

AWS Elastic Compute Cloud (EC2)

Khalid Bin Sattar

DevOps Specialist, 3Shape A/S





Agenda

- Introduction to Amazon EC2.
- Features of Amazon EC2.
- Basic Concepts of EC2.
 - 1. Instances and AMIs
 - 2. Regions and Availability Zones.
 - 3. Instance Types.
 - 4. Tags.
- Networking and Security.
- Amazon EC2 Advance Features.
- Amazon Elastic Block Store.
- Amazon EC2 Limitation.
- Amazon EC2 Pricing.
- Demo



Introduction to Amazon EC2

What Is Amazon EC2?

Amazon Elastic Compute Cloud (Amazon EC2) provides scalable computing capacity in the Amazon Web Services (AWS) cloud. Using Amazon EC2 eliminates your need to invest in hardware up front, so you can develop and deploy applications faster. You can use Amazon EC2 to launch as many or as few virtual servers as you need, configure security and networking, and manage storage. Amazon EC2 enables you to scale up or down to handle changes in requirements or spikes in popularity, reducing your need to forecast traffic.



Features of Amazon EC2

Amazon EC2 provides the following features:

- Virtual computing environments, known as instances
- Preconfigured templates for your instances, known as Amazon Machine Images (AMIs), that package the bits you need for your server (including the operating system and additional software)
- Various configurations of CPU, memory, storage, and networking capacity for your instances, known as instance types
- Secure login information for your instances using key pairs (AWS stores the public key, and you store the private key in a secure place)
- Storage volumes for temporary data that's deleted when you stop or terminate your instance, known as instance store volumes
- Persistent storage volumes for your data using Amazon Elastic Block Store (Amazon EBS), known as Amazon EBS volumes



Features of Amazon EC2

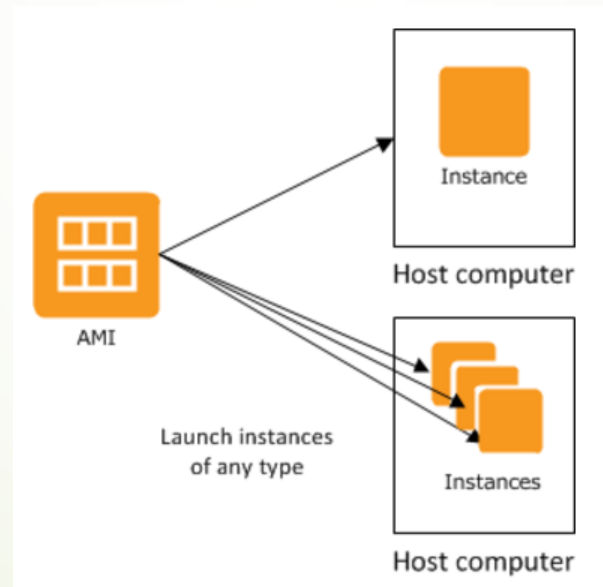
Amazon EC2 provides the following features:

- Multiple physical locations for your resources, such as instances and Amazon EBS volumes, known as regions and Availability Zones
- A firewall that enables you to specify the protocols, ports, and source IP ranges that can reach your instances using security groups
- Static IPv4 addresses for dynamic cloud computing, known as Elastic IP addresses
- Metadata, known as tags, that you can create and assign to your Amazon EC2 resources
- Virtual networks you can create that are logically isolated from the rest of the AWS cloud, and that you can optionally connect to your own network, known as virtual private clouds (VPCs)

Basic Concepts of Amazon EC2

► Instances and AMIs:-

An Amazon Machine Image (AMI) is a template that contains a software configuration (for example, an operating system, an application server, and applications). From an AMI, you launch an instance, which is a copy of the AMI running as a virtual server in the cloud. You can launch multiple instances of an AMI, as shown in the following figure.





Basic Concepts of Amazon EC2

► Instances:-

You can launch different types of instances from a single AMI. An instance type essentially determines the hardware of the host computer used for your instance. Each instance type offers different compute and memory capabilities. Select an instance type based on the amount of memory and computing power that you need for the application or software that you plan to run on the instance.

After you launch an instance, it looks like a traditional host, and you can interact with it as you would any computer. You have complete control of your instances; you can use `sudo` to run commands that require root privileges.

Your AWS account has a limit on the number of instances that you can have running.



Basic Concepts of Amazon EC2

➤ Storage for Your Instance:-

The root device for your instance contains the image used to boot the instance.

Your instance may include local storage volumes, known as instance store volumes, which you can configure at launch time with block device mapping.

After these volumes have been added to and mapped on your instance, they are available for you to mount and use. If your instance fails, or if your instance is stopped or terminated, the data on these volumes is lost; therefore, these volumes are best used for temporary data. For important data, you should use a replication strategy across multiple instances in order to keep your data safe, or store your persistent data in Amazon S3 or Amazon EBS volumes.



Basic Concepts of Amazon EC2

➤ Block Device Mapping:-

Each instance that you launch has an associated root device volume, either an Amazon EBS volume or an instance store volume. You can use block device mapping to specify additional EBS volumes or instance store volumes to attach to an instance when it's launched. You can also attach additional EBS volumes to a running instance.

However, the only way to attach instance store volumes to an instance is to use block device mapping to attach them as the instance is launched.

Basic Concepts of Amazon EC2

► Block Device Mapping Concepts:-

A block device is a storage device that moves data in sequences of bytes or bits (blocks). These devices support random access and generally use buffered I/O. Examples include hard disks, CD-ROM drives, and flash drives. A block device can be physically attached to a computer or accessed remotely as if it were physically attached to the computer. Amazon EC2 supports two types of block devices:

1. Instance store volumes (virtual devices whose underlying hardware is physically attached to the host computer for the instance)
2. EBS volumes (remote storage devices)

A block device mapping defines the block devices (instance store volumes and EBS volumes) to attach to an instance. You can specify a block device mapping as part of creating an AMI so that the mapping is used by all instances launched from the AMI. Alternatively, you can specify a block device mapping when you launch an instance, so this mapping overrides the one specified in the AMI from which you launched the instance.

Basic Concepts of Amazon EC2

➤ AMI Block Device Mapping:-

Each AMI has a block device mapping that specifies the block devices to attach to an instance when it is launched from the AMI. An AMI that Amazon provides includes a root device only. To add more block devices to an AMI, you must create your own AMI.

1. Specifying a Block Device Mapping for an AMI:-

- There are two ways to specify volumes in addition to the root volume when you create an AMI. If you've already attached volumes to a running instance before you create an AMI from the instance, the block device mapping for the AMI includes those same volumes.
- For an EBS-backed AMI, you can add EBS volumes and instance store volumes using a block device mapping. For an instance store-backed AMI, you can add instance store volumes only by modifying the block device mapping entries in the image manifest file when registering the image.

2. Viewing the EBS Volumes in an AMI Block Device Mapping:-

- You can easily enumerate the EBS volumes in the block device mapping for an AMI.

Basic Concepts of Amazon EC2

➤ Instance Block Device Mapping:-

By default, an instance that you launch includes any storage devices specified in the block device mapping of the AMI from which you launched the instance. You can specify changes to the block device mapping for an instance when you launch it, and these updates overwrite or merge with the block device mapping of the AMI.

Updating the Block Device Mapping when Launching an Instance:-

- You can add EBS volumes and instance store volumes to an instance when you launch it. Note that updating the block device mapping for an instance doesn't make a permanent change to the block device mapping of the AMI from which it was launched.

Basic Concepts of Amazon EC2

Stopping, Starting, and Terminating Instances:-

➤ Stopping an instance

When an instance is stopped, the instance performs a normal shutdown, and then transitions to a stopped state. All of its Amazon EBS volumes remain attached, and you can start the instance again at a later time.

You are not charged for additional instance hours while the instance is in a stopped state. A full instance hour will be charged for every transition from a stopped state to a running state, even if this happens multiple times within a single hour. If the instance type was changed while the instance was stopped, you will be charged the rate for the new instance type after the instance is started. All of the associated Amazon EBS usage of your instance, including root device usage, is billed using typical Amazon EBS prices.

When an instance is in a stopped state, you can attach or detach Amazon EBS volumes. You can also create an AMI from the instance, and you can change the kernel, RAM disk, and instance type.

Basic Concepts of Amazon EC2

Stopping, Starting, and Terminating Instances:-

➤ Terminating an instance

When an instance is terminated, the instance performs a normal shutdown, then the attached Amazon EBS volumes are deleted unless the volume's *deleteOnTermination* attribute is set to *false*. The instance itself is also deleted, and you can't start the instance again at a later time.

To prevent accidental termination, you can disable instance termination. If you do so, ensure that the *disableApiTermination* attribute is set to true for the instance.



Basic Concepts of Amazon EC2

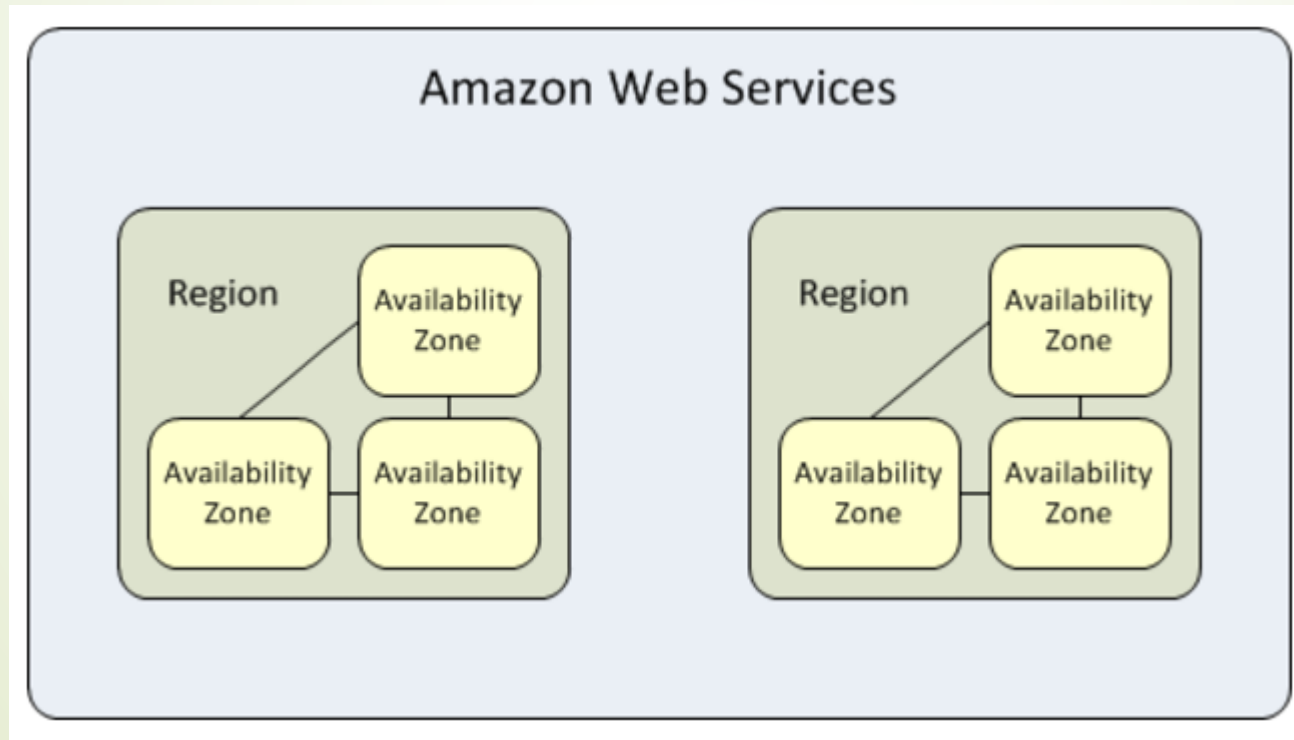
Regions and Availability Zones:-

- Amazon EC2 is hosted in multiple locations world-wide. These locations are composed of regions and Availability Zones. Each region is a separate geographic area. Each region has multiple, isolated locations known as Availability Zones. Amazon EC2 provides you the ability to place resources, such as instances, and data in multiple locations. Resources aren't replicated across regions unless you do so specifically.
- Amazon operates state-of-the-art, highly-available data centers.

Basic Concepts of Amazon EC2

Region and Availability Zone Concepts:-

Each region is completely independent. Each Availability Zone is isolated, but the Availability Zones in a region are connected through low-latency links. The following diagram illustrates the relationship between regions and Availability Zones.



Basic Concepts of Amazon EC2

Region:-

- Each region is completely independent.
- Each Amazon EC2 region is designed to be completely isolated from the other Amazon EC2 regions. This achieves the greatest possible fault tolerance and stability.
- When you view your resources, you'll only see the resources tied to the region you've specified. This is because regions are isolated from each other, and AWS don't replicate resources across regions automatically.
- When you launch an instance, you must select an AMI that's in the same region. If the AMI is in another region, you can copy the AMI to the region you're using.

Basic Concepts of Amazon EC2

Available Regions:-

Code	Name
us-east-1	US East (N. Virginia)
us-east-2	US East (Ohio)
us-west-1	US West (N. California)
us-west-2	US West (Oregon)
ca-central-1	Canada (Central)
eu-west-1	EU (Ireland)
eu-central-1	EU (Frankfurt)

Code	Name
eu-west-2	EU (London)
ap-northeast-1	Asia Pacific (Tokyo)
ap-northeast-2	Asia Pacific (Seoul)
ap-southeast-1	Asia Pacific (Singapore)
ap-southeast-2	Asia Pacific (Sydney)
ap-south-1	Asia Pacific (Mumbai)
sa-east-1	South America (São Paulo)



Basic Concepts of Amazon EC2

Availability Zones:-

- When you launch an instance, you can select an Availability Zone or AWS will choose one for you. If you distribute your instances across multiple Availability Zones and one instance fails, you can design your application so that an instance in another Availability Zone can handle requests.
- You can also use Elastic IP addresses to mask the failure of an instance in one Availability Zone by rapidly remapping the address to an instance in another Availability Zone.
- An Availability Zone is represented by a region code followed by a letter identifier; for example, us-east-1a. To ensure that resources are distributed across the Availability Zones for a region, we independently map Availability Zones to identifiers for each account. For example, your Availability Zone us-east-1a might not be the same location as us-east-1a for another account. There's no way for you to coordinate Availability Zones between accounts.



Basic Concepts of Amazon EC2

Instance Types:-

- When you launch an instance, the instance type that you specify determines the hardware of the host computer used for your instance. Each instance type offers different compute, memory, and storage capabilities and are grouped in instance families based on these capabilities. Select an instance type based on the requirements of the application or software that you plan to run on your instance.
- Amazon EC2 dedicates some resources of the host computer, such as CPU, memory, and instance storage, to a particular instance. Amazon EC2 shares other resources of the host computer, such as the network and the disk subsystem, among instances.
- Each instance type provides higher or lower minimum performance from a shared resource. For example, instance types with high I/O performance have a larger allocation of shared resources.

Basic Concepts of Amazon EC2

Available Instance Types:-

Instance Family	Current Generation Instance Types
General purpose	t2.nano t2.micro t2.small t2.medium t2.large t2.xlarge t2.2xlarge m4.large m4.xlarge m4.2xlarge m4.4xlarge m4.10xlarge m4.16xlarge m3.medium m3.large m3.xlarge m3.2xlarge
Compute optimized	c4.large c4.xlarge c4.2xlarge c4.4xlarge c4.8xlarge c3.large c3.xlarge c3.2xlarge c3.4xlarge c3.8xlarge
Memory optimized	r3.large r3.xlarge r3.2xlarge r3.4xlarge r3.8xlarge r4.large r4.xlarge r4.2xlarge r4.4xlarge r4.8xlarge r4.16xlarge x1.16xlarge x1.32xlarge
Storage optimized	d2.xlarge d2.2xlarge d2.4xlarge d2.8xlarge i2.xlarge i2.2xlarge i2.4xlarge i2.8xlarge i3.large i3.xlarge i3.2xlarge i3.4xlarge i3.8xlarge i3.16xlarge
Accelerated computing	f1.2xlarge f1.16xlarge p2.xlarge p2.8xlarge p2.16xlarge g2.2xlarge g2.8xlarge g3.4xlarge g3.8xlarge g3.16xlarge



Basic Concepts of Amazon EC2

Instance Types:-

T2 Instances:-

- T2 instances are designed to provide moderate baseline performance and the capability to burst to significantly higher performance as required by your workload. They are intended for workloads that don't use the full CPU often or consistently, but occasionally need to burst. T2 instances are well suited for general purpose workloads, such as web servers, developer environments, and small databases.
- If your account is less than 12 months old, you can use a t2.micro instance for free within certain usage limits.
- You must launch your T2 instances into a virtual private cloud (VPC); they are not supported on the EC2-Classic platform.
- You cannot change the instance type of an existing instance in EC2-Classic to a T2 instance type.
- There is a limit on the total number of instances that you can launch in a region, and there are additional limits on some instance types. By default, you can run up to 20 T2 instances simultaneously.

Basic Concepts of Amazon EC2

Instance Types:-

Compute Optimized Instances:-

Compute optimized instances are ideal for compute-bound applications that benefit from high performance processors. They are well suited for the following applications:

- Batch processing workloads
- Media transcoding
- High-traffic web servers, massively multiplayer online (MMO) gaming servers, and ad serving engines
- High performance computing (HPC) and other compute-intensive applications.

Compute Instance Features:-

	VPC only	EBS only	SSD volumes	Placement group	HVM only	Enhanced networking
C3			Yes	Yes		Intel 82599 VF
C4	Yes	Yes		Yes	Yes	Intel 82599 VF

Basic Concepts of Amazon EC2

Instance Types:-

Memory Optimized Instances:-

Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory.

R4 Instances:-R4 instances are well suited for the following applications:

- High performance relational (MySQL) and NoSQL (MongoDB, Cassandra) databases.
- In-memory databases using optimized data storage formats and analytics for business intelligence (for example, SAP HANA).
- Applications performing real-time processing of big unstructured data (financial services, Hadoop/Spark clusters).
- High-performance computing (HPC) and Electronic Design Automation (EDA) applications.



Basic Concepts of Amazon EC2

Instance Types:-

Memory Optimized Instances:-

X1 Instances:-X1 instances are well suited for the following applications:

- In-memory databases such SAP HANA, including SAP-certified support for Business Suite S/4HANA, Business Suite on HANA (SoH), Business Warehouse on HANA (BW), and Data Mart Solutions on HANA.
- Big-data processing engines such as Apache Spark or Presto.
- High-performance computing (HPC) applications.

Basic Concepts of Amazon EC2

Instance Types:-

Memory Optimized Instances:-

R3 Instances:-R3 instances are well suited for the following applications:

- High performance relational (MySQL) and NoSQL (MongoDB, Cassandra) databases.
- In-memory analytics.
- Genome assembly and analysis.
- Enterprise applications (for example, Microsoft SharePoint).

Instance Features:-

	VPC only	EBS only	SSD volumes	Placement group	Enhanced networking
R3			Yes	Yes	Intel 82599 VF
R4	Yes	Yes		Yes	ENA
X1	Yes		Yes	Yes	ENA

Basic Concepts of Amazon EC2

Instance Types:-

Storage Optimized Instances:-

Storage optimized instances are designed for workloads that require high sequential read and write access to very large data sets on local storage. They are optimized to deliver tens of thousands of low-latency, random I/O operations per second (IOPS) to applications.

D2 Instances:-D2 instances are well suited for the following applications:

- Massive parallel processing (MPP) data warehouse.
- MapReduce and Hadoop distributed computing.
- Log or data processing applications.



Basic Concepts of Amazon EC2

Instance Types:-

Storage Optimized Instances:-

I2 Instances:-I2 instances are well suited for the following applications:

- NoSQL databases.
- Clustered databases.
- Online transaction processing (OLTP) systems.

Basic Concepts of Amazon EC2

Instance Types:-

Storage Optimized Instances:-

I3 Instances:- I3 instances are well suited for the following applications:

- High frequency online transaction processing (OLTP) systems
- Relational databases & NoSQL databases.
- Cache for in-memory databases (for example, Redis).
- Data warehousing applications.
- Low latency Ad-Tech serving applications.

Storage Instance Features:-

	VPC only	SSD volumes	Placement group	Enhanced networking
D2			Yes	Intel 82599 VF
I2		SATA	Yes	Intel 82599 VF
I3	Yes	NVMe	Yes	ENA



Basic Concepts of Amazon EC2

Instance Types:-

Linux Accelerated Computing Instances:-

- If you require high processing capability, you'll benefit from using accelerated computing instances, which provide access to hardware-based compute accelerators such as Graphics Processing Units (GPUs) or Field Programmable Gate Arrays (FPGAs). Accelerated computing instances enable more parallelism for higher throughput on compute-intensive workloads.
- GPU-based instances provide access to NVIDIA GPUs with thousands of compute cores. You can use GPU-based accelerated computing instances to accelerate scientific, engineering, and rendering applications by leveraging the CUDA or Open Computing Language (OpenCL) parallel computing frameworks. You can also use them for graphics applications, including game streaming, 3-D application streaming, and other graphics workloads.
- FPGA-based instances provide access to large FPGAs with millions of parallel system logic cells. You can use FPGA-based accelerated computing instances to accelerate workloads such as genomics, financial analysis, real-time video processing, big data analysis, and security workloads by leveraging custom hardware accelerations.

Basic Concepts of Amazon EC2

Instance Purchasing Options:-

- On-Demand instances — Pay, by the hour, for the instances that you launch.
- Reserved Instances — Purchase, at a significant discount, instances that are always available, for a term from one to three years.
- Scheduled Instances — Purchase instances that are always available on the specified recurring schedule, for a one-year term.
- Spot instances — Bid on unused instances, which can run as long as they are available and your bid is above the Spot price, at a significant discount.
- Dedicated hosts — Pay for a physical host that is fully dedicated to running your instances, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs.
- Dedicated instances — Pay, by the hour, for instances that run on single-tenant hardware.



Basic Concepts of Amazon EC2

Determining the Instance Lifecycle:-

The lifecycle of an instance starts when it is launched and ends when it is terminated. The purchasing option that you choose effects the lifecycle of the instance. For example, an On-Demand instance runs when you launch it and ends when you terminate it. A Spot instance runs as long as its capacity is available and your bid price is higher than the Spot price. You can launch a Scheduled Instance during its scheduled time period; Amazon EC2 launches the instances and then terminates them three minutes before the time period ends.



Basic Concepts of Amazon EC2

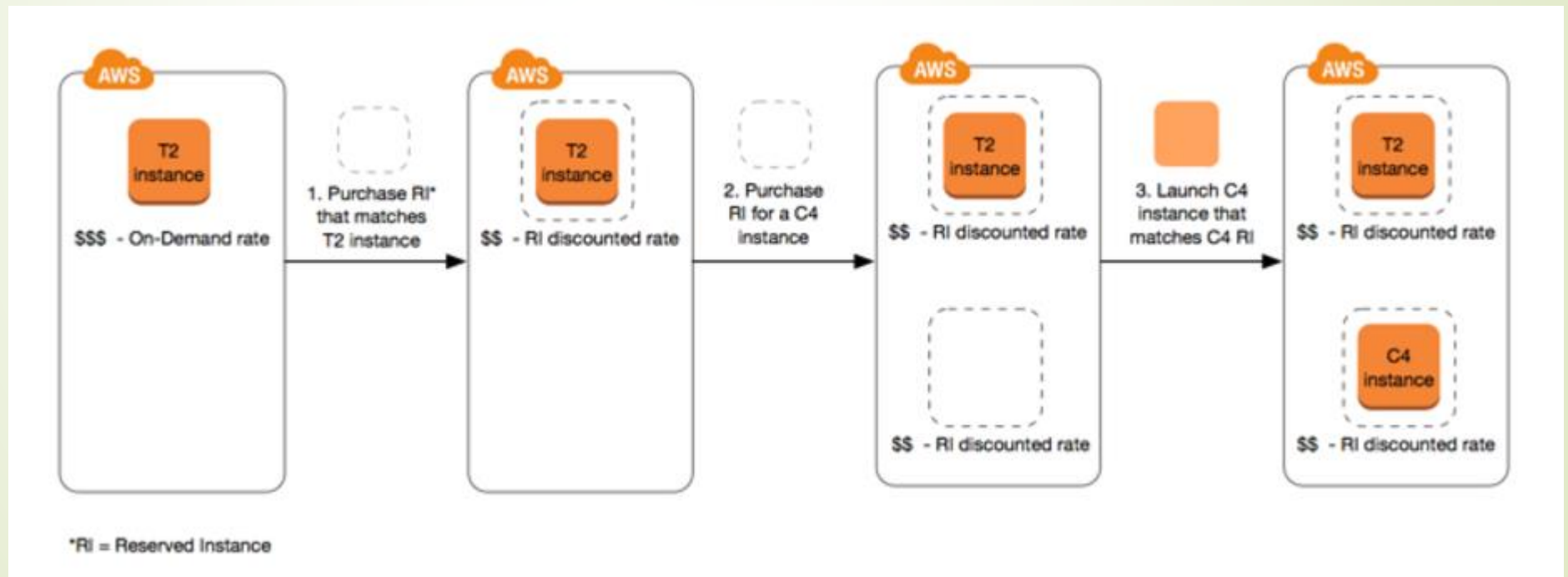
Reserved Instances:-

- Reserved Instances provide you with a significant discount compared to On-Demand Instance pricing. Reserved Instances are not physical instances, but rather a billing discount applied to the use of On-Demand Instances in your account. These On-Demand Instances must match certain attributes in order to benefit from the billing discount.
- Reserved Instances do not renew automatically; when they expire, you can continue using the EC2 instance without interruption, but you are charged On-Demand rates.

Basic Concepts of Amazon EC2

Reserved Instances:-

The following diagram shows a basic overview of purchasing and using Reserved Instances.





Basic Concepts of Amazon EC2

Reserved Instance Limits:-

- You are limited to purchasing 20 Reserved Instances per Availability Zone, per month, plus 20 regional Reserved Instances. Therefore, in a region that has three Availability Zones, you can purchase 80 Reserved Instances in total: 20 per Availability Zone (60) plus 20 regional Reserved Instances.
- Reserved Instances that are purchased for a specific Availability Zone (zonal Reserved Instances) allow you to launch as many instances that are covered by the zonal Reserved Instances, even if this results in you exceeding your On-Demand Instance limit.
- For example, your running On-Demand Instance limit is 20, and you are currently running 18 On-Demand Instances. You have five unused zonal Reserved Instances. You can launch two more On-Demand Instances with any specifications, and you can launch five instances that exactly match the specifications of your zonal Reserved Instances; giving you a total of 25 instances.



Basic Concepts of Amazon EC2

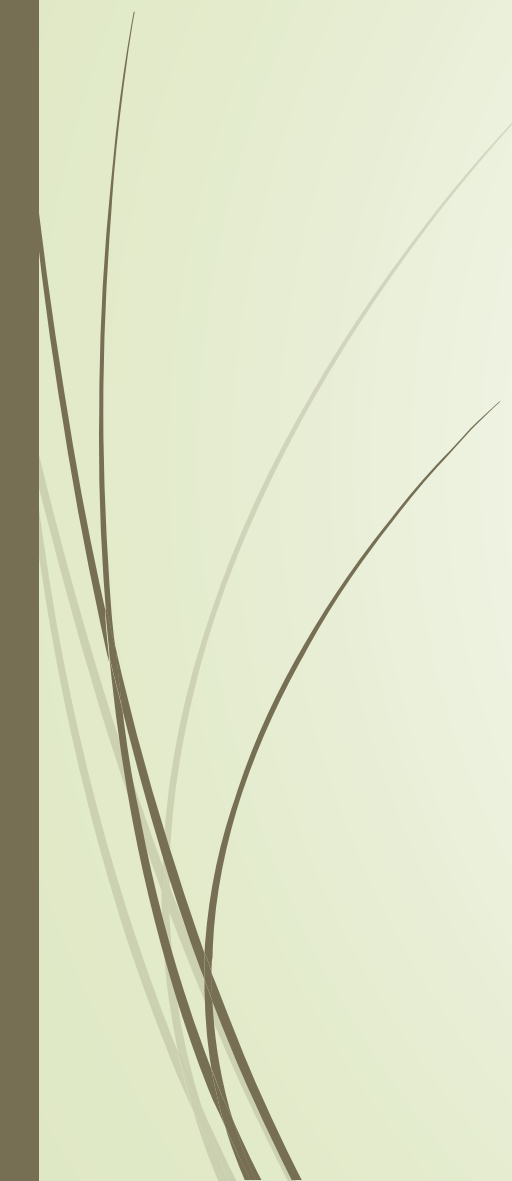
Scheduled Reserved Instances:-

- 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. You reserve the capacity in advance, so that you know it is available when you need it. You pay for the time that the instances are scheduled, even if you do not use them.
- Scheduled Instances are a good choice for workloads that do not run continuously, but do run on a regular schedule. For example, you can use Scheduled Instances for an application that runs during business hours or for batch processing that runs at the end of the week.
- If you require a capacity reservation on a continuous basis, Reserved Instances might meet your needs and decrease costs.



Basic Concepts of Amazon EC2

How Scheduled Instances Work:-

- Amazon EC2 sets aside pools of EC2 instances in each Availability Zone for use as Scheduled Instances. Each pool supports a specific combination of instance type, operating system, and network (EC2-Classic or EC2-VPC).
 - To get started, you must search for an available schedule. You can search across multiple pools or a single pool. After you locate a suitable schedule, purchase it.
 - You must launch your Scheduled Instances during their scheduled time periods, using a launch configuration that matches the following attributes of the schedule that you purchased: instance type, Availability Zone, network, and platform. When you do so, Amazon EC2 launches EC2 instances on your behalf, based on the specified launch specification.
- 

Basic Concepts of Amazon EC2

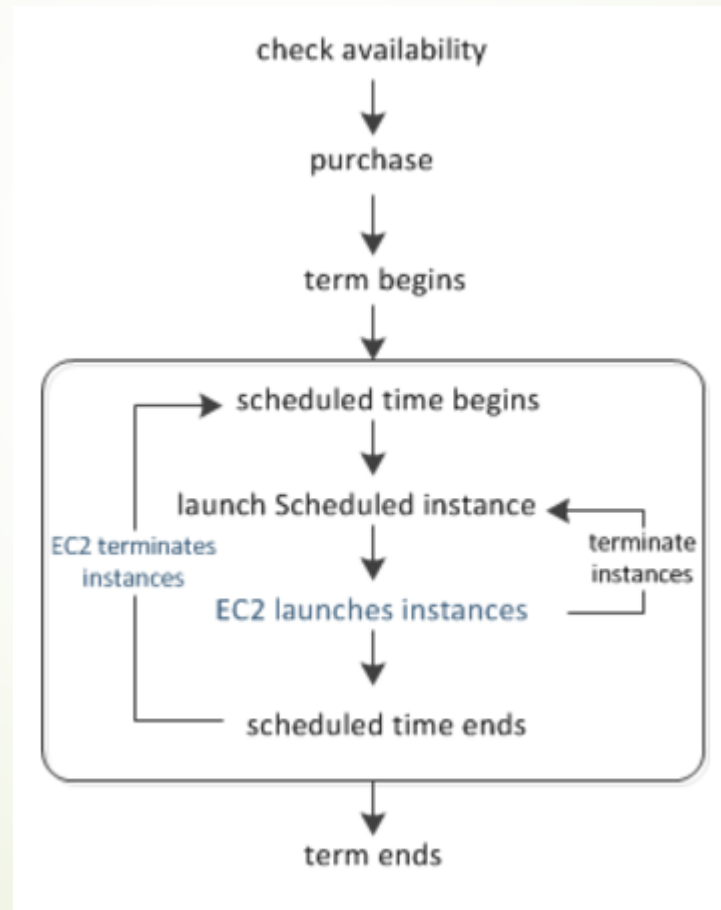
How Scheduled Instances Work:-

- Amazon EC2 must ensure that the EC2 instances have terminated by the end of the current scheduled time period so that the capacity is available for any other Scheduled Instances it is reserved for. Therefore, Amazon EC2 terminates the EC2 instances three minutes before the end of the current scheduled time period.
- You can't stop or reboot Scheduled Instances, but you can terminate them manually as needed. If you terminate a Scheduled Instance before its current scheduled time period ends, you can launch it again after a few minutes.

Basic Concepts of Amazon EC2

How Scheduled Instances Work:-

The following diagram illustrates the lifecycle of a Scheduled Instance.





Basic Concepts of Amazon EC2

Scheduled Instance Limits:-

Scheduled Instances are subject to the following limits:

- Scheduled Instance only supported instance types: C3, C4, M4, and R3.
- The required term is 365 days (one year).
- The minimum required utilization is 1,200 hours per year.
- You can purchase a Scheduled Instance up to three months in advance.



Basic Concepts of Amazon EC2

Spot Instances:-

- Spot instances enable you to bid on unused EC2 instances, which can lower your Amazon EC2 costs significantly. The hourly price for a Spot instance (of each instance type in each Availability Zone) is set by Amazon EC2, and fluctuates depending on the supply of and demand for Spot instances. Your Spot instance runs whenever your bid exceeds the current market price.
- Spot instances are a cost-effective choice if you can be flexible about when your applications run and if your applications can be interrupted. For example, Spot instances are well-suited for data analysis, batch jobs, background processing, and optional tasks.
- The key differences between Spot instances and On-Demand instances are that Spot instances might not start immediately, the hourly price for Spot instances varies based on demand, and Amazon EC2 can terminate an individual Spot instance as the hourly price for, or availability of, Spot instances changes.

Basic Concepts of Amazon EC2

Concepts Spot instances:-

- Spot instance pool—A set of unused EC2 instances with the same instance type, operating system, Availability Zone, and network platform (EC2-Classic or EC2-VPC).
- Spot price—The current market price of a Spot instance per hour, which is set by Amazon EC2 based on the last fulfilled bid. You can also retrieve the Spot price history.
- Spot instance request (or Spot bid)—Provides the maximum price (bid price) that you are willing to pay per hour for a Spot instance. When your bid price exceeds the Spot price, Amazon EC2 fulfills your request. Note that a Spot instance request is either one-time or persistent. Amazon EC2 automatically resubmits a persistent Spot request after the Spot instance associated with the request is terminated. Your Spot instance request can optionally specify a duration for the Spot instances.

Basic Concepts of Amazon EC2

Concepts Spot instances:-

- Spot fleet—A set of Spot instances that is launched based on criteria that you specify. The Spot fleet selects the Spot instance pools that meet your needs and launches Spot instances to meet the target capacity for the fleet. By default Spot fleets are set to maintain target capacity by launching replacement instances after Spot instances in the fleet are terminated. They can also be submitted as a one-time request which does not persist once instances have been terminated.
- Spot instance interruption—Amazon EC2 terminates your Spot instance when the Spot price exceeds your bid price or there are no longer any unused EC2 instances. Amazon EC2 marks the Spot instance for termination and provides a Spot instance termination notice, which gives the instance a two-minute warning before it terminates.
- Bid status—Provides detailed information about the current state of your Spot bid.

Basic Concepts of Amazon EC2

How Spot Instances Work:-

- To use Spot instances, create a Spot instance request or a Spot fleet request. The request includes the maximum price that you are willing to pay per hour per instance (your bid price), and other constraints such as the instance type and Availability Zone. If your bid price is greater than the current Spot price for the specified instance, and the specified instance is available, your request is fulfilled immediately. Otherwise, the request is fulfilled whenever the Spot price falls below your bid price or the specified instance becomes available. Spot instances run until you terminate them or until Amazon EC2 must terminate them (also known as a Spot instance interruption).
- When you use Spot instances, you must be prepared for interruptions. Amazon EC2 can interrupt your Spot instance when the Spot price rises above your bid price, when the demand for Spot instances rises, or when the supply of Spot instances decreases.
- When Amazon EC2 marks a Spot instance for termination, it provides a Spot instance termination notice, which gives the instance a two-minute warning before it terminates. Note that you can't enable termination protection for Spot instances.
- Note that you can't stop and start an Amazon EBS-backed instance if it is a Spot instance, but you can reboot or terminate it.



Basic Concepts of Amazon EC2

Spot Instance Limits:-

1. Unsupported Instance Types:-

- T2 and HS1 Instance Types are not supported for Spot Instance.

2. Spot Request Limits:-

- By default, there is an account limit of 20 Spot instances per region. If you terminate your Spot instance but do not cancel the request, the request counts against this limit until Amazon EC2 detects the termination and closes the request.



Basic Concepts of Amazon EC2

Spot Instance Limits:-

3. Spot Bid Price Limit:-

- The bid price limit for Spot instances is ten times the On-Demand price. This limit is designed to help you control costs.

4. Spot Fleet Limits:-

- The usual Amazon EC2 limits apply to instances launched by a Spot fleet, such as Spot bid price limits, instance limits, and volume limits.

Basic Concepts of Amazon EC2

Dedicated Hosts:-

- An Amazon EC2 Dedicated Host is a physical server with EC2 instance capacity fully dedicated to your use. Dedicated Hosts allow you to use your existing per-socket, per-core, or per-VM software licenses, including Windows Server, Microsoft SQL Server, SUSE, Linux Enterprise Server, and so on.

Differences between Dedicated Hosts and Dedicated Instances

- Dedicated Hosts and Dedicated Instances can both be used to launch Amazon EC2 instances onto physical servers that are dedicated for your use.
- There are no performance, security, or physical differences between Dedicated Instances and instances on Dedicated Hosts. However, Dedicated Hosts give you additional visibility and control over how instances are placed on a physical server.
- When you use Dedicated Hosts, you have control over instance placement on the host using the Host Affinity and Instance Auto-placement settings. With Dedicated Instances, you don't have control over which host your instance launches and runs on. If your organization wants to use AWS, but has an existing software license with hardware compliance requirements, this allows visibility into the host's hardware so you can meet those requirements.



Basic Concepts of Amazon EC2

Dedicated Hosts Limitations and Restrictions:-

- RHEL, SUSE Linux, and Windows AMIs offered by AWS or on the AWS Marketplace cannot be used with Dedicated Hosts
- Amazon EC2 instance auto recovery is not supported.
- The instances that run on a Dedicated Host can only be launched in a VPC.
- Host limits are independent from instance limits. Instances that you are running on Dedicated Hosts do not count towards your instance limits.
- Auto Scaling groups are not supported.
- Amazon RDS instances are not supported.
- The AWS Free Usage tier is not available for Dedicated Hosts.
- Instance placement control refers to managing instance launches onto Dedicated Hosts. Placement groups are not supported for Dedicated Hosts.



Basic Concepts of Amazon EC2

Dedicated Instances:-

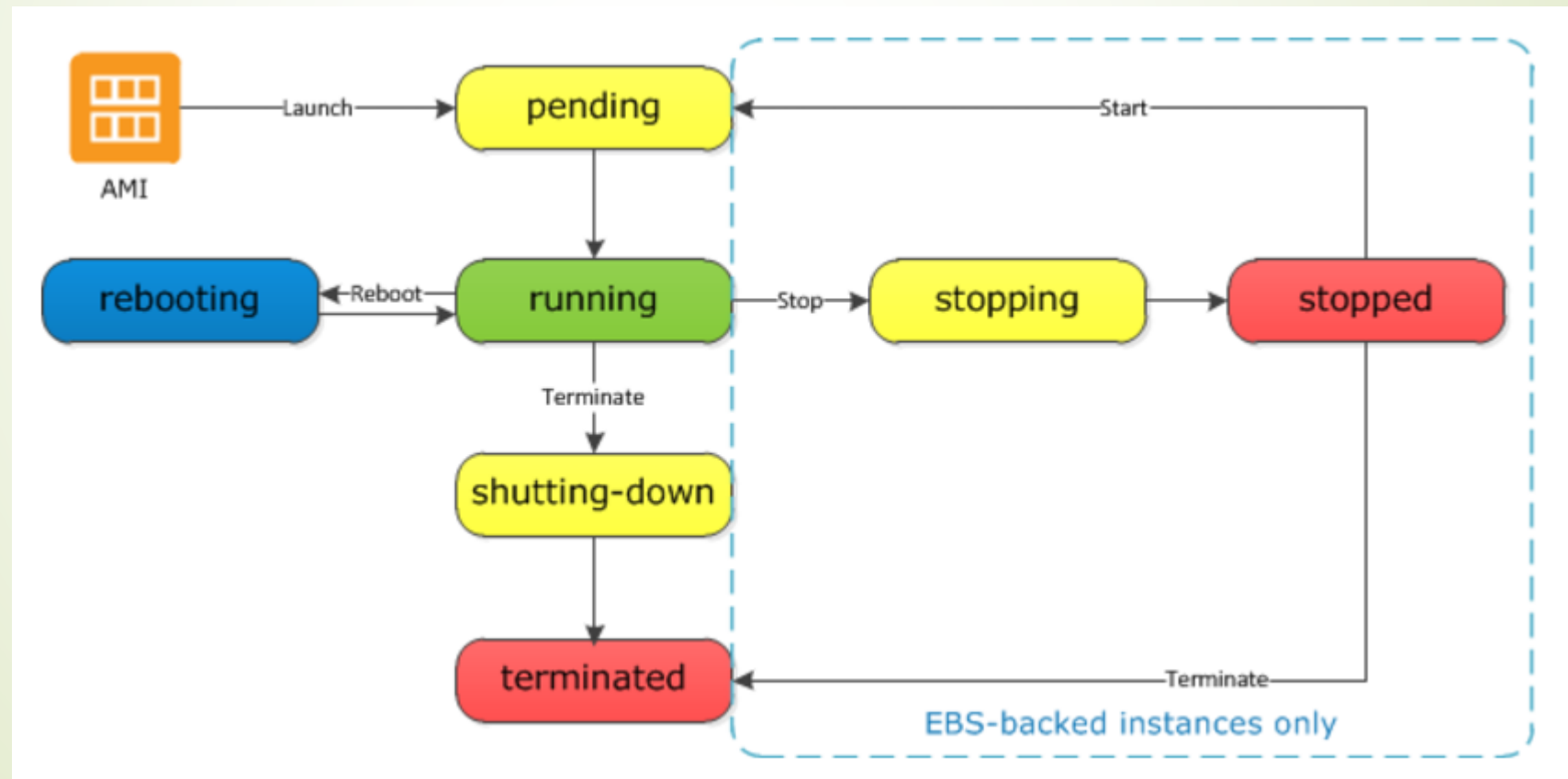
- Dedicated instances are Amazon EC2 instances that run in a virtual private cloud (VPC) on hardware that's dedicated to a single customer. Your Dedicated instances are physically isolated at the host hardware level from instances that belong to other AWS accounts. Dedicated instances may share hardware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

To create Dedicated instances, you can do the following:-

- Create the VPC with the instance tenancy set to dedicated (all instances launched into this VPC are Dedicated instances).
- Create the VPC with the instance tenancy set to default, and specify a tenancy of dedicated for any instances when you launch them.

Basic Concepts of Amazon EC2

Instance Lifecycle:- By working with Amazon EC2 to manage your instances from the moment you launch them through their termination, you ensure that your customers have the best possible experience with the applications or sites that you host on your instances.





Basic Concepts of Amazon EC2

Instance Launch:-

- When you launch an instance, it enters the pending state. The instance type that you specified at launch determines the hardware of the host computer for your instance. AWS use the Amazon Machine Image (AMI) you specified at launch to boot the instance. After the instance is ready for you, it enters the running state. You can connect to your running instance and use it the way that you'd use a computer sitting in front of you.
- As soon as your instance transitions to the running state, you're billed for each hour or partial hour that you keep the instance running; even if the instance remains idle and you don't connect to it.



Basic Concepts of Amazon EC2

Instance Stop and Start (Amazon EBS-backed instances only):-

- If your instance fails a status check or is not running your applications as expected, and if the root volume of your instance is an Amazon EBS volume, you can stop and start your instance to try to fix the problem.
- When you stop your instance, it enters the stopping state, and then the stopped state. We don't charge hourly usage or data transfer fees for your instance after you stop it, but we do charge for the storage for any Amazon EBS volumes. While your instance is in the stopped state, you can modify certain attributes of the instance, including the instance type.

Basic Concepts of Amazon EC2

Instance Stop and Start (Amazon EBS-backed instances only):-

- When you start your instance, it enters the pending state, and in most cases, AWS move the instance to a new host computer. (Your instance may stay on the same host computer if there are no problems with the host computer.) When you stop and start your instance, you'll lose any data on the instance store volumes on the previous host computer.
- If your instance is running in EC2-Classic, it receives a new private IPv4 address, which means that an Elastic IP address (EIP) associated with the private IPv4 address is no longer associated with your instance. If your instance is running in EC2-VPC, it retains its private IPv4 address, which means that an EIP associated with the private IPv4 address or network interface is still associated with your instance. If your instance has an IPv6 address, it retains its IPv6 address.
- Each time you transition an instance from stopped to running, AWS charge a full instance hour, even if these transitions happen multiple times within a single hour.



Basic Concepts of Amazon EC2

Instance Reboot:-

- You can reboot your instance using the Amazon EC2 console, a command line tool, and the Amazon EC2 API. AWS recommends that you use Amazon EC2 to reboot your instance instead of running the operating system reboot command from your instance.
- Rebooting an instance is equivalent to rebooting an operating system; the instance remains on the same host computer and maintains its public DNS name, private IP address, and any data on its instance store volumes. It typically takes a few minutes for the reboot to complete, but the time it takes to reboot depends on the instance configuration.
- Rebooting an instance doesn't start a new instance billing hour.

Basic Concepts of Amazon EC2

Instance Termination:-

- When you've decided that you no longer need an instance, you can terminate it. As soon as the status of an instance changes to shutting-down or terminated, you stop incurring charges for that instance.
- After you terminate an instance, it remains visible in the console for a short while, and then the entry is automatically deleted. You can also describe a terminated instance using the CLI and API. Resources (such as tags) are gradually disassociated from the terminated instance, therefore may no longer be visible on the terminated instance after a short while. You can't connect to or recover a terminated instance.
- Each Amazon EBS-backed instance supports the *InstanceInitiatedShutdownBehavior* attribute, which controls whether the instance stops or terminates when you initiate a shutdown from within the instance itself (for example, by using the shutdown command on Linux). The default behavior is to stop the instance. You can modify the setting of this attribute while the instance is running or stopped.
- Each Amazon EBS volume supports the *DeleteOnTermination* attribute, which controls whether the volume is deleted or preserved when you terminate the instance it is attached to. The default is to delete the root device volume and preserve any other EBS volumes.



Network and Security of Amazon EC2

Amazon EC2 Key Pairs:-

- Amazon EC2 uses public-key cryptography to encrypt and decrypt login information. Public-key cryptography uses a public key to encrypt a piece of data, such as a password, then the recipient uses the private key to decrypt the data. The public and private keys are known as a key pair.
- To log in to your instance, you must create a key pair, specify the name of the key pair when you launch the instance, and provide the private key when you connect to the instance. On a Linux instance, the public key content is placed in an entry within `~/.ssh/authorized_keys`. This is done at boot time and enables you to securely access your instance using the private key instead of a password.



Network and Security of Amazon EC2

Creating a Key Pair:-

- You can use Amazon EC2 to create your key pair.
- Alternatively, you could use a third-party tool and then import the public key to Amazon EC2.
- Each key pair requires a name. Be sure to choose a name that is easy to remember. Amazon EC2 associates the public key with the name that you specify as the key name.
- Amazon EC2 stores the public key only, and you store the private key. Anyone who possesses your private key can decrypt your login information, so it's important that you store your private keys in a secure place.
- The keys that Amazon EC2 uses are 2048-bit SSH-2 RSA keys. You can have up to five thousand key pairs per region.



Network and Security of Amazon EC2

Deleting Your Key Pair:-

When you delete a key pair, you are only deleting Amazon EC2's copy of the public key. Deleting a key pair doesn't affect the private key on your computer or the public key on any instances already launched using that key pair. You can't launch a new instance using a deleted key pair, but you can continue to connect to any instances that you launched using a deleted key pair, as long as you still have the private key (.pem) file.

How to Convert .pem Key Pair to .ppk key pair





Network and Security of Amazon EC2

Amazon EC2 Security Groups:-

- A security group acts as a virtual firewall that controls the traffic for one or more instances.
- When you launch an instance, you associate one or more security groups with the instance. You add rules to each security group that allow traffic to or from its associated instances.
- You can modify the rules for a security group at any time; the new rules are automatically applied to all instances that are associated with the security group.
- When AWS decide whether to allow traffic to reach an instance, AWS evaluate all the rules from all the security groups that are associated with the instance.



Network and Security of Amazon EC2

Security Group Rules:-

- By default, security groups allow all outbound traffic.
- You can't change the outbound rules for an EC2-Classic security group.
- Security group rules are always permissive; you can't create rules that deny access.
- Security groups are stateful — if you send a request from your instance, the response traffic for that request is allowed to flow in regardless of inbound security group rules. For VPC security groups, this also means that responses to allowed inbound traffic are allowed to flow out, regardless of outbound rules.
- You can add and remove rules at any time. Your changes are automatically applied to the instances associated with the security group after a short period.

Network and Security of Amazon EC2

Security Group Rules:-

For each rule, you specify the following:

- Protocol: The protocol to allow. The most common protocols are 6 (TCP) 17 (UDP), and 1 (ICMP).
- Port range : For TCP, UDP, or a custom protocol, the range of ports to allow. You can specify a single port number (for example, 22), or range of port numbers (for example, 7000-8000).
- ICMP type and code: For ICMP, the ICMP type and code.
- Source or destination: The source (inbound rules) or destination (outbound rules) for the traffic. Specify one of these options.

When you specify a security group as the source or destination for a rule, the rule affects all instances associated with the security group. Incoming traffic is allowed based on the private IP addresses of the instances that are associated with the source security group.



Network and Security of Amazon EC2

Working with Security Groups:-

You can create, view, update, and delete security groups and security group rules using the Amazon EC2 console.

Contents:-

- Creating a Security Group.
- Describing Your Security Groups.
- Adding Rules to a Security Group.
- Updating Security Group Rules.
- Deleting Rules from a Security Group.
- Deleting a Security Group.



Network and Security of Amazon EC2

Controlling Access to Amazon EC2 Resources:-

- Your security credentials identify you to services in AWS and grant you unlimited use of your AWS resources, such as your Amazon EC2 resources.
- You can use features of Amazon EC2 and AWS Identity and Access Management (IAM) to allow other users, services, and applications to use your Amazon EC2 resources without sharing your security credentials.
- You can use IAM to control how other users use resources in your AWS account, and you can use security groups to control access to your Amazon EC2 instances. You can choose to allow full use or limited use of your Amazon EC2 resources.



Network and Security of Amazon EC2

Amazon EC2 and Amazon Virtual Private Cloud:-

- Amazon Virtual Private Cloud (Amazon VPC) enables you to define a virtual network in your own logically isolated area within the AWS cloud, known as a virtual private cloud (VPC).
- You can launch your AWS resources, such as instances, into your VPC. Your VPC closely resembles a traditional network that you might operate in your own data center, with the benefits of using AWS's scalable infrastructure.
- You can configure your VPC; you can select its IP address range, create subnets, and configure route tables, network gateways, and security settings. You can connect instances in your VPC to the internet.
- You can connect your VPC to your own corporate data center, making the AWS cloud an extension of your data center. To protect the resources in each subnet, you can use multiple layers of security, including security groups and network access control lists.



Network and Security of Amazon EC2

Amazon EC2 and Amazon Virtual Private Cloud:-

- Your account may support both the EC2-VPC and EC2-Classic platforms, on a region-by-region basis. If you created your account after 2013-12-04, it supports EC2-VPC only. If your account supports EC2-VPC only, AWS creates a default VPC for you.
- A default VPC is a VPC that is already configured and ready for you to use. You can launch instances into your default VPC immediately.
- If your account supports EC2-Classic and EC2-VPC, you can launch instances into either platform. Regardless of which platforms your account supports, you can create your own nondefault VPC, and configure it as you need.



Network and Security of Amazon EC2

Benefits of Using a VPC:-

- Assign static private IPv4 addresses to your instances that persist across starts and stops.
- Assign multiple IPv4 addresses to your instances.
- Define network interfaces, and attach one or more network interfaces to your instances.
- Change security group membership for your instances while they're running.
- Control the outbound traffic from your instances (egress filtering) in addition to controlling the inbound traffic to them (ingress filtering).
- Add an additional layer of access control to your instances in the form of network access control lists (ACL).
- Run your instances on single-tenant hardware.
- Assign IPv6 addresses to your instances.



Network and Security of Amazon EC2

Elastic IP Addresses:-

- An Elastic IP address is a static IPv4 address designed for dynamic cloud computing. An Elastic IP address is associated with your AWS account. With an Elastic IP address, you can mask the failure of an instance or software by rapidly remapping the address to another instance in your account.
- An Elastic IP address is a public IPv4 address, which is reachable from the Internet. If your instance does not have a public IPv4 address, you can associate an Elastic IP address with your instance to enable communication with the Internet; for example, to connect to your instance from your local computer.
- AWS currently do not support Elastic IP addresses for IPv6.



Network and Security of Amazon EC2

Elastic IP Address Basics:-

- To use an Elastic IP address, you first allocate one to your account, and then associate it with your instance or a network interface.
- When you associate an Elastic IP address with an instance or its primary network interface, the instance's public IPv4 address (if it had one) is released back into Amazon's pool of public IPv4 addresses. You cannot reuse a public IPv4 address.
- You can disassociate an Elastic IP address from a resource, and reassociate it with a different resource.
- A disassociated Elastic IP address remains allocated to your account until you explicitly release it.



Network and Security of Amazon EC2

Elastic IP Address Basics:-

- To ensure efficient use of Elastic IP addresses, we impose a small hourly charge if an Elastic IP address is not associated with a running instance, or if it is associated with a stopped instance or an unattached network interface. While your instance is running, you are not charged for one Elastic IP address associated with the instance, but you are charged for any additional Elastic IP addresses associated with the instance.
- An Elastic IP address is for use in a specific region only.
- When you associate an Elastic IP address with an instance that previously had a public IPv4 address, the public DNS hostname of the instance changes to match the Elastic IP address.
- AWS resolve a public DNS hostname to the public IPv4 address or the Elastic IP address of the instance outside the network of the instance, and to the private IPv4 address of the instance from within the network of the instance.



Network and Security of Amazon EC2

Working with Elastic IP Addresses:-

- Allocating an Elastic IP Address.
- Describing Your Elastic IP Addresses.
- Associating an Elastic IP Address with a Running Instance.
- Disassociating an Elastic IP Address and Reassociating it with a Different Instance.
- Moving an Elastic IP Address.
- Releasing an Elastic IP Address.
- Recovering an Elastic IP Address.

Network and Security of Amazon EC2

Elastic Network Interfaces:-

An elastic network interface (referred to as a network interface in this documentation) is a virtual network interface that you can attach to an instance in a VPC. Network interfaces are available only for instances running in a VPC.

A network interface can include the following attributes:

- A primary private IPv4 address.
- One or more secondary private IPv4 addresses.
- One Elastic IP address (IPv4) per private IPv4 address.
- One public IPv4 address.
- One or more IPv6 addresses.
- One or more security groups.
- A MAC address.
- A source/destination check flag.
- A description.



Network and Security of Amazon EC2

Elastic Network Interfaces:-

- You can create a network interface, attach it to an instance, detach it from an instance, and attach it to another instance. The attributes of a network interface follow it as it's attached or detached from an instance and reattached to another instance. When you move a network interface from one instance to another, network traffic is redirected to the new instance.
- Every instance in a VPC has a default network interface, called the primary network interface (eth0). You cannot detach a primary network interface from an instance. You can create and attach additional network interfaces. The maximum number of network interfaces that you can use varies by instance type.

Network and Security of Amazon EC2

Placement Groups:-

- A placement group is a logical grouping of instances within a single Availability Zone. Placement groups are recommended for applications that benefit from low network latency, high network throughput, or both. To provide the lowest latency, and the highest packet-per-second network performance for your placement group, choose an instance type that supports enhanced networking.
- First, you create a placement group and then you launch multiple instances into the placement group. AWS recommend that you launch the number of instances that you need in the placement group in a single launch request and that you use the same instance type for all instances in the placement group. If you try to add more instances to the placement group later, or if you try to launch more than one instance type in the placement group, you increase your chances of getting an insufficient capacity error.
- There is no charge for creating a placement group. If you stop an instance in a placement group and then start it again, it still runs in the placement group. However, the start fails if there isn't enough capacity for the instance.
- If you receive a capacity error when launching an instance in a placement group that already has running instances, stop and start all of the instances in the placement group, and try the launch again. Restarting the instances may migrate them to hardware that has capacity for all the requested instances.



Network and Security of Amazon EC2

Placement Group Limitations:-

- A placement group can't span multiple Availability Zones.
- The name you specify for a placement group must be unique within your AWS account.
- The maximum network throughput speed of traffic between two instances in a placement group is limited by the slower of the two instances. For applications with high-throughput requirements, choose an instance type with 10 Gbps or 20 Gbps network connectivity.
- Although launching multiple instance types into a placement group is possible, this reduces the likelihood that the required capacity will be available for your launch to succeed. AWS recommend using the same instance type for all instances in a placement group.
- You can't merge placement groups. Instead, you must terminate the instances in one placement group, and then relaunch those instances into the other placement group.



Network and Security of Amazon EC2

Placement Group Limitations:-

- A placement group can span peered VPCs; however, you will not get full-bisection bandwidth between instances in peered VPCs.
- You can't move an existing instance into a placement group. You can create an AMI from your existing instance, and then launch a new instance from the AMI into a placement group.
- Reserved Instances provide a capacity reservation for EC2 instances in an Availability Zone. The capacity reservation can be used by instances in a placement group that are assigned to the same Availability Zone. However, it is not possible to explicitly reserve capacity for a placement group.
- To ensure that network traffic remains within the placement group, members of the placement group must address each other via their private IPv4 addresses or IPv6 addresses (if applicable). If members address each other using their public IPv4 addresses, throughput drops to 5 Gbps or less.
- Network traffic to and from resources outside the placement group is limited to 5 Gbps.

How to Create Placement Group and Launch EC2 in Placement Group

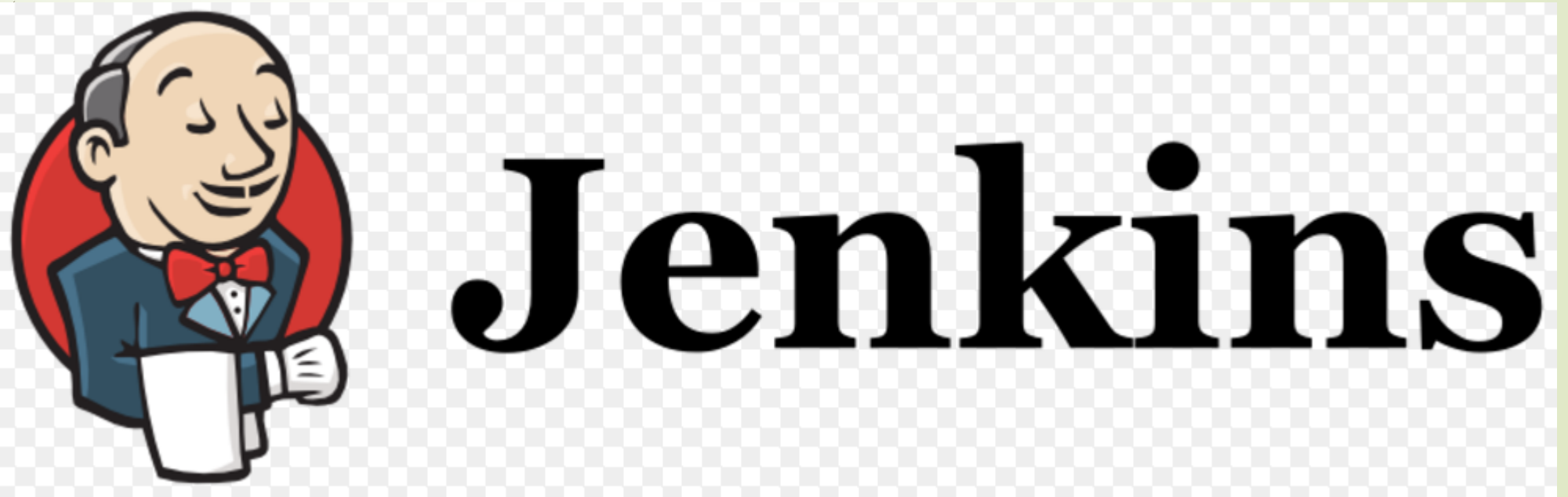




How to Launch SAS Software using AWS Market Place AMI

Amazon EC2(Elastic Compute Cloud)Demonstration

Demonstration of Hosting Jenkins on AWS EC2 using Management Console.





Amazon EC2(Elastic Compute Cloud)Demonstration

*Demonstration of Creating AWS EC2
using Management Console.*





Amazon EC2(Elastic Compute Cloud)Demonstration



*Demonstration of Creating AWS EC2
using AWS CLI.*

Amazon EC2 Demonstration

Demonstration of AWS EC2 by following ways:-

1. AWS Management Console.
2. AWS CLI (Command Line Interface).
3. AWS SDK (Software development Kits).





Amazon EC2 Pricing

Amazon EC2 Pricing:- Check the Video
for Pricings

