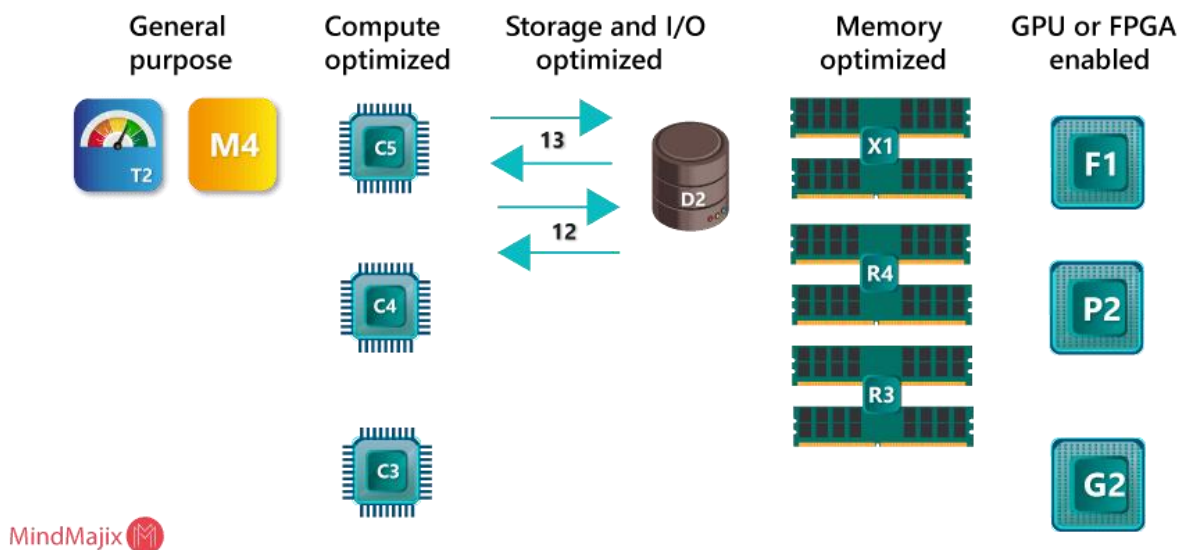# AWS EC2 Instance Types

Wondering which among the various AWS EC2 Instance Types is right for your application? In simple terms, the instance type that you choose determines the hardware specification of the cloud-hosted computer that you will utilize for running your EC2 instance. Here is a complete introduction to the top 5 Amazon EC2 instances that you can choose from.

In general, Amazon EC2 instance types provide you with flexibility and scalability with different computing, memory, and storage specifications within each instance family. AWS EC2 instances are broadly categorized on the basis of budget and workload requirements.

AWS EC2 Instance Types

The following are the list of AWS EC2 Instances Types:

- **General Purpose Instances**

- **Compute Optimized Instances**

- **Memory-Optimized Instances**

- **Accelerated Computing Instances**

- **Storage Optimized Instances**

Now, let us take a look at what each of these five EC2 instance types offers in more detail:

Type 1 - General Purpose Instances

Among the most popular and widely used EC2 instance types, the General Purpose instance is a good choice if you are new to cloud computing or AWS in general. As this instance type offers a wide balance of computing power, memory, and storage, it is suited for a majority of AWS workloads.

General-purpose instances are mostly utilized in services related to web servers, mobile or gaming development environments or apps, or enterprise-level applications like ERP or CRM. Another major distinction among general-purpose instances is the use of Fixed EC2 instances and Burstable instances – where you can scale up your overall computing power at an extra cost.

EC2 General-purpose instance Types

Here are the details of the various general-purpose instances that you can choose from:

i. A1 instance

A1 instances are the ARM-based offering of EC2 instance types – as compared to others that use either Intel or AMD processing. This type of instance is more suited for web servers and containerized microservices – along with applications that are run on open source tools like Java or Python.

ii. M5 instance

This is the latest generation of general-purpose instances that is powered by the Intel Xeon Platinum 8175 3.1Ghz processor. With their cloud computing power, M5 instances provide a balance of computing, memory, and networking power. This instance type is suited for small-to-midsize databases, data processing tasks, and as a backend server for enterprise applications like SAP or SharePoint.

iii. T3/ T3a instance

Available with burstable instances, T3 and T3a are the respective general-purpose instance types powered with Intel and AMD processors. These instance types are a good fit if you are looking for a cheaper and less powerful option than the M5 fixed instance. They are commonly used for long-lasting application instances such as websites, web applications, and code repositories.

Type 2 - Compute Optimized Instances

As the name suggests, compute-optimized instances are used during compute-intensive workloads that can benefit from processors with high computing power. Compute-optimized

[instances](#) deliver high performance at a cost-effective price and are typically used in applications like web servers and scientific modeling.

EC2 Compute Optimized instances Types

Here are the details of the two types of Compute-optimized instances that you can choose from:

i. C5/ C5n instance

The C5 series of instances are suitable for applications like online gaming, scientific modeling, media transcoding – which require raw computing power. C5 instances are executed on the Intel Xeon Platinum processor and have recorded a 25% improvement in speed as compared to the previous C4 instance generation. With the C5d instance type, you can physically connect the NVME-based SSD device to the host server to provide block-level storage for the entire instance lifetime.

ii. C6/ C6g instance

This family of compute-optimized instances is powered by the AWS Graviton2 series of processors and is suited for highly intensive and advanced applications such as high-performance computing, video encoding, ad serving, and distributed analytics. With the C6g instance type, you can get a 40% improvement in price-performance as compared to the C5 instance family.

Type 3 - Memory-Optimized Instances

As the name suggests, memory-optimized instances are used for memory-intensive workloads that are required to process large datasets at a fast speed. Examples of memory-intensive applications include Big Data analytics or those running on Hadoop or Apache Spark.

EC2 Memory Optimized instances Types

Here are the details of the different types of memory-optimized instances that you can choose from:

i. R5/ R5a/ R5n instance

The R5 series of instances are suited for workloads with high memory consumption such as high-performance databases, real-time Big data analytics, and large in-memory cache applications. Among these, R5 and R5a are respectively powered by Intel and AMD processors – and also benefit from the [AWS Nitro system](#) that provides easy access to the computing and memory resources of the server.

As compared to the previous R4 type, R5 instances offer 5% more memory for each vCPU along with a 10% price improvement per GiB over R4.

ii. R6g/ R6gd instance

Powered by the AWS Graviton2 processor, R6 instances are suited for high memory workloads such as open-source databases (example, MySQL) and in-memory caching (example, KeyDB). These instances use a custom-built AWS Graviton2 processor equipped with a 64-bit ARM Neoverse core. In the R6gd instance, the local NVME-based SSD drive is physically connected to the host server, thus enabling block-level storage.

iii. X1/ X1e instance

Powered by the Intel Xeon processor, the X1 family of memory-optimized instances are designed to provide high computational memory for memory-intensive applications like SAP HANA, Apache Spark, and for high-performance computing. Among all EC2 instances, the X1e instance type provides the highest memory-to-compute ratio at the lowest price calculated for each GiB of RAM.

iv. High memory instance

As the name suggests, high memory instances provide the highest capacity of RAM – ranging from 6TB to 24TB in a single instance. High memory instances are used to run high in-memory databases – including the deployment of the SAP HANA database to the cloud platform.

High memory instances are only available on dedicated hosts – where you need to commit to running instances for a 3-year period.

Type 4 - Accelerated Computing Instances

Accelerated Computing instances use additional hardware accelerators like Graphics Processing Units (or GPUs) and Field Programmable Gate Arrays (or FPGAs) that enable higher throughput in compute-intensive applications with more parallelism. For example, with GPU-powered instances, applications can access NVIDIA GPUs that have thousands of computing cores.

Similarly, FPGA-powered instances provide applications with access to large FPGAs with millions of parallel logic cells.

This instance type is suitable for applications that require parallel processing. This includes graphic processing, floating-point calculators, and data pattern matching.

EC2 Accelerated Computing instances Types

Here are the details of the different types of accelerated computing instances that you can choose from:

i. P3 instance

The latest generation of GPU-based instances uses up to 8 NVIDIA Tesla GPUs and is powered by a high-frequency Intel Xeon processor. This instance type supports NVLink for peer-based GPU communication and provides up to 100Gbps of network bandwidth.

ii. P2 instance

Designed for general-purpose GPU compute applications, P2 instances are powered by the Intel Xeon processor and feature high-performance NVIDIA K80 GPUs. P2 instances provide capabilities like high-performance networking and double-precision floating-point calculations – thus making it ideal for deep learning applications and high-performance databases.

iii. Inf1 instance

Featuring 16 AWS Inferential machine learning chips that enable low latency and cost-effective performances, Inf1 instances are powered by Intel Xeon 2nd generation processors. API developers can also work with AWS Deep Learning APIs that are bundled with AWS Inferential.

iv. G3 instance

Designed with the NVIDIA Tesla M60 GPU, G3 instances provide high-performance and cost-effective solutions for graphic-intensive applications using DirectX or OpenGL. Some of the graphics-related tasks include graphic rendering and streaming. G3 instances are powered by high-frequency Intel Xeon processors and can enable the NVIDIA Grid Virtual Workstation functionality.

v. G4 instance

Powered by a 2nd generation Intel Xeon Scalable processor, G4 instances are suited for accelerating machine learning inference and graphics-intensive workloads. Designed with the NVIDIA T4 Tensor Core GPU, this instance type also supports the NVIDIA Grid Virtual Workstation feature.

vi. F1 instance

F1 instances offer hardware acceleration using FPGAs. Powered with the high-frequency Intel Xeon processor, F1 instances feature NVMe SSD storage and support for enhanced networking. You can also use customized FPGA AMIs for quicker development and deployment of applications.

Type 5 - Storage Optimized Instances

As the name suggests, storage optimized instances are used for applications that have high storage requirements particularly with sequential read and write applications like log processing. Storage optimized instances are designed to deliver a high number of low latency and random I/O operations each second (or IOPS).

Storage optimized instances are also suitable for cloud-running applications that run high transaction and low latency workloads in use cases such as in-memory databases, data warehousing, and data analytics.

EC2 Storage Optimized instances Types

Here are the details of the different types of storage optimized instances that you can choose from:

i. D2 instance

Featuring 48TB of HDD storage, D2 instances is a storage-optimized instance with high-density storage with high sequential read/write for large datasets – including large Hadoop distributed environments. Along with high disk throughput, D2 instances are available for the lowest cost per disk throughput on Amazon EC2. These instances are designed for applications like MPP data warehousing, MapReduce and Hadoop computing, and log processing.

ii. H1 instance

As compared to D2 instances, H1 instances offer lesser density storage with a maximum of 16TB HDD. Powered by an Intel Xeon processor, H1 instances offer high disk throughput and enhanced networking of up to 25Gbps. These instances are most suitable for data-intensive applications like MapReduce, and for applications that require high throughput and sequential access to large data volumes.

iii. I3/ I3en instance

The I3 instance family offers SSD storage that has lower latency as compared to HDD-based instances. The I3 instance features an NVMe SSD storage that provides high IOPS at low costs thanks to its offering of low latency and high random I/O performance. This instance family is most suitable for high-frequency OLTP systems, relational databases, and caching for in-memory databases like Redis.

Conclusion

Depending on your workload and application requirements, you can choose from the different Amazon EC2 instance types that we have explained in this guide. To cater to different workloads, the Amazon EC2 platform keeps expanding its offered instance types with new releases.