

# Amazon Elastic Compute Cloud (EC2)



# What You'll Learn



Define Elastic Cloud Computing

Identify EC2 benefits and related terminology

Describe Amazon Machine Images

Identify the main events in the history of AWS

Describe the AWS platform

Differentiate Root Volume in Instance Store and EBS

Define Elastic File System and its usage

Describe and demonstrate ELB and Auto-scaling

# Basic Concepts of EC2

# What is EC2?

---



Scalable computing capacity

Manage resources and servers

Pay for what you use

Boot-up instances in minutes

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Instances are virtual computing environments.

Every virtual server in EC2 is called an instance.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

AMIs (Amazon Machine Images) are preconfigured templates used to create instances.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Instance types are optimized instances that fit various use cases.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

EC2 key pairs provide a secured login for your instances.

AWS stores the public key and you store the private key in your client's machine.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs



# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Instance stores are storage volumes for temporary data that are physically attached to an EC2 instance.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Elastic Block Store is a persistent block level storage volume.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Security Group is a software firewall that enables you to specify the protocols, ports, and source IP ranges to access your instances.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Public IP address is used to reach the instance over the internet.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Private IP address is used to communicate between instances.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

The Elastic IP address is a static IP address used as a Public IP address to reach the instance over the internet.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

Tags are metadata that can be assigned to your Amazon EC2 resources.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs

# EC2 Terminology

Instances

AMIs

Instance Types

EC2 Key Pair

Instance Store

Elastic Block Store

VPCs are virtual networks that are logically isolated from the rest of the AWS cloud.

Security Group

Public IP

Private IP

Elastic IP

Tags

VPCs



# Benefits of EC2

---

Elastic Web Scale Computing

Complete Control

Flexible Cloud Hosting Services

Reliability and Security

Inexpensiveness

Designed to work with other AWS services

# EC2 Instance Types and Resource Usage



## Instance Types

- Determine the hardware of the underlying host.
- Each type offers different compute, memory, and storage capabilities.
- Grouped into instance families based on capabilities

## Instance Family

**Type A**

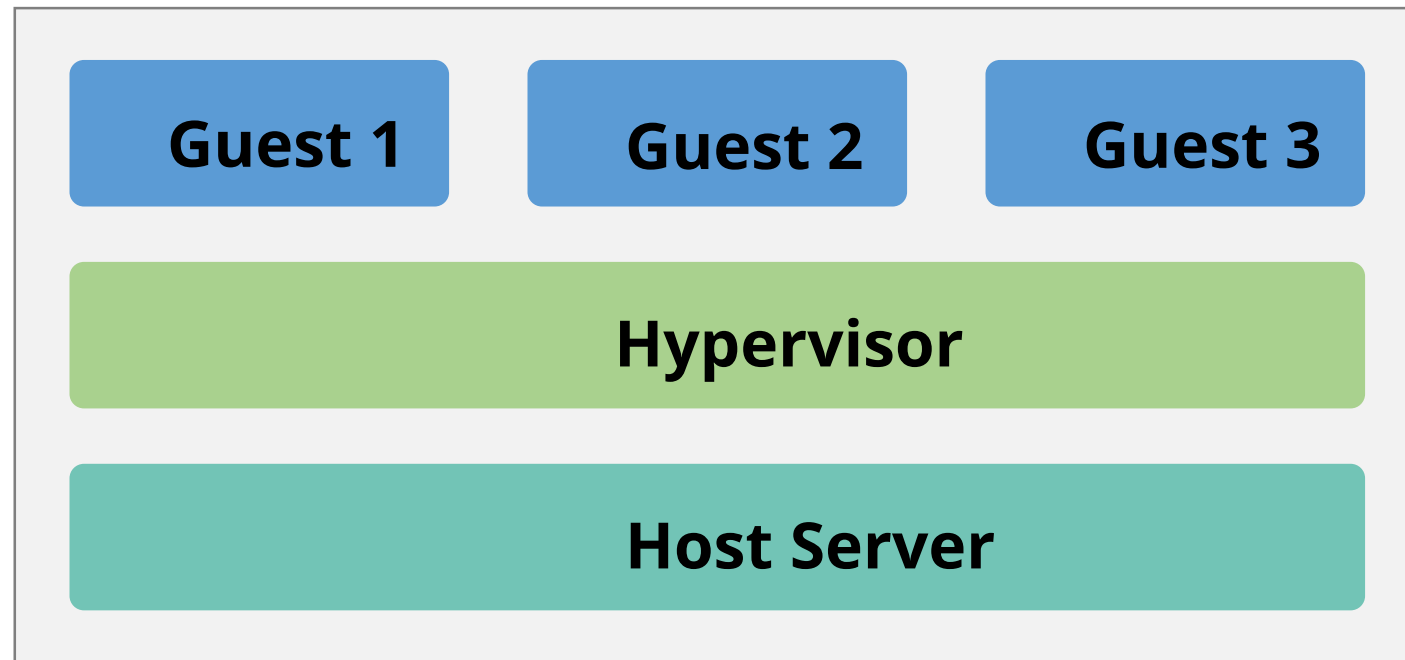
**Type B**

**Type C**

# EC2 Instance Families

Instance Family	Current Generation Instance Type	Previous Generation Instance Type
General Purpose	T2, M4 and M3	M1
Compute Optimized	C4 and C3	C1
Memory Optimized	R3	M2
Storage Optimized	I2 and D2	Hi1 and Hs4
GPU Instances	G2	cg1

# Virtualization Types



## Virtualization Types

Para-Virtual (PV)

Hardware Virtual Machine (HVM)

HVM is supported by all new generation instance types.

Difference: Boot and use of special hardware extensions for better performance.

**HVM** makes the operating system run directly on top of a VM without any modification

**HVM** AMIs support enhanced networking, GPU processing, and hardware-assist or extensions

# Instance Types and Storage

---



Some Instance Types support EBS (**Elastic Block Store**) and **Instance Store** volumes, but others support only EBS volumes.



Some instance types use SSD based instance store volumes to deliver very high random I/O performance.



Some types are **EBS-optimized** by default, to get dedicated capacity for Amazon EBS I/O.



**EBS optimization** supports dedicated block storage throughput of 500 Mbps to 10,000 Mbps.

# Instance Types—Purchase Options

---

On Demand



Reserved Instances



Spot Instances



Dedicated Instances



Scheduled Reserved Instances





# Instance Types – Purchase Options

## On Demand



- On demand Instances
  - Pay as you go and Pay per hour
  - No long-term commitment or upfront payment
  - Capacity can be increased or decreased
  - May not be available for short periods
- Use cases:
  - Short term, temporary use
  - Unpredictable workloads that cannot be interrupted
  - Development and Test Environments

# Instance Types – Purchase Options

## Reserved Instances



- Reserved Instance
  - Up to 75% discount, unlike On demand instance
  - Upfront partial or full payment
  - Always available
  - Can Move between AZs, Modify instance types, and Resell
- Use cases:
  - Applications with steady state or predictable usage
  - Applications that require reserved capacity
  - Make upfront payments to reduce total computing costs



# Instance Types – Purchase Options

## Spot Instances



- Spot Instances
  - Compute capacity with no upfront commitment
  - Allow you to specify(Bid) the maximum hourly price
  - Spot Price fluctuates based on supply and demand
  - Lose your instance when spot price > customer price
- Use cases:
  - Applications that have flexible start and end times
  - Applications that can afford only low compute cost
  - Low priority but high compute needs

# Instance Types—Purchase Options

## Dedicated Instances



- Dedicated Instances
  - Single customer tenancy on customer hardware
  - Physically isolated at host hardware level
  - Useful for customers with unique compliance requirements (like HIPAA PHI).
- Use cases:
  - Applications that have flexible start and end times
  - Applications that can afford only low compute cost
  - Low priority but high compute needs

# Instance Types—Purchase Options

## Scheduled Reserved Instances



- Scheduled Reserved Instances
  - Reserve capacity that you need on scheduled interval
  - Purchase capacity reservation that reoccurs on a daily, weekly, or monthly basis
  - Set Start time and duration of running resources and that too for complete one year
- Use cases:
  - For instances that don't run for entire 24 hours but run only for fixed hours in a day
  - For an application that runs during business hours
  - For a batch processing that runs at the end of the week



# Knowledge Check

KNOWLEDGE  
CHECK

What is the use of EBS Optimized volume?

- a. Durability of data up to 99.99%
- b. Dedicated capacity for Amazon EBS I/O
- c. Supports throughput of 500 Mbps to 10,000 Mbps
- d. Automatic snapshots



KNOWLEDGE  
CHECK

What is the use of EBS Optimized volume?

- a. Durability of data up to 99.99%
- b. Dedicated capacity for Amazon EBS I/O
- c. Supports throughput of 500 Mbps to 10,000 Mbps
- d. Automatic snapshots



The correct answers are **Dedicated capacity for Amazon EBS I/O & Supports throughput of 500 Mbps to 10,000 Mbps**

**Explanation :** Optimization provides the best performance for your EBS volumes by minimizing conflict between Amazon EBS I/O and other traffic from your instance. EBS-optimized instances deliver dedicated bandwidth to EBS, with options between 500 Mbps to 10,000 Mbps, depending on the instance type used.

KNOWLEDGE  
CHECK

Which of the following instances are always available on the specified recurring schedule for a one-year term?

- a. Spot Instance
- b. Reserved Instances
- c. Scheduled Dedicated instances
- d. Scheduled Reserved Instances



KNOWLEDGE  
CHECK

Which of the following instances are always available on the specified recurring schedule for a one-year term?

- a. Spot Instance
- b. Reserved Instances
- c. Scheduled Dedicated instances
- d. Scheduled Reserved Instances



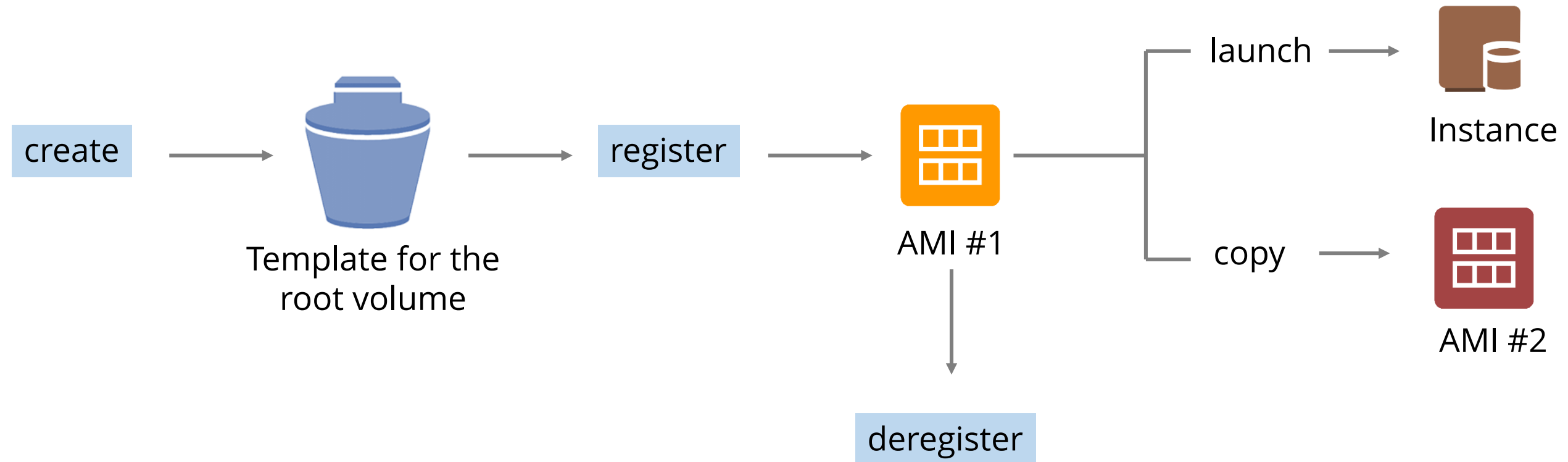
The correct answers are **d**

Scheduled Reserved Instances (Scheduled Instances) enable you to purchase capacity reservations that recur on a daily, weekly, or monthly basis, with a specified start time and duration, for a one-year term.

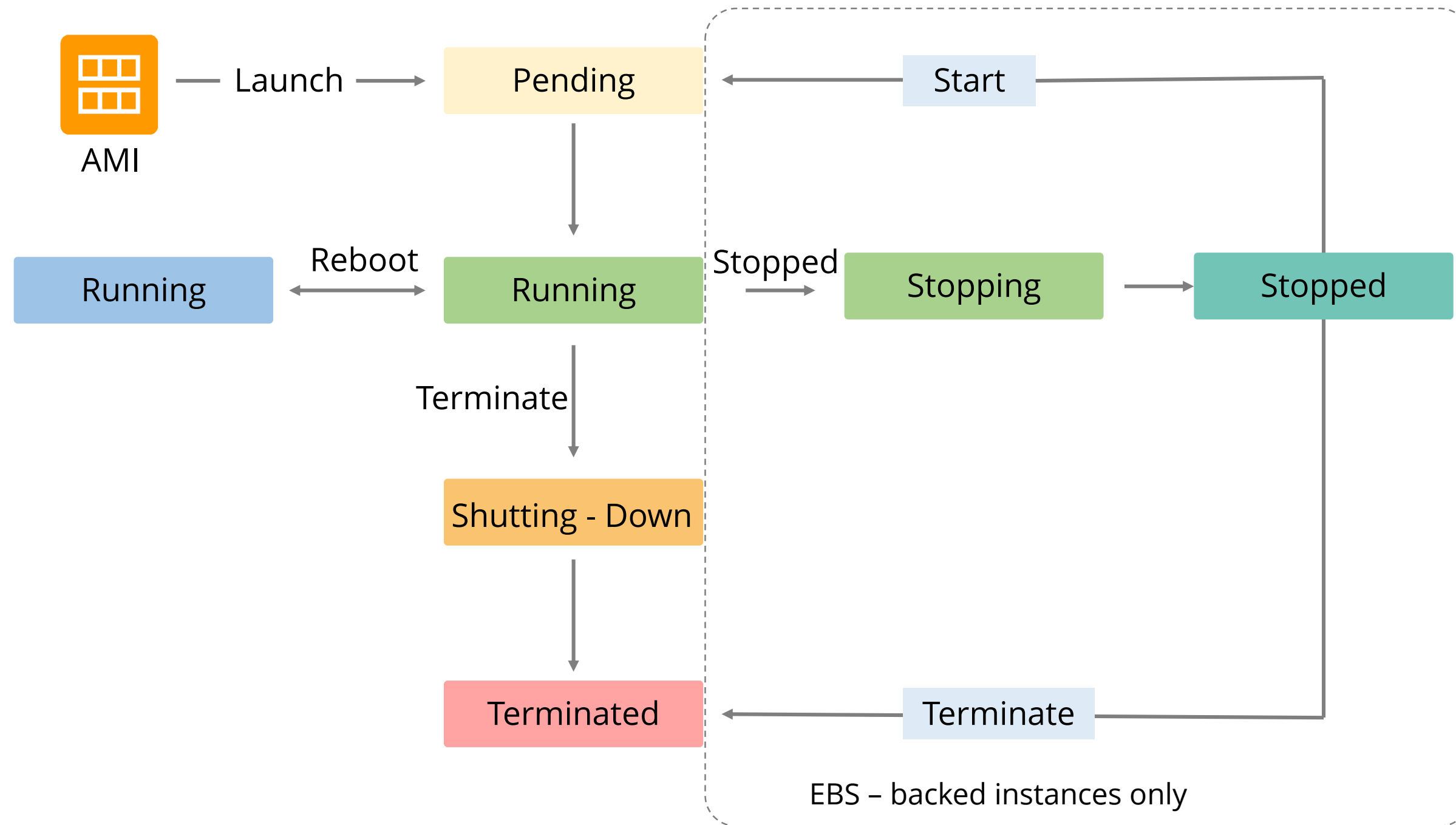


# EC2 Instance Creation Process

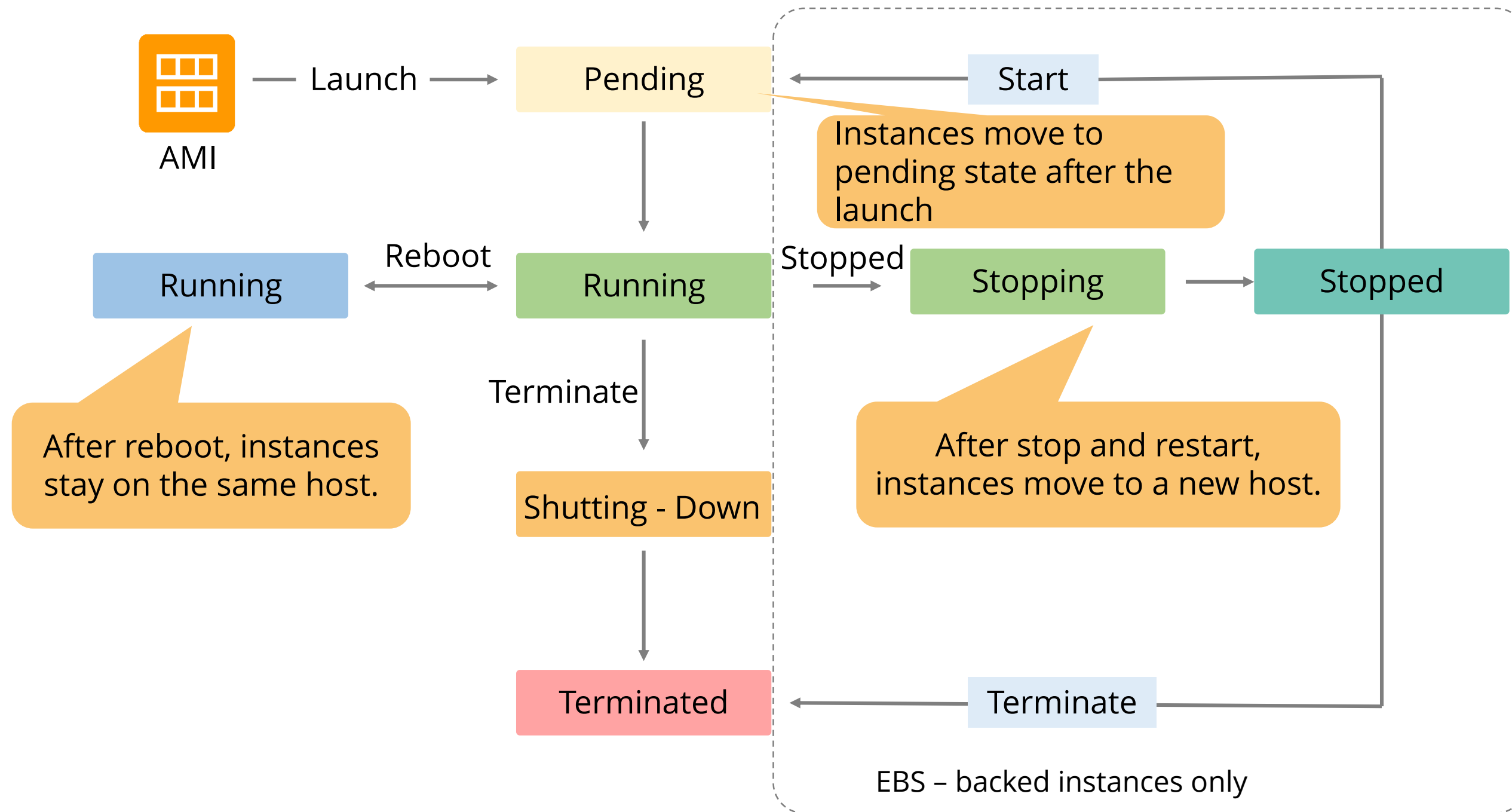
# Amazon Machine Images (AMIs)



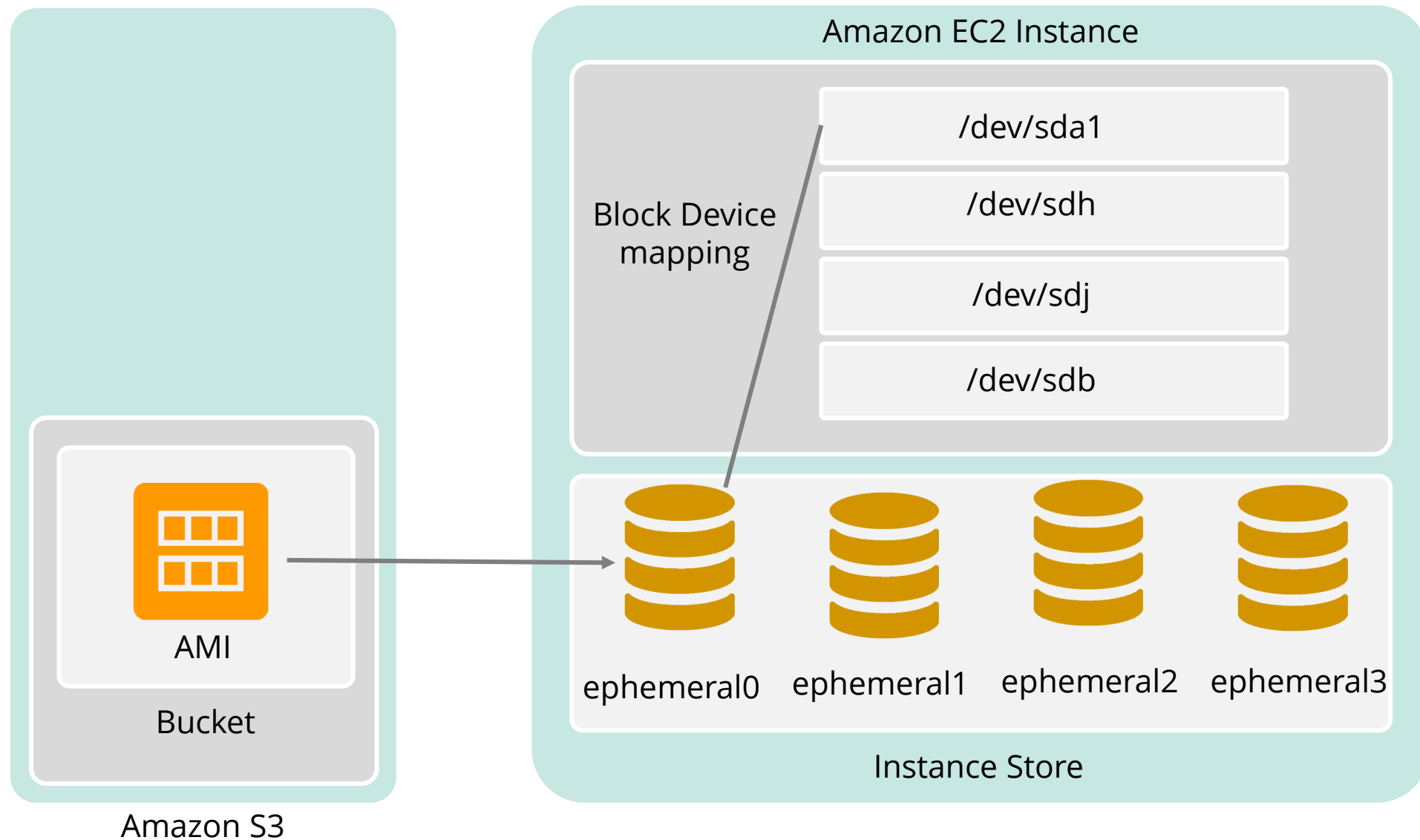
# Instance Lifecycle



# Instance Lifecycle



# Instance Store as Root Volume



Data in an instance store will exist as long as the associated instance exists:

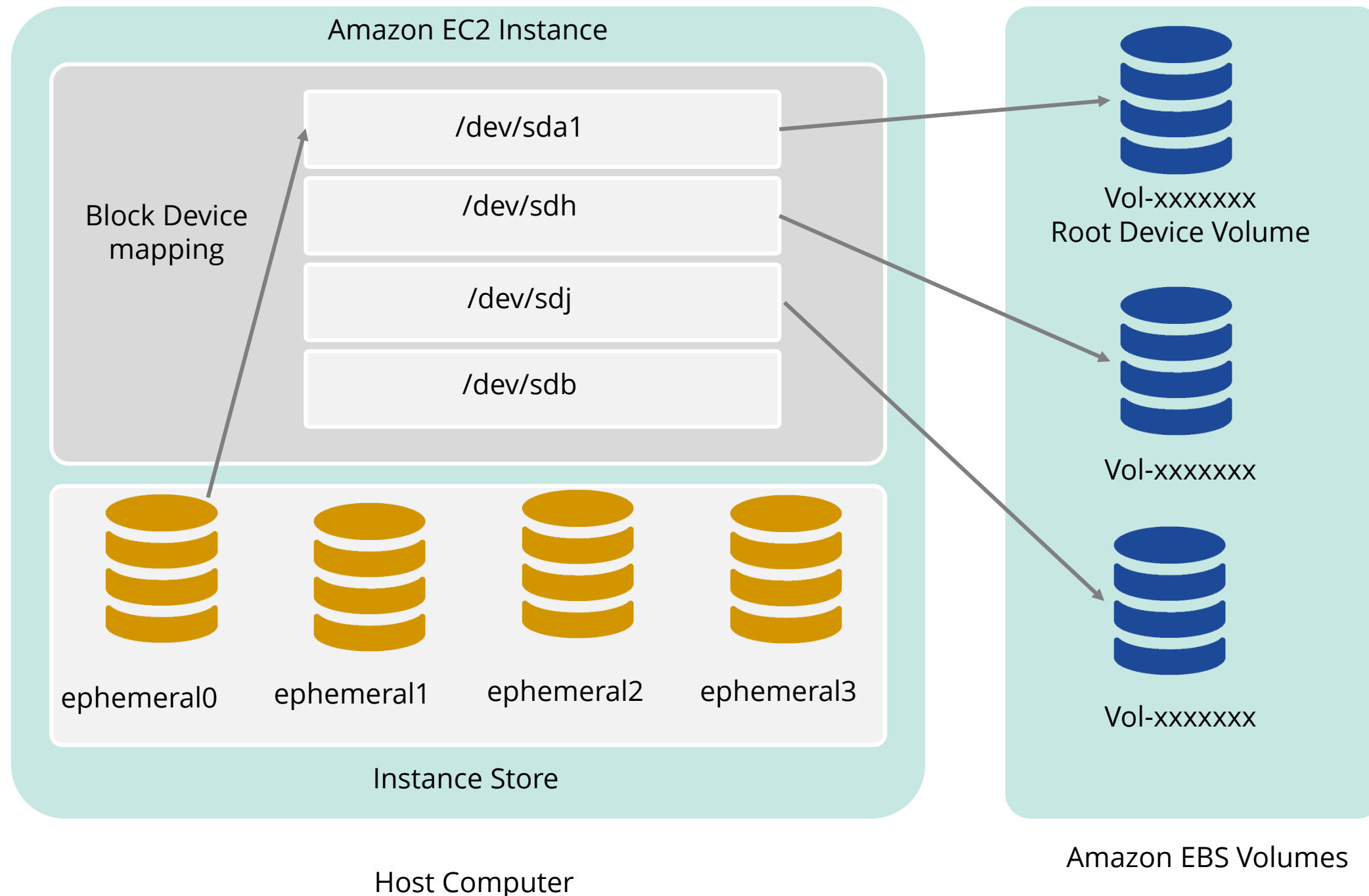
## Ephemeral Storage

Data is lost when instance is stopped or terminated

Data will continue to exist when the instance is rebooted

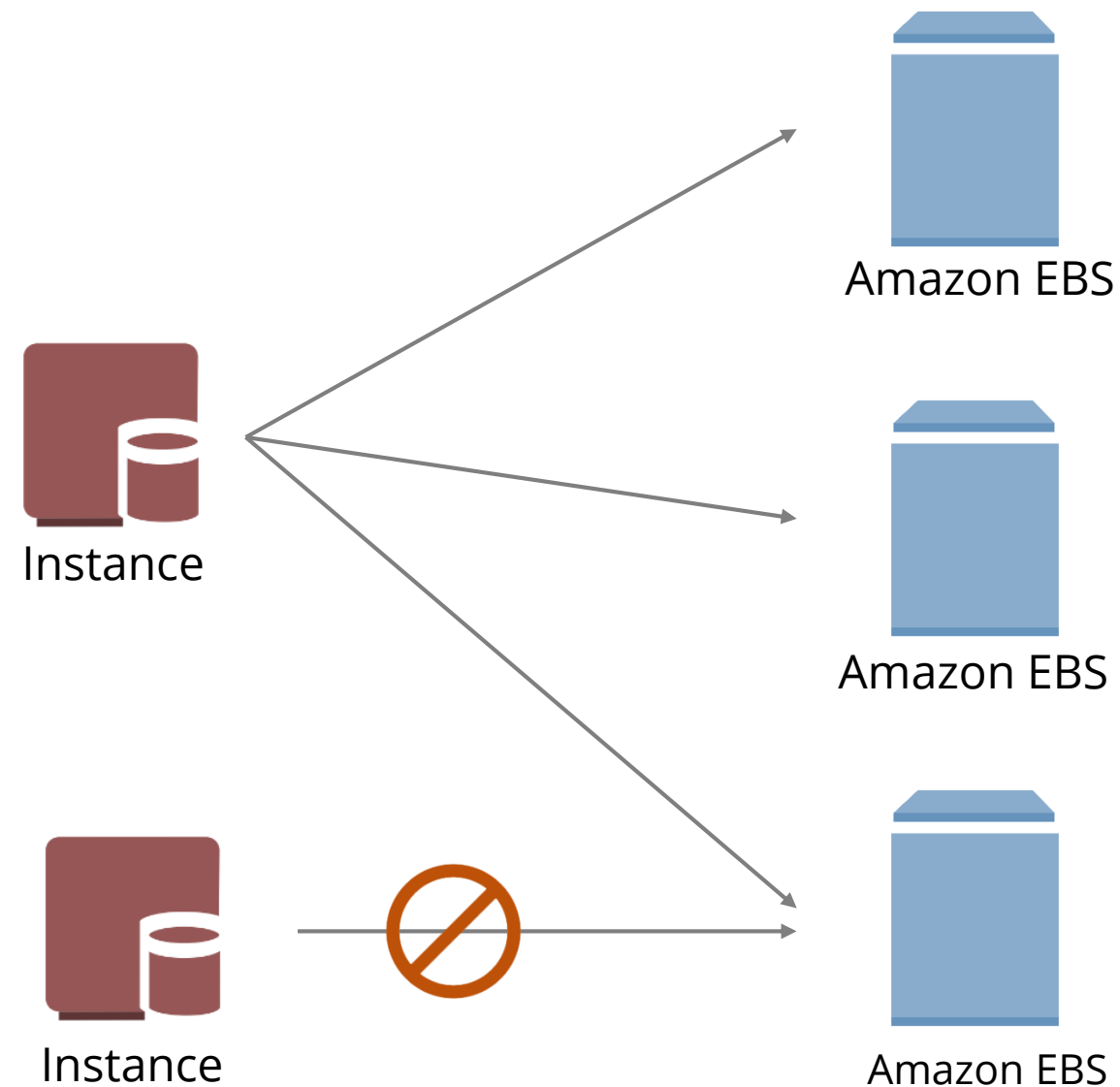
Back up the data to persistent storage like S3

# EBS as Root Volume



- If instance is stopped or rebooted, data in EBS is retained
- Root volume will be deleted on termination
- Retain by deselecting the **“Delete On Termination”** Checkbox

# EC2 Storage: EBS Volume



## EBS Volume

- Durable, block-level storage
- Primary storage solution for EC2
- Works as a physical hard drive
- Automatically replicated
- Is only available inside the availability zone
- Attached to only one instance at a time
- Multiple volumes can be attached to a single instance
- Users can stripe data across the volumes
- Uses 256-bit Advanced Encryption Standard algorithms (AES-256)

# EBS Volume Types

## Solid State Drives (SSD)

Backed volumes are optimized for best IOPS for smaller workloads

It delivers single digit millisecond latencies and the ability to burst to 3000 IOPS

Throughput Optimized volume is a low-cost magnetic storage ability to burst throughput

## Hard Disk Drives (HDD)

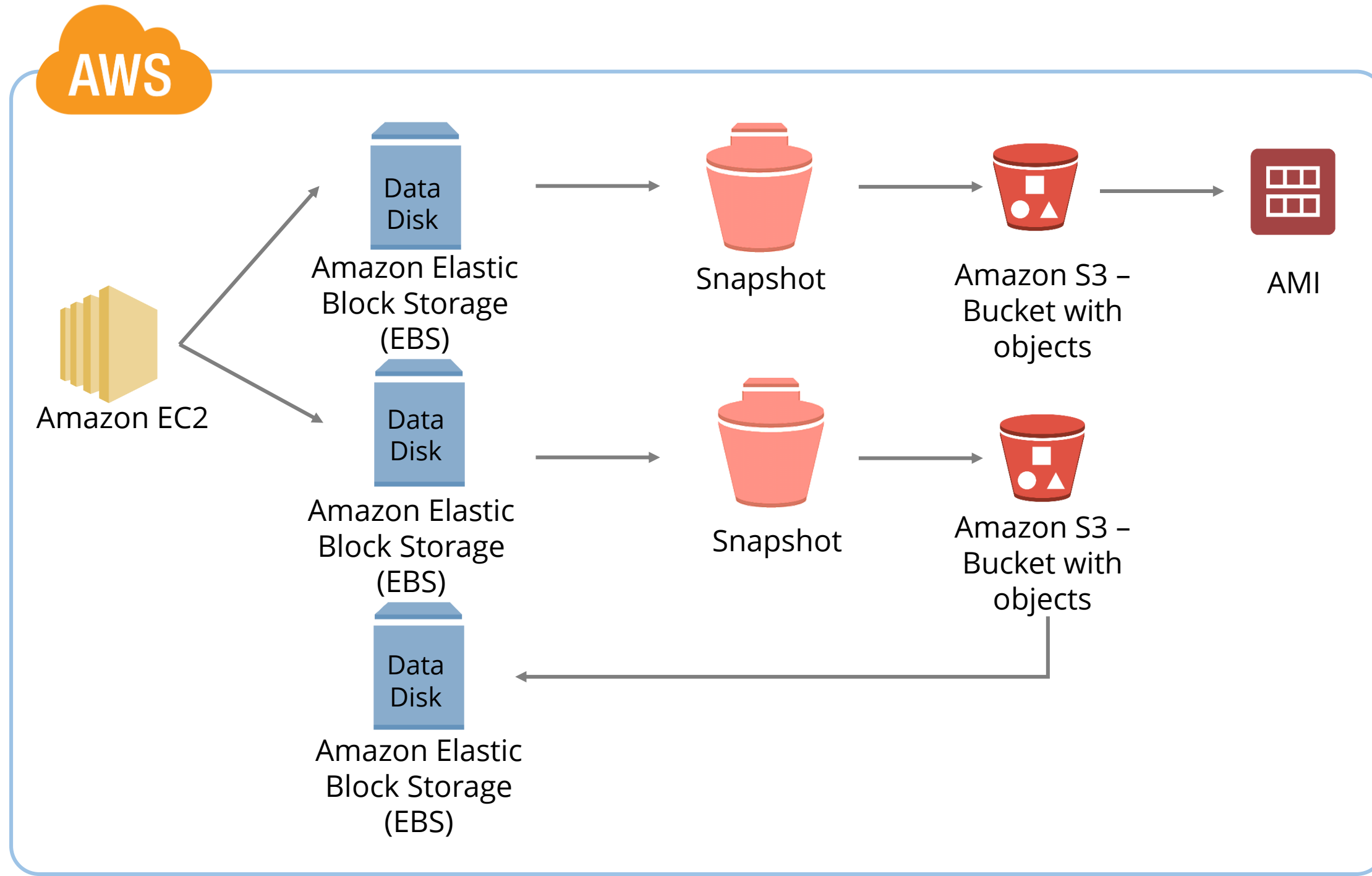
HDD based volumes are optimized for larger workload where throughput is more important than IOPS

Provisioned IOPS supports a volume size between 4 GB to 16 TB and 20,000 per volume

Cold HDD volumes are best suited for less frequently accessed workloads

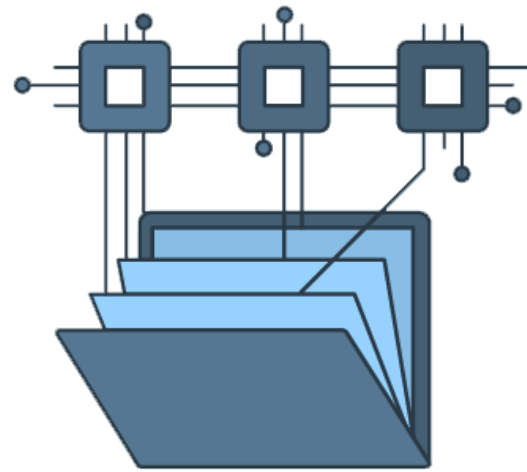


# EBS Volume Snapshots



# Amazon (Elastic File System)

---



New Service

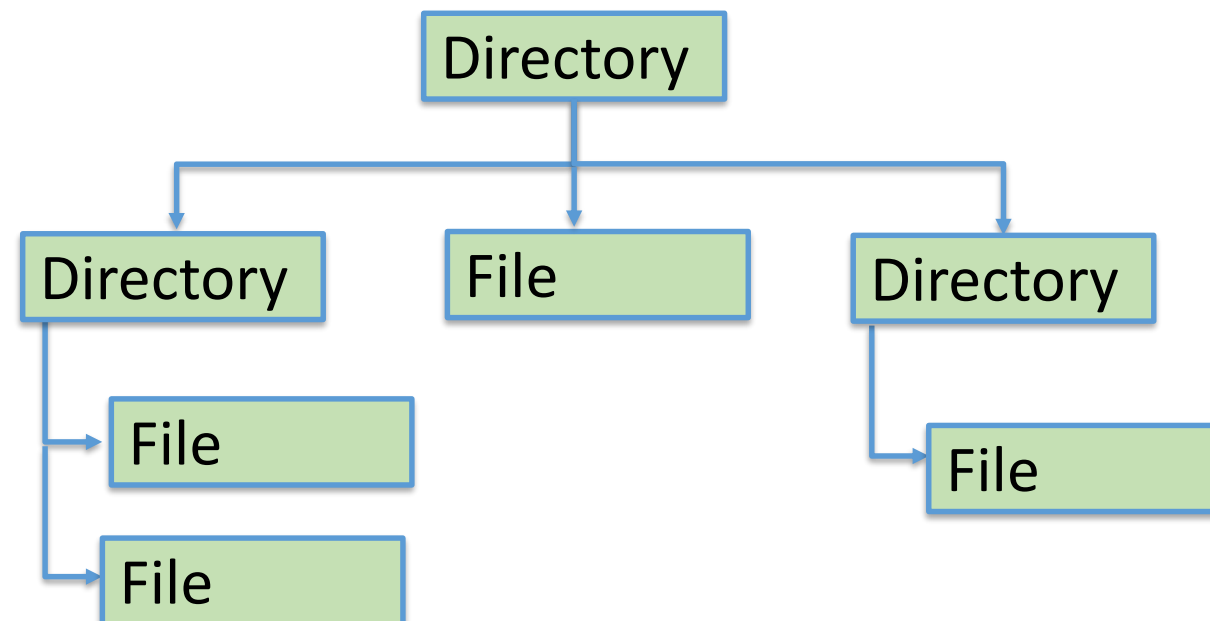
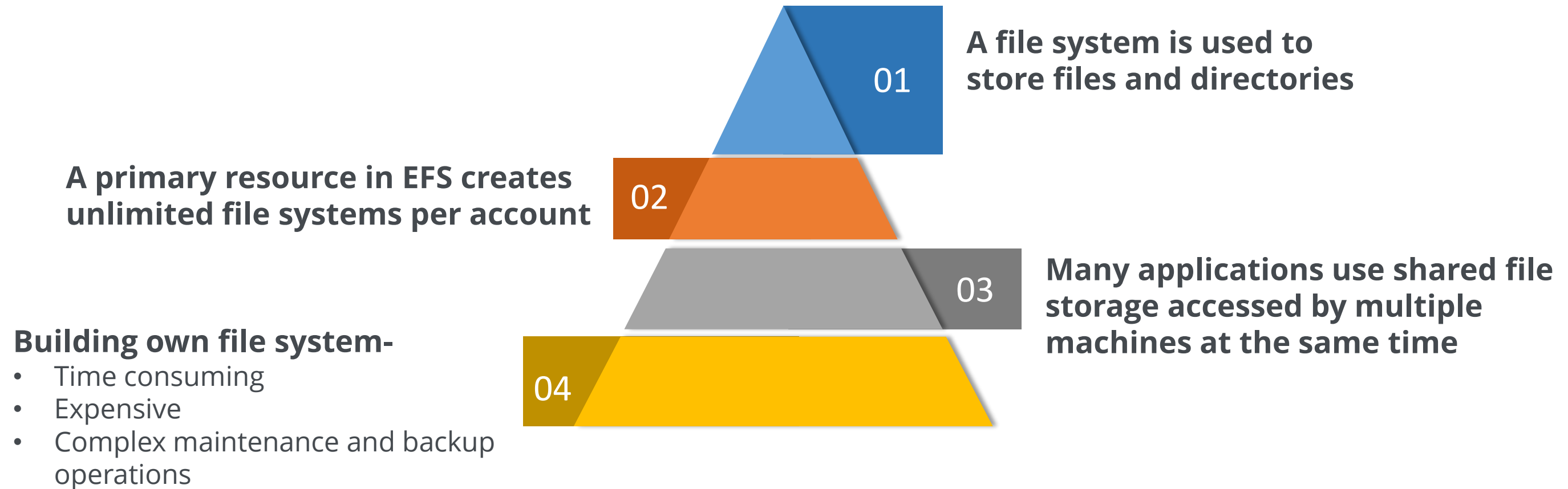
Elastic Storage Capacity

Shared Low Latency Access

Integrates with IAM for security

Highly available and durable data

# Elastic File System in Detail



# Amazon EFS Benefits



Fully managed sharable file storage system used for applications running on multiple EC2 instances (from few to thousands of instances)



Lets you run applications that use shared file system in cloud



NFS (Network File System) v4 based server in AWS cloud



Based on SSD (Solid State Disk)



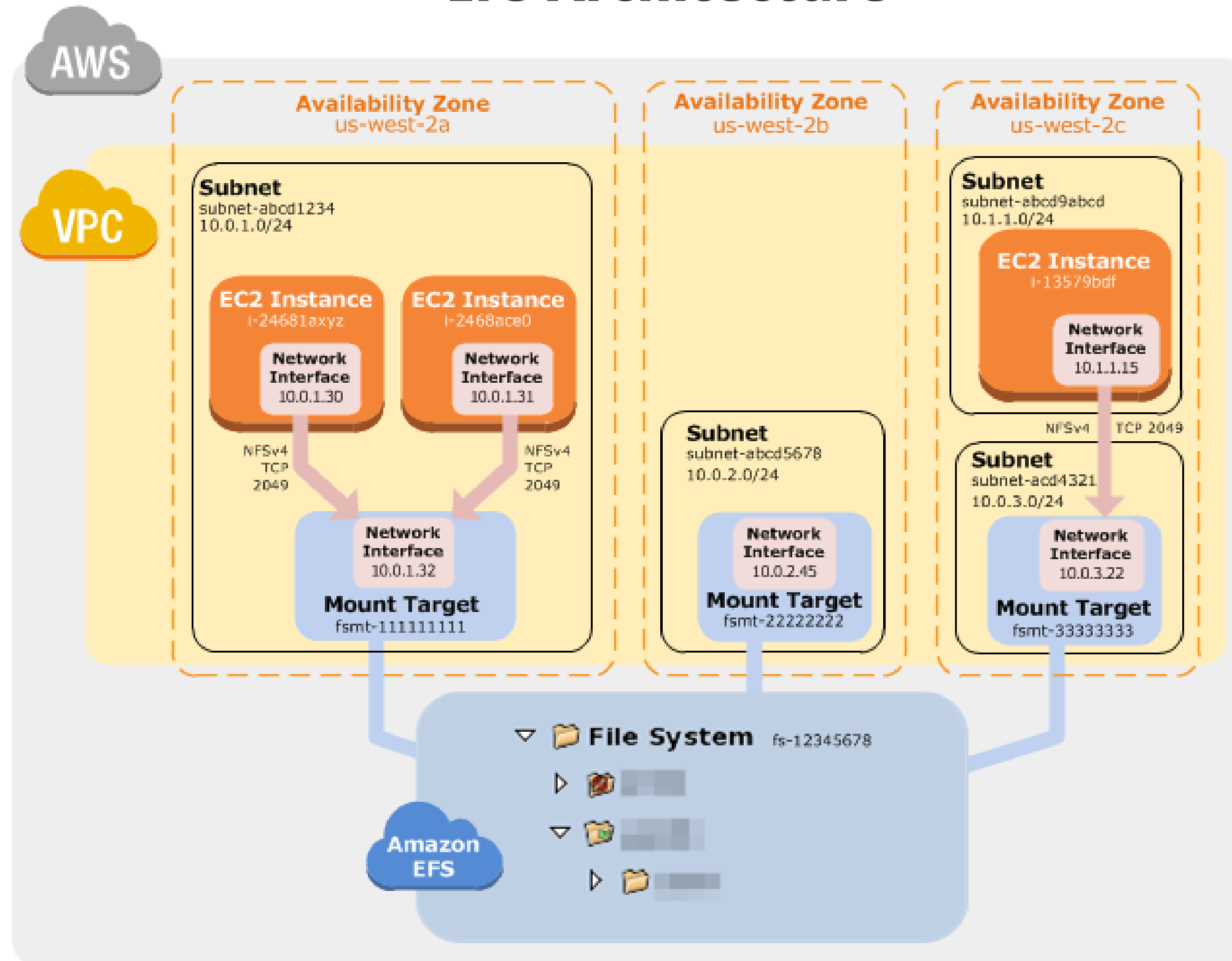
Mount Amazon EFS file systems on your on-premises datacenter servers when connected to your Amazon VPC with AWS Direct Connect



Set up and manage a file system using

- AWS Management console
- AWS CLI
- AWS SDK

# EFS Architecture



Source: <http://docs.aws.amazon.com/efs/latest/ug/how-it-works.html>

# EFS Benefits and Security Mechanisms

---

1 Scalable

2 Elastic

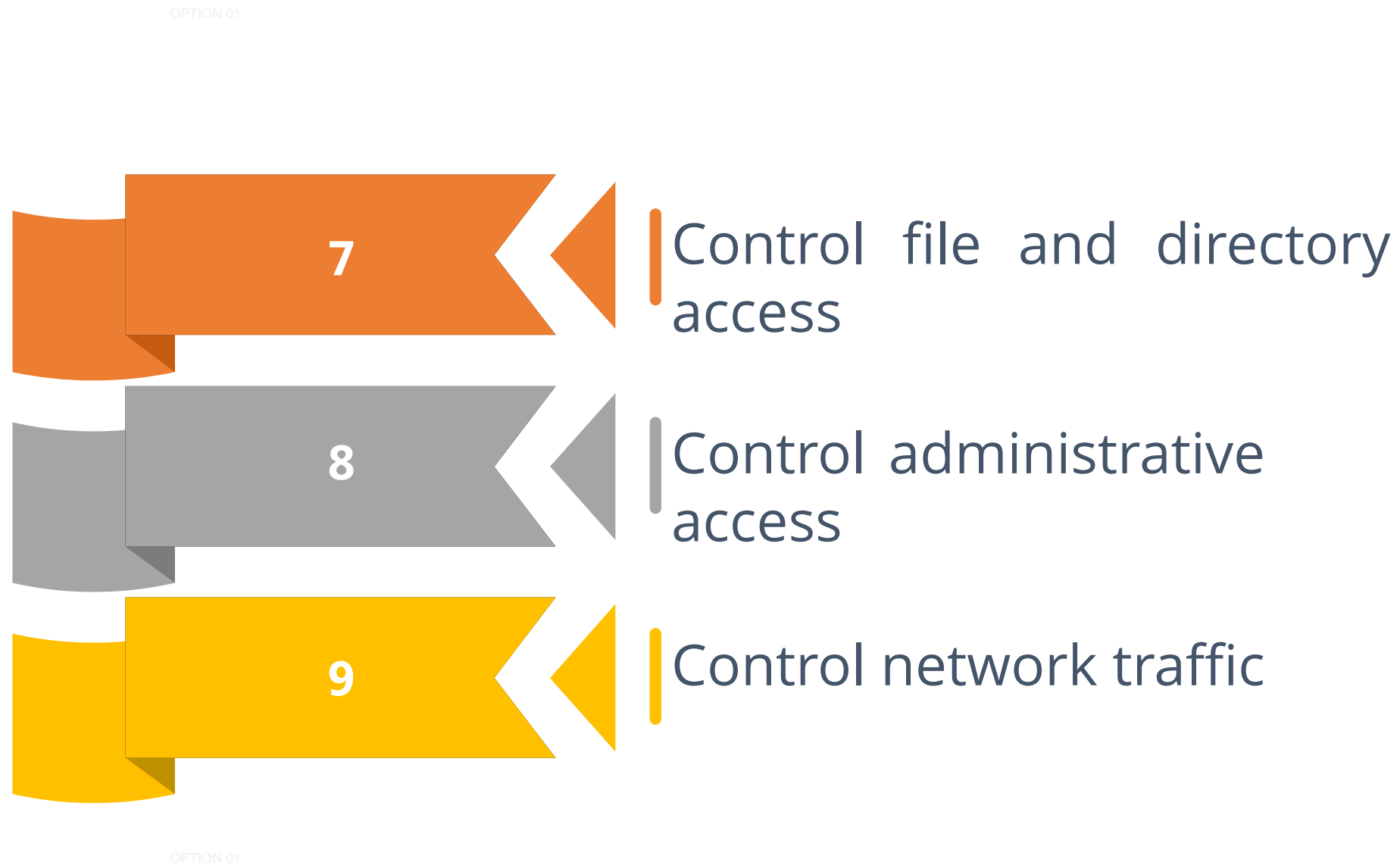
3 Simple

4 Fully Managed

5 Durable

6 Low Cost and Concurrent Access

# EFS Benefits and Security Mechanisms (Contd.)





# Knowledge Check



KNOWLEDGE  
CHECK

Which of the following statements are true about Instance Store?

- a. Data will be erased every 30 days.
- b. Data exists as long as the associated instance exists.
- c. Data is lost when instance is stopped or terminated.
- d. Data is lost when instance is rebooted.



KNOWLEDGE  
CHECK

Which of the following statements are true about Instance Store?

- a. Data will be erased every 30 days.
- b. Data exists as long as the associated instance exists.
- c. Data is lost when instance is stopped or terminated.
- d. Data is lost when instance is rebooted.



The correct answers are **B and C**

Explanation : The data in an instance store will exist as long as the associated instance exists. So, it is referred to as ephemeral storage. Data is lost when instance is stopped or terminated or when the underlining disk fails.

KNOWLEDGE  
CHECK

With Amazon EFS, you can mix and match the instance types connected to a single file system for file sharing.

- a. True
- b. False



KNOWLEDGE  
CHECK

With Amazon EFS, you can mix and match the instance types connected to a single file system for file sharing

- a. True
- b. False



The correct answers are **True**

**Explanation:** It is true that with Amazon EFS you can mix and match the instance types connected to a single file system for file sharing.

# EC2 Advanced Features

# IAM Roles with EC2

---

IAM roles are considered to be secure, universal, and easy to manage.

Define which account or AWS service can assume the role.

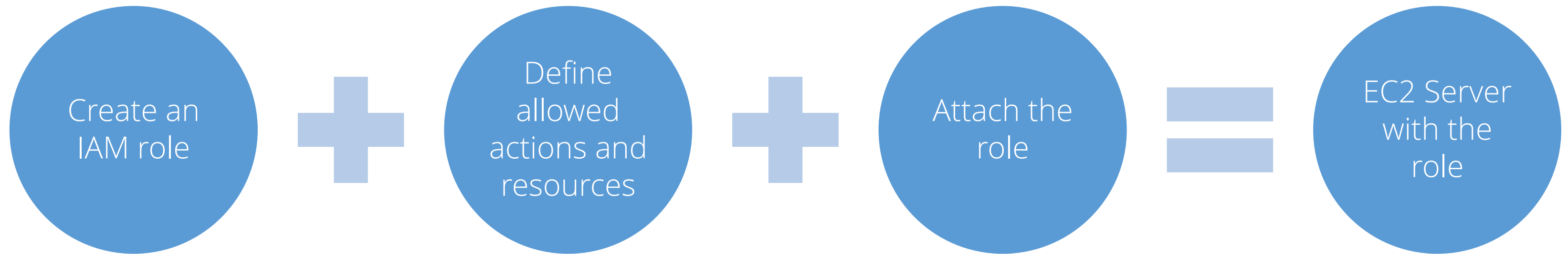
Define the actions and resources a role can do and attach them to EC2 instance.

Permissions on the IAM role associated with a running instance can be changed.

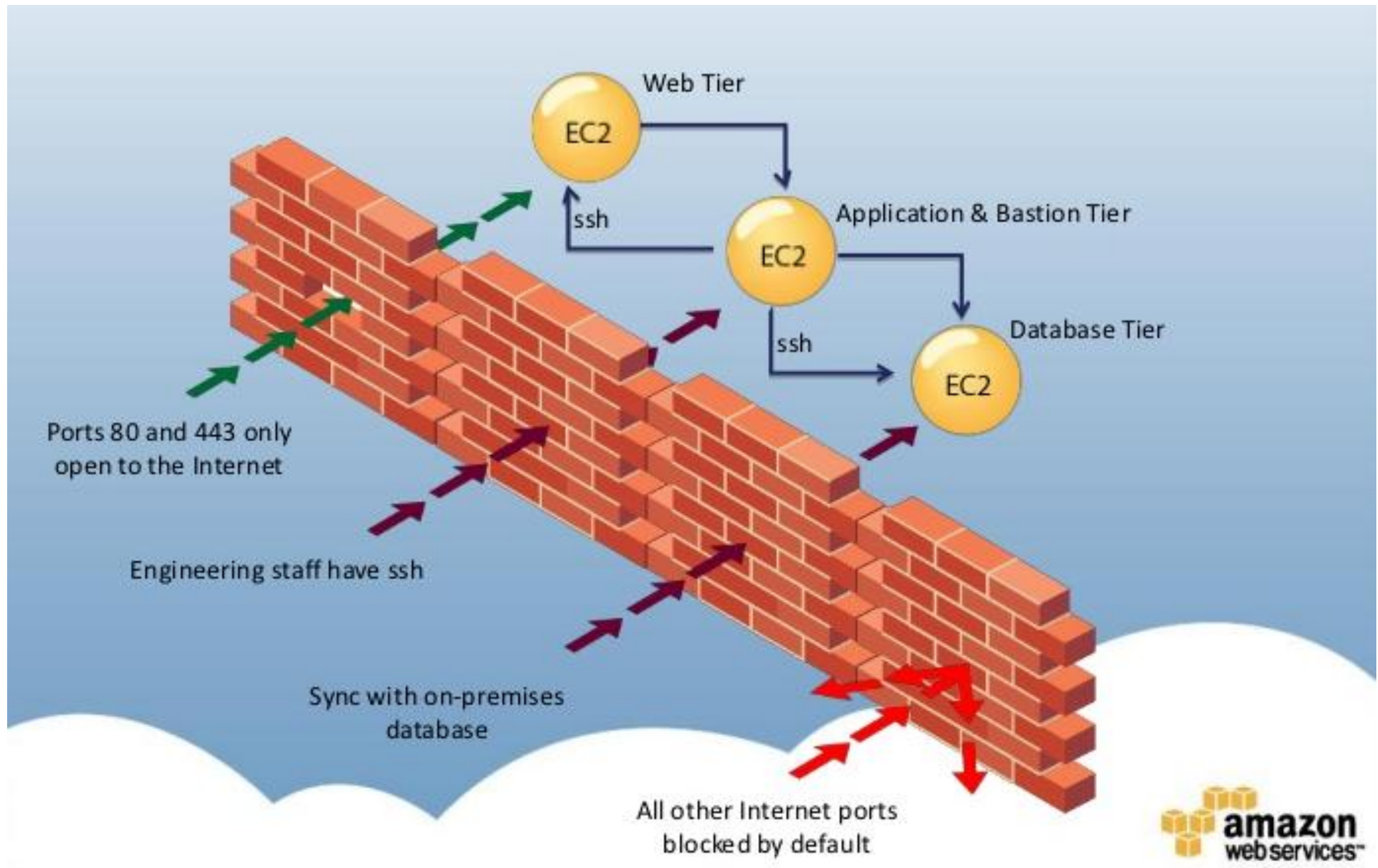


# IAM Roles with EC2 - Steps

---



# Security Groups with EC2





# Characteristics of Security Groups with EC2



Allows all outbound traffic from the instances



Does NOT allow users to write a rule to deny access



Rules can be added or removed at any time



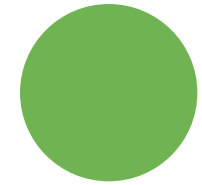
Rules can be copied to another security group



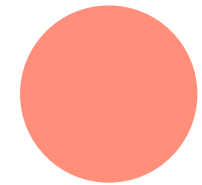
Security groups are statefull

# Elastic Network Interface

# What Is Elastic Network Interface (ENI)?

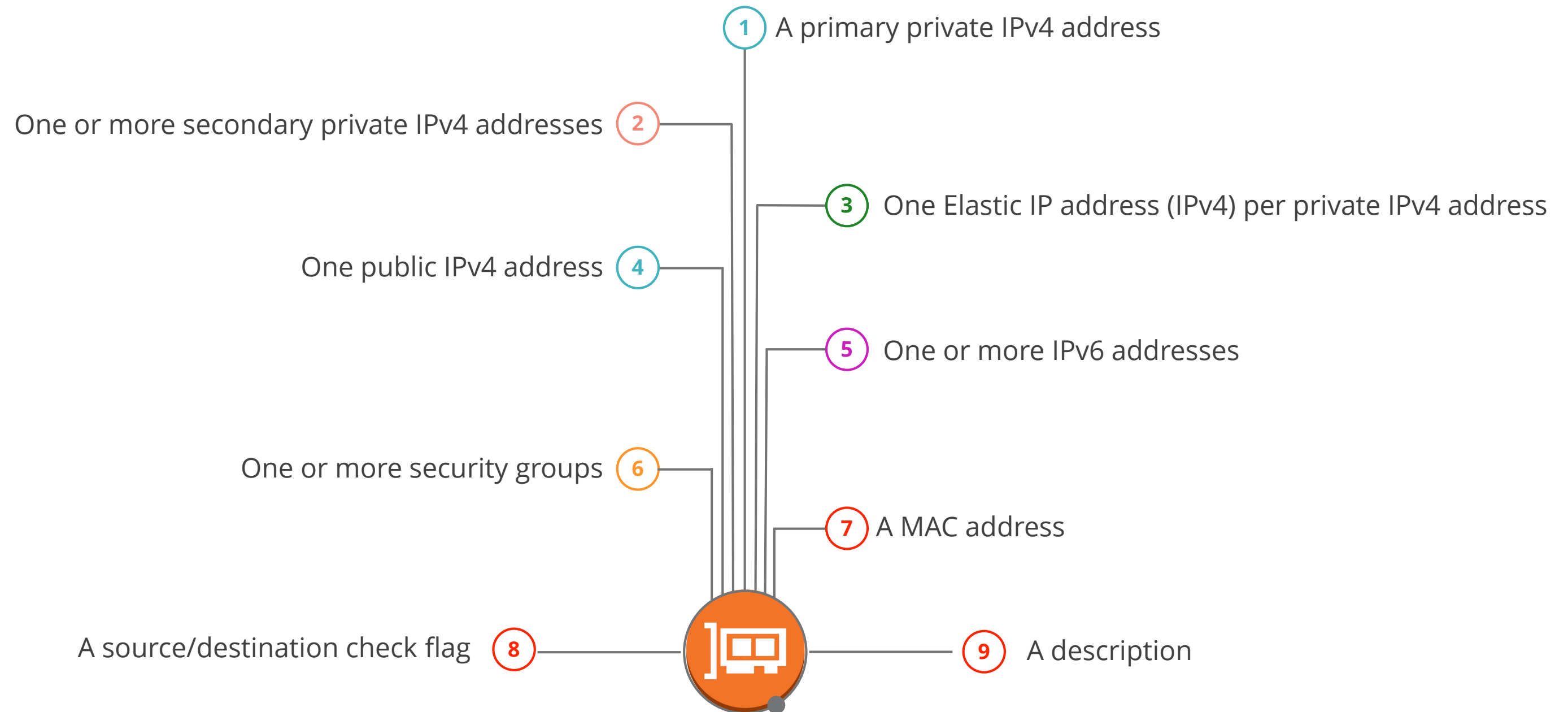


It is a virtual network interface that can be connected to instances in a VPC.

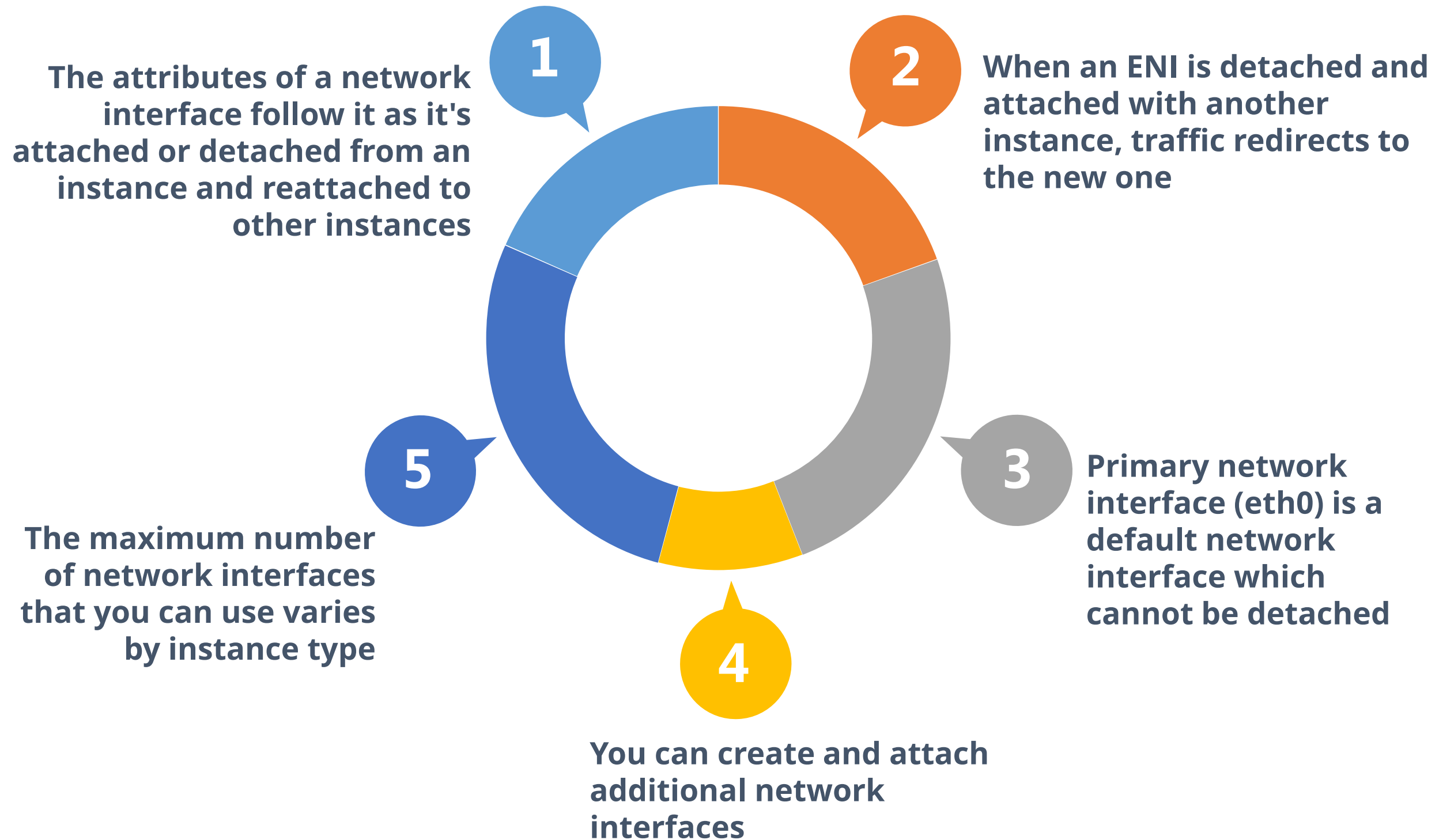


It is a set of network components that can be easily attached and detached.

# Elastic Network Interface Attributes



# Elastic Network Interface Characteristics



## EC2: Meta Data

---



- Data about an instance that can be used to manage the instance
- Can be viewed by anyone who has access to the instance
- Must not be used to store sensitive information such as passwords
- Curl or GET commands can be used
- Results are returned as plain text and data is listed in separate lines
- Dynamic data retrieves Instance identity information
- Can be accessed using the URL  
<http://169.254.169.254/latest/dynamic/>

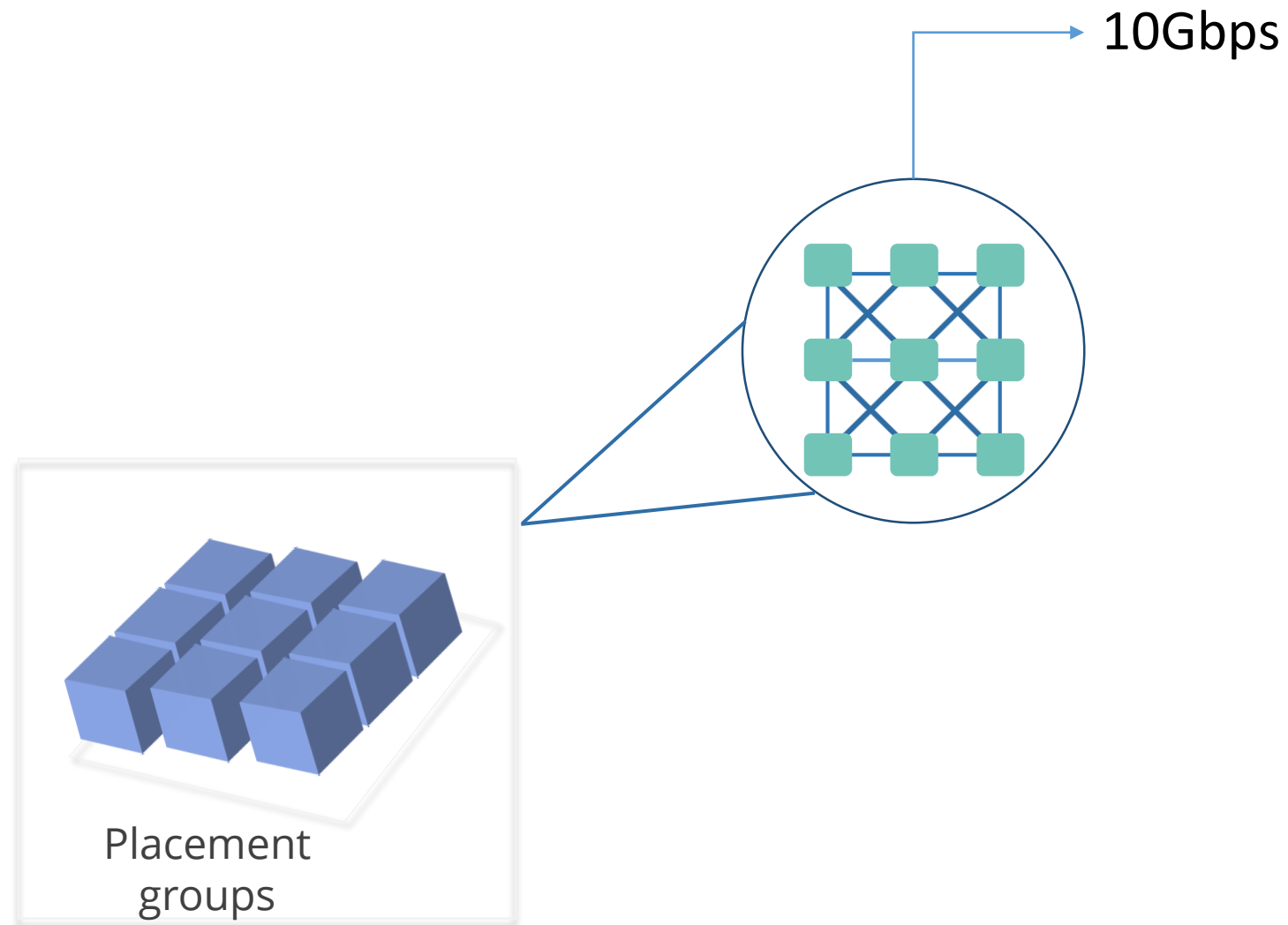
## EC2: User Data

---



- Used when an instance is launched; executed only at launch time
- Commonly known as bootstrap script
- If an instance is restarted with modified user data, it will not be executed
- Accessed using URL: <http://169.254.169.254/latest/user-data>
- 16KB limit in raw form or text file
- API submission requires base64 encoding

# Placement Groups



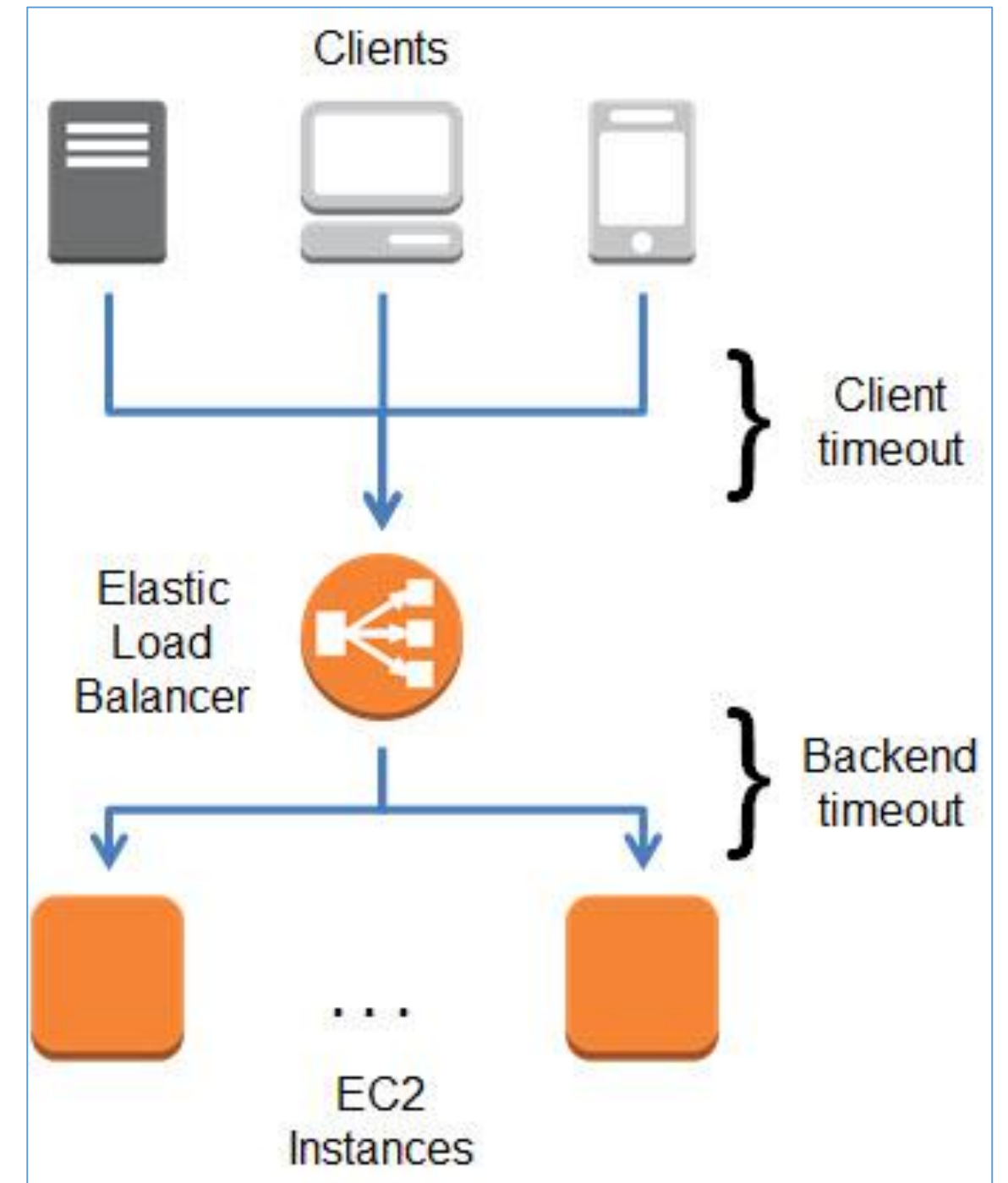
- Logical collection of instances inside a single Availability Zone
- 10 Gigabits per second network communication
- Use the same instance types or families within a placement group
- Placement Groups cannot span across multiple availability zones
- Placement groups do not support all instance types
- Placement groups cannot be merged
- Existing instance cannot be moved into a placement group



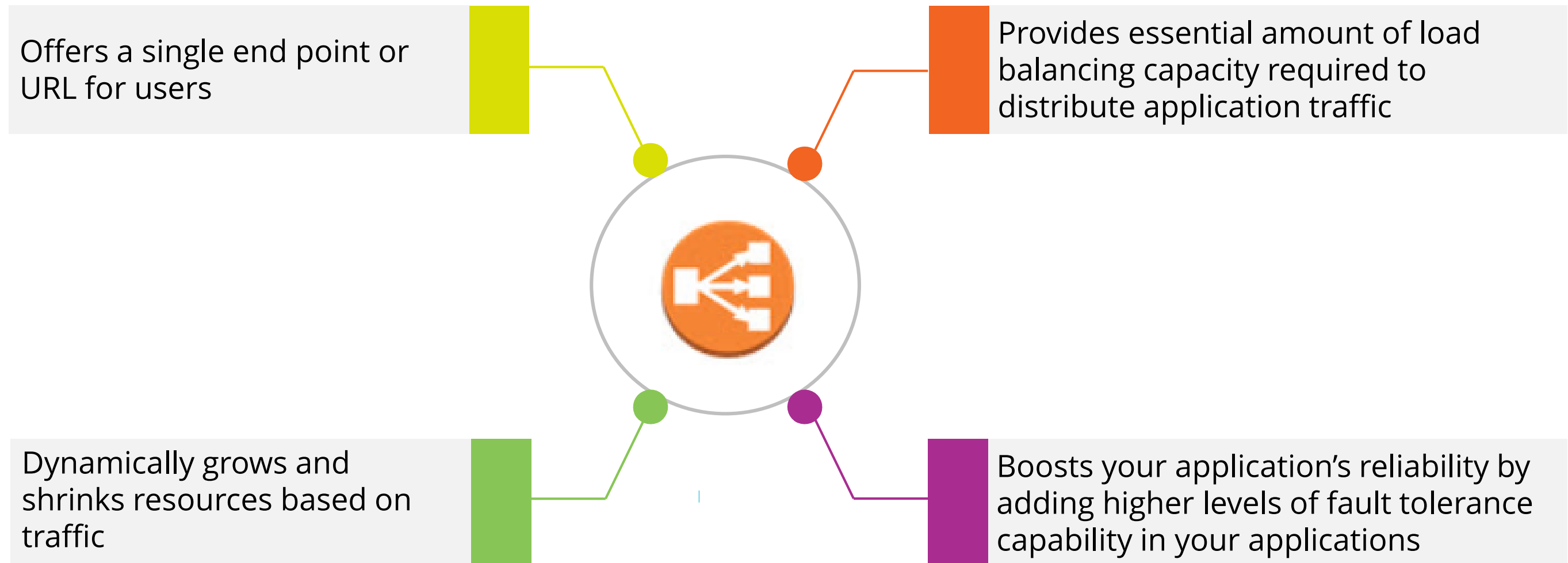
# Elastic Load Balancer Overview

## Elastic Load Balancer (ELB)

ELB helps to distribute incoming application traffic steadily and automatically to the various Amazon EC2 instances or Amazon Availability zones.

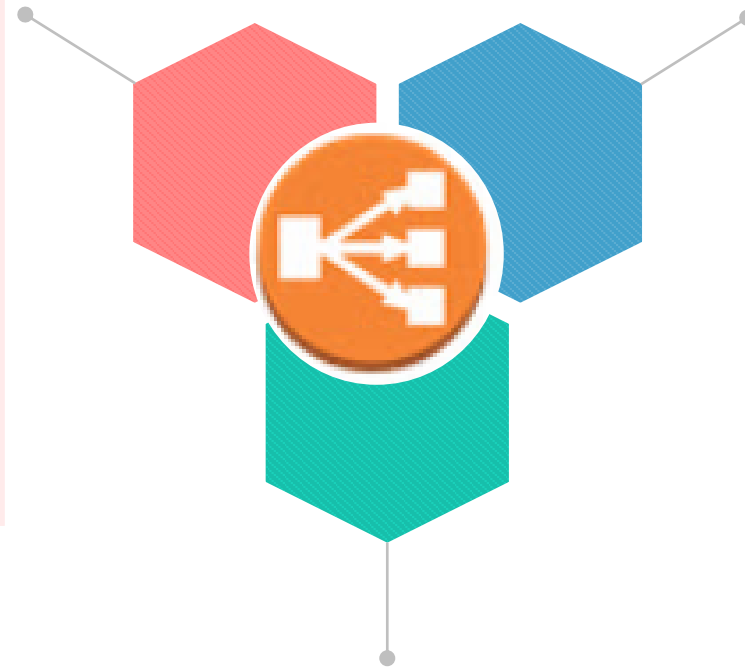


# Elastic Load Balancer Characteristics



# Elastic Load Balancer Types

Application load balancer provides you with a flexible feature set for your web applications with HTTP and HTTPS traffic. Operating at the request level, Application load balancers provide advanced routing, TLS termination, and visibility features targeted at application architectures including microservices and containers.



Network load balancer provides you with ultra-high performance and static IP addresses for your application. Operating at the connection level, Network load balancers are capable of handling millions of requests per second while maintaining ultra-low latencies.

Choose Classic load balancer when you have an existing application running in the EC2-Classic network.

# Classic Load Balancer

---



It distributes incoming application traffic across multiple EC2 instances.



Configured with health checks, Classic load balancer makes sure that the registered instance is healthy while routing the traffic to it.



Classic load balancer routes traffic either at application or network level.

# Application Load Balancer

---



Application Load Balancer makes routing decisions based on the application content.



It is a layer 7 load balancer which routes the traffic based on the target groups (applications)



Applications can have different rules to decide upon routing.

# Network Load Balancer

---



NLB can handle millions of requests per second and functions at fourth layer of OSI model.



It provides support for static IP addresses.



In NLB, one elastic IP address per subnet can be assigned.



It provides support for registering targets by IP address, including targets outside the VPC for the load balancer.

# ELB Facts

---

## COST

- ELB is charged by hours of usage and amount of data transfer

- Protocol and Ports supported by ELB

- HTTP
- HTTPs(Secured HTTP)
- TCP
- SSL(Secured TCP)
- Back-end Ports: 1-65535
- Front-end Ports: 25,80,443,465,587,1024-65535

## PROTOCOL AND PORTS

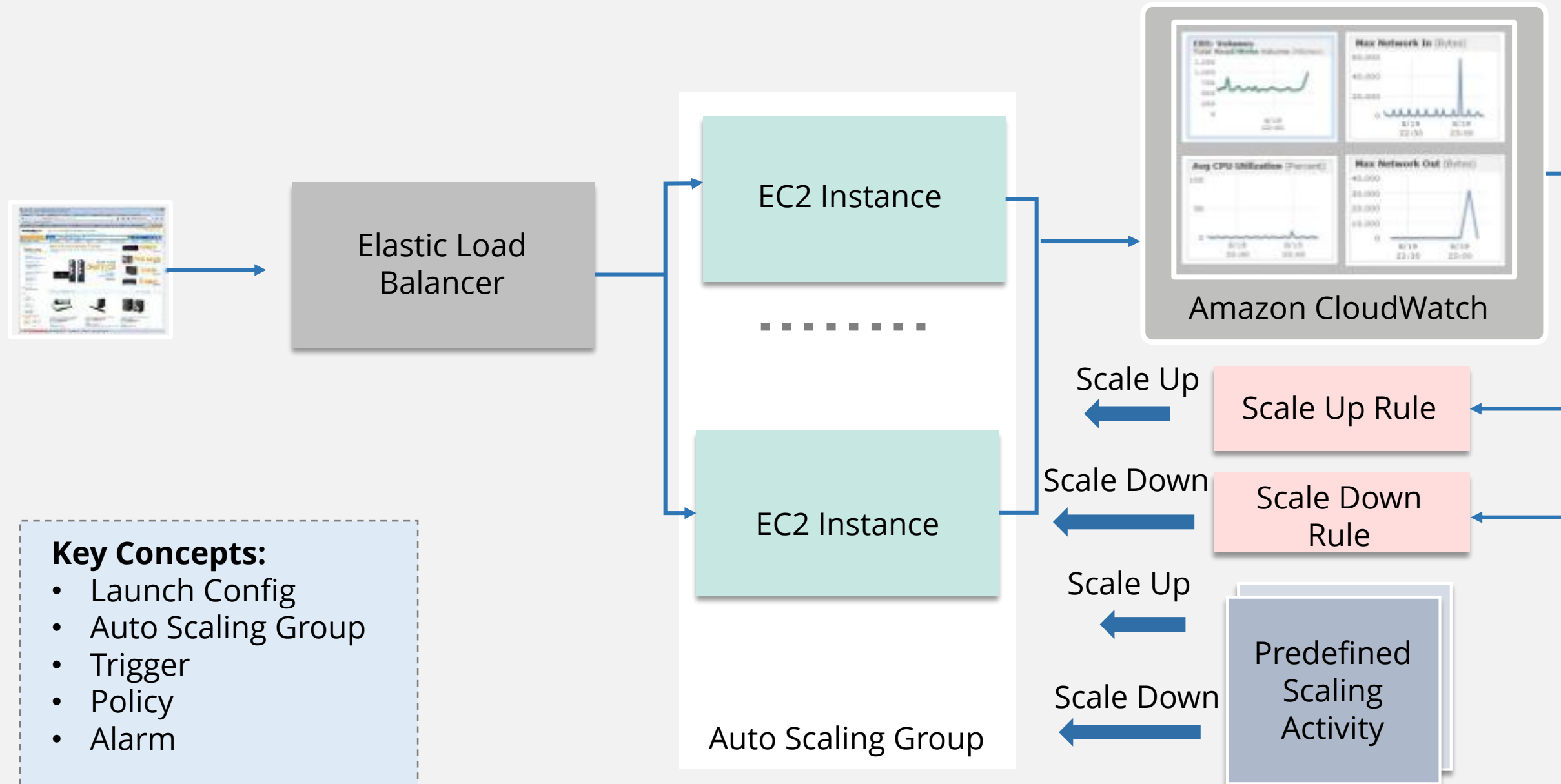
- HTTP Response Codes:

- 200 - Successful
- 3xx - Redirection
- 4xx - Client side error
- 5xx - Server side error

## RESPONSE CODES

# Amazon Auto Scaling

## Amazon Auto Scaling Architecture





# Amazon Auto Scaling, Cont..

---

## Health Checks

Relies on regular health checks to maintain the desired number of instances

## Manual Scaling

Allows users to change the size of the auto scaling group manually.

## Scheduled Scaling

Scheduled scaling performs a scaling action at a certain time in the future.

## Dynamic Scaling

Dynamic scaling scales dynamically and responds to changing demand.

## Across AZs

Scales across Availability Zones in the same region,  
Not across multiple regions



# Knowledge Check

KNOWLEDGE  
CHECK

## How does the security group work with EC2?

- a. Changes to rules are applied to all associated instances.
- b. Data exists as long as the associated instance exists.
- c. Data is lost when instance is stopped or terminated.
- d. Allows all outbound traffic from instances by default.



KNOWLEDGE  
CHECK

## How does the security group work with EC2?

- a. Changes to rules are applied to all associated instances.
- b. Data exists as long as the associated instance exists.
- c. Data is lost when instance is stopped or terminated.
- d. Allows all outbound traffic from instances by default.



The correct answer is **A and D**

**Explanation :** The default security groups allow inbound traffic only from other instances associated with the default security group and allow all outbound traffic from the instance. Permission to IAM role attached to an instance can be changed, which will immediately impact the running instances; however, you cannot change the IAM role itself.

# Practice Assignment: EC2 Instance

To install WordPress on your windows server.

# Install and Deploy a webpage on EC2



You need to install and deploy a webpage on Amazon EC2 instance running on Linux server.

## **Prerequisites:**

- Amazon EC2 instance should be launched from Linux base AMI.
- Launch Windows t2.micro instance, which is available on free tier (note that when you launch a regular instance to deploy WordPress website, you will incur the standard Amazon EC2 usage fees for the instance)
- Security group in which instance is launched has port 80(HTTP), 443(HTTPS), and 22(SSH) open for all inbound traffic. Ports 80 and 443 allow computers outside of the instance to connect with HTTP and HTTPS. WordPress site can't be accessed from outside the instance if these ports are not opened.
- Ensure that you Connect to your instance.

## **Task:**

Install a webpage on your Linux server.



## QUIZ

1

Your are deploying an application that requires very high random I/O performance, and high IOPS at a low cost. What kind of EC2 instance type you should select for this application?

- a. t2.micro
- b. i2.2xlarge
- c. m4.large
- d. c4.large





## QUIZ

1

Your are deploying an application that requires very high random I/O performance, and high IOPS at a low cost. What kind of EC2 instance type you should select for this application?

- a. t2.micro
- b. i2.2xlarge
- c. m4.large
- d. c4.large



The correct answer is **i2.2xlarge**

**Explanation:** When you see keywords such as “high random I/O” performance and “high ops,” you can easily figure out that i2 is the right option.

## QUIZ

### 2

You have an audio transcoding application running on Amazon EC2. Each time an instance runs, it polls a queue to find out which audio should be transcoded and then runs a transcoding process. When the process is interrupted, the audio will be transcoded by another instance based on the queuing system. You have a large backlog of audios that need to be transcoded, and you need additional instances to reduce the backlog. These instances will only be required until the backlog is reduced. Which cost effective instance type can be used to reduce the backlog in the shortest time?

- a. On-demand instances
- b. Dedicated instances
- c. Spot instances
- d. Reserved Instances



## QUIZ

2

You have an audio transcoding application running on Amazon EC2. Each time an instance runs, it polls a queue to find out which audio should be transcoded and then runs a transcoding process. When the process is interrupted, the audio will be transcoded by another instance based on the queuing system. You have a large backlog of audios that need to be transcoded, and you need additional instances to reduce the backlog. These instances will only be required until the backlog is reduced. Which cost effective instance type can be used to reduce the backlog in the shortest time?

- a. On-demand instances
- b. Dedicated instances
- c. Spot instances
- d. Reserved Instances



The correct answer is **Spot instances**

Key points in this question are “When the process is interrupted, the audio will be transcoded by another instance based on the queuing” and “These instances will only be required until the backlog is reduced” and “cost effective instance type.” Spot instances are a great choice if your application is flexible enough about when it can run or be interrupted.

## QUIZ

3

What charges will apply when using a Amazon EBS shared snapshot, which you are sharing?

- a. No charges are incurred when you share or you make a copy of the shared snapshot.
- b. You and the user will both be charged when you make a copy of the snapshot.
- c. You will not be charged when others users make a copy of your shared snapshot.
- d. You will be charged each time others make a copy of your snapshot.



## QUIZ

3

What charges will apply when using a Amazon EBS shared snapshot, which you are sharing?

- a. No charges are incurred when you share or you make a copy of the shared snapshot.
- b. You and the user will both be charged when you make a copy of the snapshot.
- c. You will not be charged when others users make a copy of your shared snapshot.
- d. You will be charged each time others make a copy of your snapshot.



The correct answer is **You will not be charged when other users make a copy of your shared snapshot**

**Explanation:** By default, snapshots are private and available to the AWS account where it is created. You can either provide access to other AWS accounts to use your snapshot or make a copy for them. When others make a copy and use, they have to pay for it.

## QUIZ

5

Which of the following components are used for dynamic scaling?

- a. Alarms
- b. Timer
- c. Policies
- d. ELB



## QUIZ

5

Which of the following components are used for dynamic scaling?

- a. Alarms
- b. Timer
- c. Policies
- d. ELB



The correct answer is **Alarms & Policies**

**Dynamic scaling helps you to scale dynamically and respond to changing demand. Auto-scaling uses a combination of alarms and policies to decide if the scaling conditions are met. Alarms are generated when a single metric breaches a defined threshold. Policy refers to a set of instructions to the scaling group on how to react or respond to alarms.**

## QUIZ

6

How much data can you store with Amazon EFS?

- a. You need to specify the size of EFS at the time of creation
- b. You don't need to specify the upfront size, it scales automatically as load increases
- c. You can choose any one of the above two options while creating EFS
- d. None of the above





## QUIZ

6

How much data can you store with Amazon EFS?

- a. You need to specify the size of EFS at the time of creation
- b. You don't need to specify the upfront size, it scales automatically as load increases
- c. You can choose any one of the above two options while creating EFS
- d. None of the above



The correct answer is **You don't need to specify the upfront size, it scales automatically as load increases**

**Dynamic scaling helps you to scale dynamically and respond to changing demand. Auto-scaling uses a combination of alarms and policies to decide if the scaling conditions are met. Alarms are generated when a single metric breaches a defined threshold. Policy refers to a set of instructions to the scaling group on how to react or respond to alarms.**

## Key Takeaways

---

- Each type of instance offers a different size of compute, memory, and storage features and capabilities.
- HVM-based VMs can take advantage of hardware extensions that have features such as enhanced networking and clustering.
- Reserved instances are ideal for applications with predictable usage or steady state.
- Spot instances are good for applications that can be stopped at any time and resumed without any data or application state loss.
- The data in an instance store will exist as long as the associated instance exists.
- When the instance is stopped or rebooted, the data in EBS volume still remains. Root volume can be retained by deselecting the “Delete On Termination” check box.
- Understand the HTTP response code provided by the ELBs to identify the source of errors.



**This concludes 'Elastic Compute Cloud (EC2).'**

The next lesson is "S3"