

# CAPITALIZING ON EC2 ADVANCEMENTS

## M6a general purpose instance types

### AT A GLANCE

Resilient enterprise architecture is built from platform advancements, including the pillars of compute, storage, database, networking, and security. Over the years, each of these pillars has been greatly modernized, enabling businesses to advance their go-to-market strategies better than ever before. In healthcare, these advancements result in the creation of architectures that elevate patient care and bring advanced analytics capable of analyzing datasets of large sizes from a variety of sources, including electronic health records (EHRs), wearables, etc., at a better cost and with increased security. Cloud offers the best performance and enables advancements across platforms, allowing industries such as healthcare to capitalize on the latest and greatest technology stack advancements.

Compute is a critical pillar, where rightsizing is vital to bringing architecture advancements and cost efficiencies. Cloud brings significant advantages, as customers can pick up modernized computing configurations throughout their engagement.

Amazon Web Services (AWS) has one of the largest in both volume and variety of compute instances, commonly referred to as Amazon Elastic Compute Cloud (EC2). Each is optimized for various use cases and purposes. The instance types span the categories of general purpose, compute optimized, memory optimized, accelerated computing, storage optimized, and high performance compute optimized. With each instance type, there are various options based on the configuration requirements.

When building a cloud architecture that includes EHR or application hosting, it's important to choose the right computing type based on what type of workload will run on that system. This will result in optimal performance, productivity, and cost savings. For example, the compute optimized instances run on a high-performance processor, and these instance types are best suited for batch processing workloads, web servers, scientific modeling, ML inferences, etc. *(Continued on next page.)*

### WHAT THIS MEANS FOR YOU

One of the greatest advantages of cloud computing is the constant platform advancements across every component of cloud architecture. With EC2 instances, AWS brings consistent and ongoing enhancements to all types of instances, offering the benefit of the latest technological advancements to its customers.

In addition to new instance releases, AWS works with respective software providers to get their instances approved, which includes those that run distinct workloads such as EHRs. AWS-dedicated solution architects work alongside them throughout the testing and release of all new instance types. When considering new cloud architectures, including when hosting an EHR on AWS, it's recommended to do a total cost of ownership (TCO) assessment and cloud sizing at that point in time, irrespective of if an AWS TCO has been done previously. This process ensures consideration of the latest and greatest AWS advanced EC2 instance types.

## AT A GLANCE (CONT'D.)

These instances are constantly updated at every layer, including silicon and processor, down to the hardware and software components that power the system. Each upgrade brings significant productivity uplift, resulting in faster cycle time for applications, and reducing the total runtime. Cloud provides substantial benefits when it comes to time to market at every significant update, along with options to run application workloads on a much more updated instance type without having to pay for the backend upgrades. As part of solution and design, and to make the most out of a cloud investment, it's recommended to consider the different financial options, such as reserved pricing with either compute or EC2 savings plans.

AWS recently announced achieving a higher benchmark towards EHR hosting, made possible with the general purpose instance called M6idn. In this article, we will dive deeper into the recent announcement of the general purpose instance type, specifically the M6a instance type, and the advantages of using cloud for running complex and special workloads such as the EHR.

## GENERAL PURPOSE INSTANCE TYPES

The general purpose instance type provides a balance of compute, memory, and networking resource. The systems are powered by various processors, including Intel Xeon, 3rd generation AMD Arm-based, and multiple Graviton generation processors, making this EC2 instance widely fitting to various business and enterprise requirements. This general purpose instance type would be most commonly used when it comes to EHR hosting. With the hypervisor that protects the hardware, enables virtualization, and provides a rich set of virtual management capabilities, these instances are powered by AWS Nitro. This is AWS's proprietary hypervisor that brings the best possible innovation and cost savings for customers.

When it comes to the processors that are powering these instances, AWS follows the best engineering across the families of processors comprising Intel, AMD, and the AWS-owned Graviton processors. They not only bring the same performance as the other family processors, but with the AWS proprietary advantage, customers get the compute performance with significant cost savings. As of today, there are 16 types of EC2 general purpose instances available.

**FIG. 1** General purpose instance types

M7g	Mac	M6g	M6i	M6in	M6a	M5	M5n	M5zn	M5a
M4	A1	T4g	T3	T3a	T2				

When the new instance types are released for production, customers get the choice to upgrade to the advanced systems with just the flip of a switch. When compared with an on-premises infrastructure, a large advantage of the cloud is the point-in-time upgrade, allowing customers to get the best of systems, bringing performance, security, and significant cost savings.

## INSTANCE RELEASES

At the re:Invent 2022 AWS conference, AWS released several new instances (in preview and production):

- M6in and M6idn: For network-intensive workloads, such as application development environments, real-time big data analytics, and gaming servers.
- C6in: For network-intensive workloads, such as distributed computing applications, data and analytics, and HPC.
- C7gn: Currently in preview. These are best suited for network-intensive workloads, such as load balancers, tightly coupled cluster computing jobs, and CPU-based AI/ML inference.
- R6in and R6idn: For workloads that require high networking throughput, such as processing large data sets
- R7iz: Released in preview mode, now in production state. A good option for simulation workloads, as well as workloads for relational databases and other commercial software licensed on a per-core basis.
- Hpc7g: Currently in preview and powered by AWS Graviton3E processors. For tightly coupled compute-intensive HPC and distributed computing workloads.
- Hpc6id: For tightly coupled HPC workloads, such as finite element analysis for various types of simulations.
- Inf2: Currently in preview. For deep learning (DL) inference applications, such as language translation, speech recognition, personalization, and fraud detection.

## TECHNICAL DETAILS

The M6a instance is powered by the 3rd generation AMD EPYC processor and is built on the AWS Nitro hypervisor that brings the best of virtualization productivity and security. M6a has extensive features, including:

- Up to 3.6 GHz 3rd generation AMD EPYC processors (AMD EPYC 7R13)
- Up to 35% better compute price performance over M5a instances
- Up to 50 Gbps of networking speed
- Up to 40 Gbps of bandwidth to the Amazon EBS
- Instance size with up to 192 vCPUs and 768 GiB of memory
- Supports Elastic Fabric Adapter on the 48xlarge size
- Built on the AWS Nitro System, a combination of dedicated hardware and lightweight hypervisor

M6a instance delivers up to 35% better price performance compared to the previous generation M5a instance and 10% lower cost than comparable x86-based EC2 instances. At this time, M6a comes in 11 different options (see Figure 2) with varied virtual CPUs, memory, storage, and networking configurations, making it the best fit for most general purpose workloads, backend servers supporting enterprise applications, and application development environments.

**FIG. 2** M6a configuration options

Instance Size	vCPU	Memory (GiB)	Instance Storage (GB)	Network Bandwidth (Gbps)**	EBS Bandwidth (Gbps)
m6a.large	2	8	EBS-Only	Up to 12.5	Up to 10
m6a.xlarge	4	16	EBS-Only	Up to 12.5	Up to 10
m6a.2xlarge	8	32	EBS-Only	Up to 12.5	Up to 10
m6a.4xlarge	16	64	EBS-Only	Up to 12.5	Up to 10
m6a.8xlarge	32	128	EBS-Only	12.5	10
m6a.12xlarge	48	192	EBS-Only	18.75	15
m6a.16xlarge	64	256	EBS-Only	25	20
m6a.24xlarge	96	384	EBS-Only	37.5	30
m6a.32xlarge	128	512	EBS-Only	50	40
m6a.48xlarge	192	768	EBS-Only	50	40
m6a.metal	192	768	EBS-Only	50	40

In AWS's latest announcement on EHR hosting on the its platform, the M6a instance has resulted in higher benchmarks and scalability of the EHR database performance as compared to the previous M5a instances. If considering hosting an EHR on AWS, it's recommended to do a TCO assessment, as this will allow for the proper rightsizing and cost for the EHR environments. With the latest AWS announcement, having legacy systems will cost more, as additional virtual servers will be needed to meet the application and database performance requirements. It's recommended to use the latest generation's systems to take full advantage of performance and cost savings.

## NORDIC IN ACTION

Advancements in EC2 instances are constant. However, there are times when there may be application or other level dependencies that require the need to continue to run on previous-generation instances. AWS does allow customers to retain previous generation instances, and the timelines to upgrade to the most recent versions will change depending on the age of the previous generation.

The EC2 instance upgrade process is very straightforward, with most cases as easy as a flip of a switch. However, the steps are different depending on existing instances' root volume and compatibility to the new instance type. Cloud also offers resiliency and migration techniques for legacy systems and those that are not compliant. Here are the technical steps to follow if an instance is backed by an elastic block store (EBS) root volume that is compatible. Detailed technical documents are available for both compatible and non-compatible instances.

1. Open the Amazon EC2 console at <https://console.aws.amazon.com/ec2/>.
2. In the navigation pane, choose **Instances**.
3. Select the instance and choose **Instance state, Stop instance**. When prompted for confirmation, choose **Stop**. It can take a few minutes for the instance to stop.
4. With the instance still selected, choose **Actions, Instance settings, Change instance type**. This option is grayed out if the instance state is not stopped.
5. On the **Change instance type** page, do the following:
  - a. For **Instance, type**, select the instance type that you want. If the instance type is not in the list, then it's not compatible with the configuration of your instance.
  - b. Choose **Apply** to accept the new settings.
6. To start the instance, select the instance and choose **Instance state, Start instance**. It can take a few minutes for the instance to enter the running state.

In this cloud environment, changing or upgrading instance types is real-time, and seemingly instantaneous. In an on-premises environment, upgrading a platform level is a tedious and time-consuming procurement process that has various dependencies, including supply chain, rack and stack capabilities, etc. The ability to quickly upgrade instances to modernized and use case-customized instances allows customers to gain greater productivity and increase resiliency and security without spending CapEx/OEx budgets. It is one of many significant reasons to consider the cloud.



**SARAVANAN SONAISAMY**  
Cloud Solutions Lead, Digital Health



**KEVIN ERDAL**  
Practice Leader, Digital Health

---

CloudPress@NordicGlobal.com

## REFERENCES

1. Amazon Web Services. (2023). *Amazon EC2*. <https://aws.amazon.com/ec2/instance-types/>
2. Amazon Web Services. (2023). *General purpose instances*. <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/general-purpose-instances.html>
3. DeLurgio, J., Pradel, A., & Dinger, M. Amazon Web Services. (2023, April 27). *AWS increases scalability of Epic database performance*. <https://aws.amazon.com/blogs/industries/aws-increases-scalability-of-epic-database-performance/>
4. Amazon Web Services. (2023). *Amazon EC2 M6a Instances*. <https://aws.amazon.com/ec2/instance-types/m6a/>
5. Amazon Web Services. (2023). *AWS Nitro System*. <https://aws.amazon.com/ec2/nitro/>
6. Amazon Web Services. (2023). *Change the instance type*. <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ec2-instance-resize.html>
7. Amazon Web Services. (2023). *Previous Generation Instances*. <https://aws.amazon.com/ec2/previous-generation/>