AWS EC2 (Amazon Elastic Compute Cloud)

What is EC2 in AWS Cloud :

Amazon EC2 is a web service which provides secure and resizable compute capacity in cloud. It is designed in such a way that anyone can provision a Virtual Machine within minutes and start working on it. It provides you complete control of your computing resources to let you run and work according to your requirement on Amazon's secure environment.

Some Important Features of EC2 :

- Compute capacity can be scaled Horizontally or Vertically within minutes.
- EC2 provides the availability with 99.99% of SLA for each region and each region consists of at least 3 availability zones.
- More than 250 types of EC2 instances are available which helps in building any types of Workloads for a better performance with a cost effective solution.
- EC2 instances are available with the choice of Memory, Processor, Storage, Networking Options, Operating System and Purchase Model.
- AWS supports 89 security standards and compliance certifications including PCI-DSS, HIPAA/HITECH, FedRAMP, GDPR, FIPS 140-2, and NIST 800-171, which is meaningfully more than any other cloud provider

Key Benefits of an Amazon EC2 :

- **Auto-Scaling :**
  - With the help of Autoscaling you can achieve a solution where the EC2 Instances can be Scale-In or Scale-Out easily as per the demand automatically. So this feature can help you to build a Crash Free solution along with a cost optimized solution.
- **Pay-As-You-Go :**
  - You will be charged only for the time what you have used. For an example if your requirement is to create an EC2 instance which is needed 3 hours in a day, so in that case you can keep the instance in stopped state for rest of the times so that you will be charged only for that 3 hours but not for the entire day or month.
- **Reliability :**
  - As AWS is spread their data centers across the world with more than 20 regions and more than 61 AZs, it is easier to expand the business whenever needed. Also it is easier to build a DR solution with minimum RTO and RPO and recover the site within minutes or even seconds.
- **Elasticity :**
  - This feature will allow you to run any kind of EC2 instances at the beginning when you are not sure about the required configuration and once you are done with the setup and testing you can resize the instances within minutes and continue your work without any complex changes or higher downtime.

Types of EC2 Instances and Use Cases :

There are many types of EC2 Instances are available in AWS with different types of configuration and the instance types determines the underlying Hardware of the instances which will be launched.

General Purpose (A1, T2, T3, T3a, M4, M5, M5a, M5n, M6g) :

- General purpose instances provide a balance of compute, memory and networking resources, and can be used for a variety of diverse workloads.

A Series Instances :

- A1 instances are used in applications that work in synchrony with the Arm ecosystem and are suitable for scale-out workloads. These instance types are suitable for the below workloads.
  - Web Servers
  - Containerized Applications
  - Caching Fleets
  - Distributed Data Stores

T2 and T3 Instances :

- These Instances offers Clock up or Down of CPU performance and these are well suited for the below work loads
  - Web Sites and Web Applications
  - Code Repository
  - Development, Build, Test and Staging Environments
  - Containerized Applications

M Series Instances :

- These instances offers a balance among Compute, Memory and Networking performance and these instances are best suited for below workloads.
  - Web and Application Servers
  - Small to Medium databases
  - Gaming Servers
  - Caching Fleets

Compute Optimized (C4, C5, C5n) :

- These instances are mostly useful for Compute/CPU dependent application which required high performance of processors and these well suited for below workloads.
  - Batch Processing Workloads
  - High-Performance Web Servers
  - High-Performance Computing
  - Scientific Modeling
  - Dedicated Gaming Servers and AD Serving Servers

Memory Optimized (R4, R5, R5a, R5n, X1, X1e, z1d, High Memory) :

- These instances are useful for the application which consume more Memory power for delivering fast performance for processing large Data sets in Memory and these instance are best suited for the below workloads.
    - High Performance Web Servers
    - Application Servers which process Big Data Sets like Hadoop, Spark Clusters etc.
    - MySQL, NoSQL (MongoDB, Cassandra) Databases

Accelerated / Graphics Computing (P2, P3, G3, G4, F1, inf1) :

- These Instances are the latest generation instances where they provide an accelerated performance with a higher CPU clock rates and these instances are suitable for the below requirement.
    - P2, P3, G3 and F1 instances are used for accelerated computing
    - G3 Instances are used for Graphic-Intensive applications

Storage Optimized (I3, I3en, H1, D2) :

- These instances are designed for the workloads which handle very large data Sets where it has to be written in memory and require high sequential read and right access.

    D2 Series :

    - These instances feature up to 48TB of HDD-based local storage and offer the lowest price-per-disk throughput performance.

    H1 Series :

    - These instances provide a balance between Compute and Memory features up to 16TB of HDD-based local storage which delivers high-disk throughput.

    I3 Series :

    - These instances provide Non-Volatile Memory Express (NVMe) SSD-backed instance storage optimized for low latency and very high I/O performance, high sequential read throughput, and high IOPS at low cost.

EC2 Instance Purchasing Options :

- Amazon EC2 purchasing options enable you to select the appropriate option based on the requirement to optimize your cost.

On-Demand Instances :

- Pay by the second for the instances that you launched and used.

Saving Plans :

o Reduce the EC2 cost by making a commitment to a consistent amount of usage, in USD per hour, for a term of 1 or 3 years.

Reserved Instances :

o Reduce the EC2 cost by making a commitment to a consistent instance configuration, including instance type and region for a term of 1 or 3 years.

Scheduled Instances :

o Purchase instances that are always available on the specified recurring schedule (daily, weekly, or monthly) for a term of 1 year.

Spot Requests :

o Request spare/unused EC2 instances which can reduce EC2 costs significantly with a highest discount rate. These instances are available with a bidding process where you can bid with a hourly price.
o Use cases are, stateless, fault-tolerant, or flexible applications such as big data, containerized workloads, CI/CD, web servers, high-performance computing (HPC), and test & development workloads
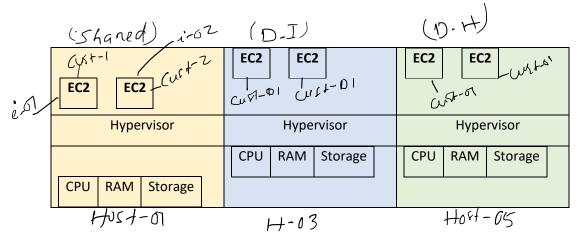
Shared Tenancy :

- Different instances from different customers can share the same physical box/server
- On reboot underlying H/W remain same
- With stop and start underlying H/W will get changed (from host-01 to host-02)

Dedicated Hosts :

o Pay for a physical host that is fully dedicated to run your instances on a dedicated physical server, and bring your existing per-socket, per-core, or per-VM software licenses to reduce costs. These are mostly beneficial for running Windows Server, Microsoft SQL Server, SUSE, and Linux Enterprise workloads.
o Even if you stop/start your instance it will remain on the same host, which ensure your all licenses are intact with the same instance
o Reserving whole host, so cost will be higher than other two, whether running one or more than one on the same host

Dedicated Instances :

o Pay, by the hour, for instances that run on single-tenant hardware. Instances are physically isolated from instances that belong to other AWS accounts/vpc.
o Until the instance is running it will remain on the same host-03, but when you stop and start it may go and launch on another host-04
o No other instance from any other customer or any other instance from your shared tenancy can come to host-03

Capacity Reservations :

- o Reserve capacity for your EC2 instances in a specific Availability Zone for any duration. This will guaranty you that the instance/capacity will be available at any cost at that period of time.

Other Important EC2 Services (Side Panel) :

Bundle Tasks :

- o Bundle Task is a process designed for creating an AMI and getting the Windows Instance ready for booting from Instance Store.

Lifecycle Manager :

- o It provides a simple and automated way to back up data stored on Amazon EBS volumes. You can define backup and retention schedules for EBS snapshots by creating lifecycle policies based on tags. With this feature, you no longer have to rely on custom scripts to create and manage your EBS backups.

Placement Groups :

- o Placement group is a logical grouping or clustering of instances within a single Availability Zone that's provide low Network Latency and High Network Throughput. Placement group places the instances on distinct hardware. A partition placement group places groups of instances in different partitions, where instances in one partition do not share the same hardware with instances in another partition

Side Panel Options :

Events : Any scheduled events or actions on EC2 instances or EBS volumes such as stop, start, reboot, retirement or I/O issues will be visible under this option.

Tags : This option will allow you to view all available Tags on your account assigned with the available resources. It also allows you to manage by adding/removing tags from here.

Limits : It shows all current limit on your account on EC2 services and you can also create a request to increase your limit if you have exceeded the default limit.

Instances: It contains whole list of provisioned instances on your account with any state.

Instance Types: It lists all available instance type on Amazon with their configuration details including price information.

Images (AMIs): This will allow you to create an Image of your existing instances which helps you to create the backup of whole instance as well as helps to recreate another instance as a mirror of the existing one.

Elastic Block Store (EBS): It provides block-level storage which can be attached and used by EC2 and RDS instances. This has a wide range of storage options based on performance and cost.

Volumes: This is an exact definition of EBS as stated above.
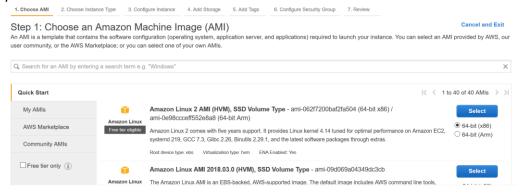
Other Important Terms :

- **Hibernation :** Hibernation stops your instance and saves the contents of the instance's RAM to the root volume. You cannot enable hibernation after launch
- **CPU Credit :** available for T2, T3, and T3a instances. Selecting Unlimited for the credit specification allows applications to burst beyond the baseline for as long as needed at any time. If the average CPU utilization of the instance is at or below the baseline, the hourly instance price automatically covers all usage. Otherwise, all usage above baseline is billed.
- **EBS Optimized :** Enables additional, dedicated throughput between Amazon EC2 and Amazon EBS. Supported on certain instance types only

How to Launch or Create an EC2 Instance :

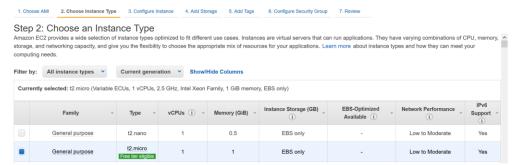There are total 8-Steps to be followed to create or launch an EC2 instance in AWS Cloud.

1. **Choose AMI :**
   - An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.
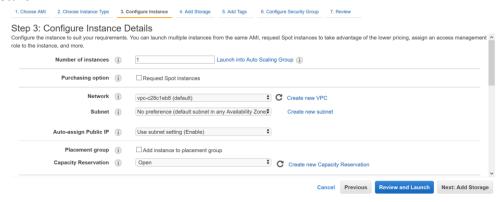

2. **Choose Instance Type :**
   - There are a wide range of Instance Type with various combination of CPU, Memory, Storage and Networking capacity where you can select one of them according to your workload requirement.

3. Configure Instance :

- In this phase you can select and enter your inputs based on your requirement like number of Instances, Networking Option, IAM role, Monitoring Option, Tenancy and many others details.
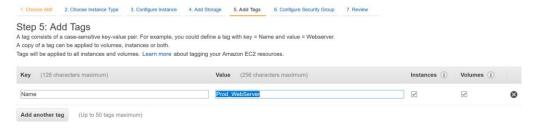

4. Add Storage :

- Here you can modify the default Root Volume and attach more additional EBS Volumes and Instance Store Volumes based on the requirement. Also you can attach the additional EBS volumes after launching the Instance but not the Instance Store Volumes.
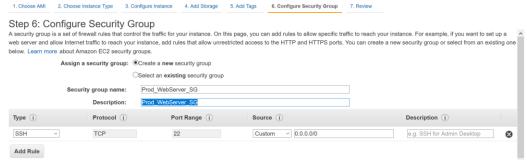

5. Add Tags :

- Tag consists of a case-sensitive key-value pair. For example key = Name and Value = App Server. You can add more Tags as per your requirement. A copy of Tags can be applied to Volumes and Instances or both.
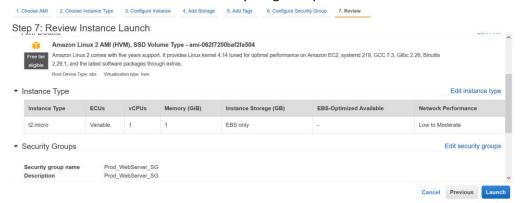
6. **Configure Security Group :**
   - A security group is a set of firewall rules that control the traffic for your instance. On this page, you can add rules to allow specific traffic to reach your instance. For example, if you want to set up a web server and allow Internet traffic to reach your instance. You can create a new security group or select from an existing one from the drop-down box.
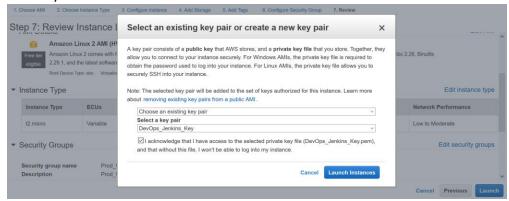

7. **Review :**
   - In this page you can review all of your configuration details before you actually launch the EC2 instance and make the correction if anything is required.

8. **Launch :**
   - Once you reviewed and confirmed your required settings you can click on Launch button and in this page you have to select the key-pair option either select one from the existing one or create a new Key-Pair with which you can login into the Instance after its launched successfully.

Connect to EC2 Instance :

1. Using Putty
2. From Linux/MacOS
3. Using Session Manager

User Data :

#!/bin/bash

sudo su

yum update -y

yum install -y httpd

systemctl start httpd

systemctl enable httpd

echo "I am serving from $(hostname -f)" > /var/www/html/index.html

Attaching EBS Volume to EC2 :

1. Create a volume (on the same region as EC2)
2. Select the volume and attach to the instance
3. Check the list of volumes : # lsblk
4. Check the file system of the volume : # file -s /dev/xvdf (empty volume output : /dev/xvdf: data)
5. Format the volume : # mkfs -t ext4 /dev/xvdf
6. Create a directory : # mkdir /volume-name
7. Mount the volume to the dir : # mount /dev/xvdf /volume-name/
8. Check the volume status : # df -h
9. Create a permanent mount point : make an entry on [/etc/fstab]
10. Take a backup of config file : # cp /etc/fstab /etc/fstab.bak
11. Syntax of the fstab entry : [device_name mount_point file_system_type fs_mntops fs_freq fs_passno]
12. Example : [/dev/xvdf   /newvolume   ext4   defaults,nofail   0   0]
13. # mount -a
14. Unmount the volume : # umount /dev/xvdf
15. Then detach the volume from EC2 instance

Types of Reserved Instances (RIs) :

Standard Reserved Instance :

➢ It provides highest discount up to 72% with less flexibility on the selected 1 or 3 year term, within that Region.
➢ Attributes can be Modified :-
  1. Instance Family : No
  2. Instance Type-Size : Yes (Linux only) (M)
  3. Availability Zone (optional) : Yes (M)
  4. Region : No
  5. Platform (OS) : No
  6. Tenancy : No
  7. Scope (Zonal to Regional and vice versa) : Yes (M)
  8. Exchange : No
  9. Capacity Reservation: (Only with Zonal RI)

Convertible Reserved Instance :

➢ It provides lower discount up to 66% with more flexibility on the selected 1 or 3 year term, within that Region.
➢ Attributes can be Exchanged/Modified :-
  1. Instance Family : (m4 to c5) :Yes (E)
  2. Instance Type-Size (Same family and OS) : Yes (M)
  3. Availability Zone (optional) : Yes (M)
  4. Region : No
  5. Platform (OS : Windows to Linux) : Yes (E)
  6. Tenancy : (Shared to Dedicated) :Yes (E)
  7. Scope (Zonal to Regional and vice versa) : Yes (M)
  8. Exchange : Yes (including Family, OS, Tenancy)
  9. Capacity Reservation: (Only with Zonal RI)

AWS Savings Plan

Compute Savings Plan :

1. Provide more flexibility and helps to lower the cost by up to 66% (just like Convertible RIs) compared to On Demand pricing and automatically apply to any Instance usage regardless of
   1. Instance Family (C5 to M5, or R5, etc..)
   2. Instance Size (c5.xlarge, m5.2xlarge, etc..)
   3. Availability Zones (Az-a, Az-b, Az-c, etc..)
   4. Region (Frankfurt, Ireland, London, etc..)
   5. Operating System (Linux, Windows, etc..)
   6. Tenancy (Shared, Dedicated, etc..)
   7. Fargate and Lambda are supported
   8. Applies to EC2, ECS, EKS, Fargate & Lambda

   Example : You can change from C5 to M5, c5xlarge to m5.2xlarge, Az-a to Az-b, Frankfurt to Ireland, Windows to Linux, Shared to Dedicated or move work load from EC2 to Fargate or Lambda.

EC2 Instance Savings Plan :

1. Provide slightly less flexibility and helps to lower the cost by up to 72% (just like Standard RIs) compared to On Demand pricing and automatically apply to the Selected Instance family within that region usage regardless of
   1. Instance Family (C5 to M5 or R5, etc..)
   2. Instance Size (c5.xlarge, m5.2xlarge, etc..)
   3. Availability Zones (Az-a, Az-b, Az-c, etc..)
   4. Region (Frankfurt, Ireland, London, etc..)
   5. Operating System (Linux, Windows, etc..)
   6. Tenancy (Shared, Dedicated, etc..)
   7. Fargate and Lambda are not supported
   8. Applies to EC2, ECS and EKS

   Example : You can change from c5.xlarge to c5.2xlarge, Az-a to Az-b, Windows to Linux and Shared to Dedicated.