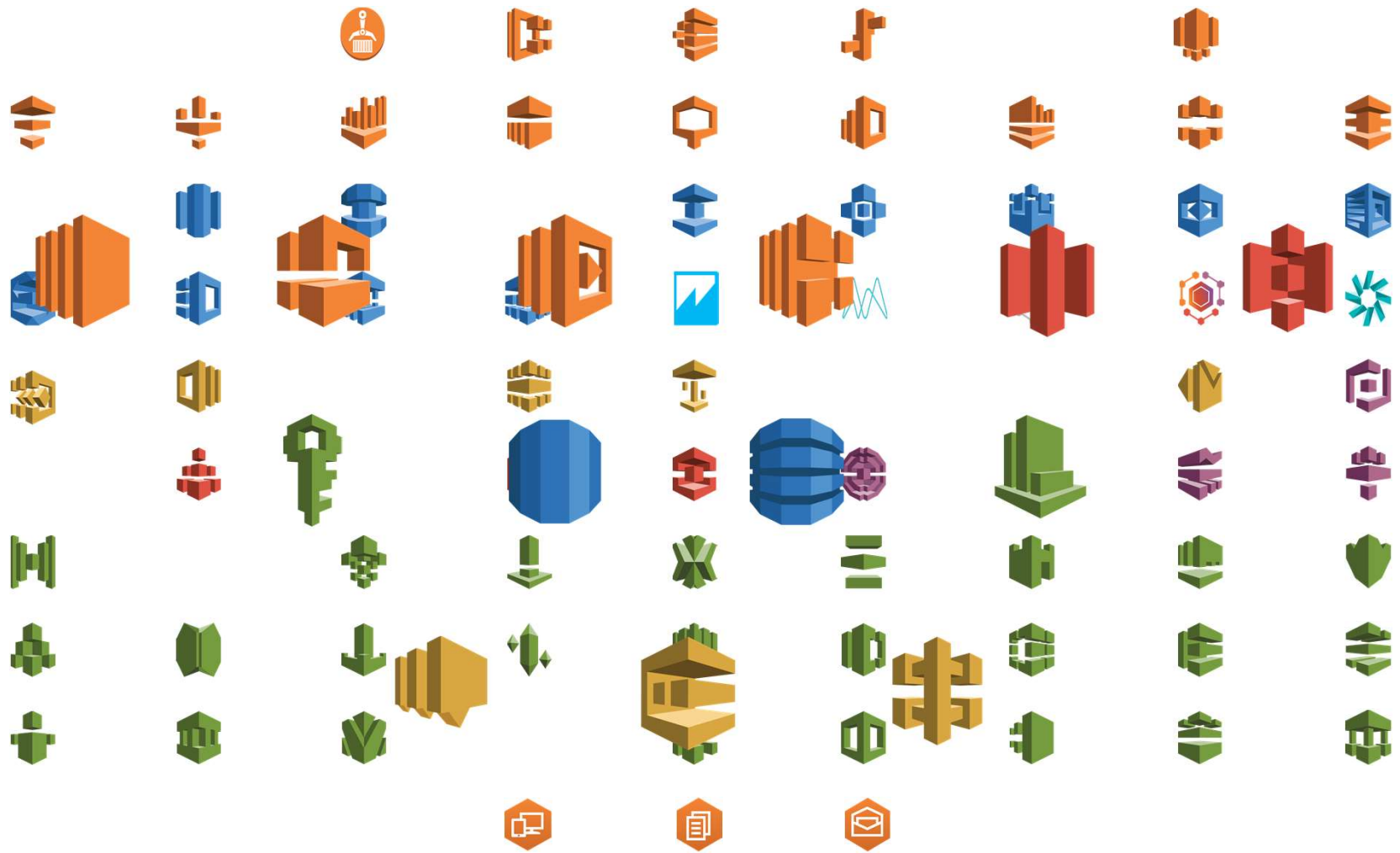


## Role-Based Certifications

## Specialty Certifications









# Thank you!