[ENGLISH]

Hello, Cloud Gurus,

and welcome to this lesson,

where we're going to take a look

at the different EC2 instance types that are available.

And there are loads of them, but don't worry,

for any of the associate-level exams,

you don't need to memorize them all.

It's only for the Solutions Architect Professional

where you will need to remember the names

of the different instance types.

So that being said,

what do we mean when we talk about instance types?

Well this is all about the underlying hardware

that is used to host your EC2 instance.

So when you launch an EC2 instance,

the instance type determines the hardware

of the underlying host computer used for your instance,

or where your instance is running.

So ultimately, our instances, or virtual machines,

are running on a physical server,

and it's this physical server that I'm talking about

and the hardware characteristics

of the underlying physical server.

Now, instance type also determines other characteristics

or capabilities when it comes to CPU, memory, and storage.

And each instance type offers different compute,

memory, and storage capabilities,

and depending on these capabilities,

they are grouped together in instance families.

So how do you know which one to choose?

Well the simple answer is to select an instance type

based on your application's requirements.

So you select an instance type based on the requirements

of the application that you plan to run on your instance.

And there is a massive choice, and it's growing steadily

as they bring out more and more instance types.

And the reason why they have so many is because

instance types are optimized to fit different use cases,

and they give you the flexibility to choose

the appropriate mix of resources for your applications.

But the best way to understand the different

instance types that are available

is to look at the latest AWS documentation,

and I've included a link to this documentation

in the resources for this lesson.

Now the first thing that you might notice

when you start creating EC2 instances

is that there are loads of different instance types,

but they are all organized into instance families.

So first of all, we've got general purpose,

which provides a balance of compute, memory, and network,

and it's a great choice for many applications,

including small or medium databases,

web servers, or code repositories.

And if you select the different tabs down here--

so we've got the Mac instance types, T2, 3, and 4,

the M instance types, and A as well--

if you select one of these tabs,

you can take a look at the specification

for each instance type in this family.

And if you scroll down here, it will give you some use cases

for this instance type as well.

So that's general purpose.

And then we've got compute optimized,

and these are designed to have a higher CPU power,

so they're designed for compute-intensive applications.

And these are going to be great for applications

which have high CPU requirements,

for instance, batch processing,

distributed analytics, media transcoding,

or high performance science or engineering applications.

And these instance types, they mainly begin with C,

but you've also got HPC,

or high-performance computing as well.

Moving on to memory optimized.

And these instance types are designed

to deliver fast performance for workloads

that process large data sets in memory,

so think any memory-intensive workloads,

or in-memory databases, for example.

Moving on to accelerated computing.

And this is a family of instances

that include special hardware accelerators,

or special processes that are optimized to perform workloads

like graphics processing, data pattern matching,

or floating point number calculations.

And these are great for machine learning workloads as well,

and there's even some instance types

in this family that include CPUs that were designed

from the ground up to support deep learning workloads

or machine learning workloads as well.

And then over here, we've also got instances,

which include field programmable gate arrays.

And this is a special processor or card

that can be programmed to run specific processing

for your application.

And it's really good for applications which need

massively parallel processing power,

so things like genomics, data analytics,

or real-time video processing, or financial computing.

And actually, a long time ago

when I used to work in investment banking,

we used field programmable gate arrays for accelerating

some of the algorithmic trading that we did.

So that's real low latency, high throughput trading,

and as you can imagine,

that is a highly competitive business.

So any competitive advantage that you can gain

by processing your algorithm and running your analytics

a few milliseconds faster than your competitor

can give you a massive advantage in the market.

And then over here, we've also got some instance types,

which are designed specifically

for real-time video transcoding workloads.

And then finally, we've got the

storage-optimized instance types,

and these basically allow you to add large amounts

of storage to your EC2 instance.

So you can select an instance based on the amount

and type of storage that you need,

how many disks you want, and how much storage in total

is needed for your application.

So this is when you need direct attached storage options

for applications with very specific disk I/O,

and storage capacity requirements.

So these are great for NoSQL databases,

data warehousing, Elasticsearch, big data, analytics,

and that kind of thing.

So that is a really quick run through

of all the different instance families.

But don't worry, as I said before,

you don't need to remember them all--

just have a general awareness that there is an instance type

for pretty much any kind of workload you can imagine.

And for my exam tips, just remember the instance type

determines the hardware configuration

and capabilities of the host computer

where your instance or virtual machine is running.

Each instance type offers different compute,

memory, and storage capabilities,

and they're grouped together in instance families.

And you should select the instance type

which is most appropriate based on the requirements

of your application.

So that's the end of this lesson.

If you have any questions, please let me know.

Otherwise, feel free to join me in the next lesson,

where we're going to be getting our hands dirty,

launching an EC2 instance.

Thank you.

[SPANISH]

Hola, gurús de la nube,

y bienvenidos a esta lección,

donde vamos a echar un vistazo

en los diferentes tipos de instancias EC2 que están disponibles.

Y hay un montón de ellos, pero no te preocupes,

para cualquiera de los exámenes de nivel asociado,

no es necesario memorizarlos todos.

Es solo para el Solutions Architect Professional

donde tendrás que recordar los nombres

de los diferentes tipos de instancias.

Así que dicho esto,

¿A qué nos referimos cuando hablamos de tipos de instancias?

Bueno, esto es todo sobre el hardware subyacente.

que se utiliza para alojar su instancia EC2.

Entonces, cuando lanza una instancia EC2,

el tipo de instancia determina el hardware

de la computadora host subyacente utilizada para su instancia,

o dónde se ejecuta su instancia.

Entonces, en última instancia, nuestras instancias, o máquinas virtuales,

se ejecutan en un servidor físico,

y es este servidor físico del que estoy hablando

y las características del hardware

del servidor físico subyacente.

Ahora, el tipo de instancia también determina otras características

o capacidades cuando se trata de CPU, memoria y almacenamiento.

Y cada tipo de instancia ofrece un cálculo diferente,

capacidades de memoria y almacenamiento,

y dependiendo de estas capacidades,

se agrupan en familias de instancias.

Entonces, ¿cómo sabes cuál elegir?

Bueno, la respuesta simple es seleccionar un tipo de instancia

en función de los requisitos de su aplicación.

Así que selecciona un tipo de instancia en función de los requisitos

de la aplicación que planea ejecutar en su instancia.

Y hay una gran variedad, y está creciendo constantemente

a medida que sacan a la luz más y más tipos de instancias.

Y la razón por la que tienen tantos es porque

los tipos de instancias están optimizados para adaptarse a diferentes casos de uso,

y te dan la flexibilidad de elegir

la combinación adecuada de recursos para sus aplicaciones.

Pero la mejor manera de entender los diferentes

tipos de instancias que están disponibles

es mirar la última documentación de AWS,

y he incluido un enlace a esta documentación

en los recursos para esta lección.

Ahora lo primero que podrías notar

cuando empiezas a crear instancias EC2

es que hay muchos tipos de instancias diferentes,

pero todos están organizados en familias de instancias.

En primer lugar, tenemos un propósito general,

que proporciona un equilibrio de cómputo, memoria y red,

y es una excelente opción para muchas aplicaciones,

incluyendo bases de datos pequeñas o medianas,

servidores web o repositorios de código.

Y si selecciona las diferentes pestañas aquí abajo--

así que tenemos los tipos de instancias de Mac, T2, 3 y 4,

los tipos de instancia M, y A también--

si selecciona una de estas pestañas,

Puedes echar un vistazo a las especificaciones.

para cada tipo de instancia en esta familia.

Y si se desplaza hacia abajo aquí, le dará algunos casos de uso.

para este tipo de instancia también.

Así que eso es propósito general.

Y luego tenemos computación optimizada,

y estos están diseñados para tener una mayor potencia de CPU,

por lo que están diseñados para aplicaciones de cómputo intensivo.

Y estos van a ser geniales para aplicaciones

que tienen altos requisitos de CPU,

por ejemplo, procesamiento por lotes,

análisis distribuido, transcodificación de medios,

o aplicaciones científicas o de ingeniería de alto rendimiento.

Y estos tipos de instancias, principalmente comienzan con C,

pero también tienes HPC,

o computación de alto rendimiento también.

Pasando a la memoria optimizada.

Y estos tipos de instancias están diseñados

para ofrecer un rendimiento rápido para las cargas de trabajo

que procesan grandes conjuntos de datos en la memoria,

así que piense en cualquier carga de trabajo intensiva en memoria,

o bases de datos en memoria, por ejemplo.

Pasando a la computación acelerada.

Y esta es una familia de instancias.

que incluyen aceleradores de hardware especiales,

o procesos especiales que están optimizados para realizar cargas de trabajo

como procesamiento de gráficos, coincidencia de patrones de datos,

o cálculos de números de coma flotante.

Y estos también son excelentes para las cargas de trabajo de aprendizaje automático,

e incluso hay algunos tipos de instancias

en esta familia que incluyen CPU que fueron diseñadas

desde cero para admitir cargas de trabajo de aprendizaje profundo

o cargas de trabajo de aprendizaje automático también.

Y luego aquí, también tenemos instancias,

que incluyen arreglos de puertas programables en campo.

Y este es un procesador o tarjeta especial

que se puede programar para ejecutar un procesamiento específico

para su aplicación.

Y es realmente bueno para aplicaciones que necesitan

poder de procesamiento masivamente paralelo,

cosas como la genómica, el análisis de datos,

o procesamiento de video en tiempo real, o computación financiera.

Y en realidad, hace mucho tiempo

cuando trabajaba en banca de inversión,

utilizamos matrices de puertas programables de campo para acelerar

algunas de las transacciones algorítmicas que hicimos.

Así que eso es latencia realmente baja, comercio de alto rendimiento,

y como puedes imaginar,

es un negocio altamente competitivo.

Entonces, cualquier ventaja competitiva que pueda obtener

procesando su algoritmo y ejecutando sus análisis

unos milisegundos más rápido que su competidor

puede darle una gran ventaja en el mercado.

Y luego aquí, también tenemos algunos tipos de instancias,

que están diseñados específicamente

para cargas de trabajo de transcodificación de video en tiempo real.

Y finalmente, tenemos el

tipos de instancias optimizadas para almacenamiento,

y estos básicamente te permiten agregar grandes cantidades

de almacenamiento a su instancia EC2.

Para que pueda seleccionar una instancia en función de la cantidad

y tipo de almacenamiento que necesita,

cuántos discos desea y cuánto almacenamiento en total

es necesario para su aplicación.

Así que aquí es cuando necesita opciones de almacenamiento adjunto directo

para aplicaciones con E/S de disco muy específicas,

y requisitos de capacidad de almacenamiento.

Estos son geniales para bases de datos NoSQL,

almacenamiento de datos, Elasticsearch, big data, análisis,

y ese tipo de cosa.

Así que ese es un repaso muy rápido

de todas las diferentes familias de instancias.

Pero no te preocupes, como te dije antes,

no es necesario recordarlos todos--

solo tenga una conciencia general de que hay un tipo de instancia

para casi cualquier tipo de carga de trabajo que puedas imaginar.

Y para mis consejos de examen, recuerde el tipo de instancia

determina la configuración del hardware

y capacidades de la computadora anfitriona

donde se ejecuta su instancia o máquina virtual.

Cada tipo de instancia ofrece una computación diferente,

capacidades de memoria y almacenamiento,

y se agrupan en familias de instancias.

Y debe seleccionar el tipo de instancia

cuál es el más adecuado en función de los requisitos

de su aplicación.

Así que ese es el final de esta lección.

Si tiene alguna pregunta, por favor hágamelo saber.

De lo contrario, no dude en unirse a mí en la próxima lección,

donde nos vamos a ensuciar las manos,

lanzamiento de una instancia EC2.

Gracias.