What is Amazon EC2 Instance

An instance is nothing but a virtual server in the cloud.

Amazon EC2 Instance Types

Amazon EC2 provides a wide selection of instance types optimized to fit different use cases. Instance types comprise varying combinations of CPU, MEMORY, STORAGE, and NETWORKING capacity and give you the flexibility to choose the appropriate mix of resources for your applications.

Each instance type includes one or more instance sizes like SMALL, MEDIUM, LARGE, XLARGE,2XLARGE, 4XLARGE, NXLARGE, METAL, NANO, MICRO etc.

EC2 instances are mainly classified into 5 categories.

1. General Purpose

2. Compute Optimized

3. Memory Optimized

4. Storage Optimized

5. Accelerated Computing

General Purpose Instances:

General purpose instances provide a balance of COMPUTE, MEMORY and NETWORKING resources, and can be used for a variety of diverse workloads.

These instances are ideal for applications that uses resources in equal proportions such as web servers and code repositories.

Instances starts with **A, T, M** comes under this category type.

Compute Optimized Instances:

Compute Optimized Instances provide more COMPUTE POWER.

These instances are ideal for applications that benefit from high performance processors. Instances belonging to this family are well suited for batch processing workloads, media transcoding, high performance web servers, high performance computing (HPC), scientific modeling, dedicated gaming servers and ad server engines, machine learning inference and other compute intensive applications.

Instances starts with **C** comes under this category type.

Memory Optimized Instances:

Memory optimized instances are designed to deliver fast performance for workloads that process large data sets in memory. This means these instances will have more RAM (RANDOM ACCESS MEMORY).

Instances starts with **R, X, Z, High Memory** comes under this category type.

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.

Instances starts with **I, D, H** comes under this category type.

Accelerated Computing Instances:

Accelerated computing instances use hardware accelerators, or co-processors, to perform functions, such as floating-point number calculations, graphics processing, or data pattern matching, more efficiently than is possible in software running on CPUs.

These instances will be equipped with high end GPU's.

Instances starts with **P, G, F, Inf** comes under this category type.

Amazon EC2 Instance Features (Additional)

Amazon EC2 instances provide a number of additional features to help you deploy, manage, and scale your applications.

We can think of these additional features like different types of toppings for pizza.

1. Burstable Performance Instances.

2. Multiple Storage Options.

3. EBS-optimized Instances.

4. Cluster Networking.

5. Intel Processor Instances.

Burstable Performance Instances:

Amazon EC2 allows you to choose between Fixed Performance Instances (e.g. M5, C5, and R5) and Burstable Performance Instances (e.g. T3).

Many applications such as web servers, developer environments and small databases don't need consistently high levels of CPU, but benefit significantly from having full access to very fast CPUs when they need them.

Burstable Performance Instances provide a baseline level of CPU performance with the ability to burst above the baseline.

Baseline performance and ability to burst are governed by CPU Credits.

A CPU Credit provides the performance of a full CPU core for one minute.

If the instance needs to run at higher CPU utilization for a prolonged period, it can do so at a flat additional charge of 5 cents per vCPU-hour.

Multiple Storage Options:

Amazon EC2 allows you to choose between multiple storage options based on your requirements.

Amazon EBS is a durable, block-level storage volume that you can attach to a single, running Amazon EC2 instance. You can use Amazon EBS as a primary storage device for data that requires frequent and granular updates.

Amazon EBS provides three volume types to best meet the needs of your workloads.

---> **General Purpose (SSD)**   # This is the default EBS

 General Purpose (SSD) volumes are suitable for a broad range of workloads, including small to medium sized databases, development and test environments, and boot volumes.

---> **Provisioned IOPS (SSD)**

Provisioned IOPS (SSD) volumes offer storage with consistent and low-latency performance, and are designed for I/O intensive applications such as large relational or NoSQL databases.

---> **Magnetic**

Magnetic volumes provide the lowest cost per gigabyte of all EBS volume types. Magnetic volumes are ideal for workloads where data is accessed infrequently, and applications where the lowest storage cost is important.

Many Amazon EC2 instances can also include storage from disks that are physically attached to the host computer. This disk storage is referred to as instance store.

Instance store provides temporary block-level storage for Amazon EC2 instances. The data persists only during the life of the associated Amazon EC2 instance.

EBS-optimized Instances:

For an additional, low, hourly fee, customers can launch selected Amazon EC2 instances types as EBS-optimized instances.

EBS-optimized instances enable EC2 instances to fully use the IOPS provisioned on an EBS volume. EBS-optimized instances deliver dedicated throughput between Amazon EC2 and Amazon EBS.

AWS recommend using Provisioned IOPS volumes with EBS-optimized instances or instances that support cluster networking for applications with high storage I/O requirements.

Cluster Networking:

Select EC2 instances support cluster networking when launched into a common cluster placement group. A cluster placement group provides low-latency networking between all instances in the cluster.

The bandwidth an EC2 instance can utilize depends on the instance type and its networking performance specification. Inter instance traffic within the same region can utilize up to 5 Gbps for single-flow and up to 100 Gbps for multi-flow traffic in each direction (full duplex).

Cluster networking is ideal for high performance analytics systems and many science and engineering applications, especially those using the MPI library standard for parallel programming.

Intel Processor Instances:

Amazon EC2 instances that feature an Intel processor may provide access to the following processor features

---> **Intel AES New Instructions (AES-NI):**

Intel AES-NI encryption instruction set improves upon the original Advanced Encryption Standard (AES) algorithm to provide faster data protection and greater security.

All current generation EC2 instances support this processor feature.

---> **Intel Advanced Vector Extensions (Intel AVX, Intel AVX2 and Intel AVX-512)**

Intel AVX instructions improve performance for applications like image and audio/video processing, scientific simulations, financial analytics, and 3D modeling and analysis.

These features are only available on instances launched with HVM AMIs.

---> **Intel Turbo Boost Technology:**

The processor is able to automatically run cores faster than the base operating frequency to help you get more done faster.

---> **Intel Deep Learning Boost (Intel DL Boost):**