[ENGLISH]

Hello, Cloud Gurus,

and welcome to this lecture, which is a summary

of the EC2 section, including all of my exam tips.

So EC2 is like a virtual machine,

which is hosted in AWS instead of your own data center.

You can select the capacity that you need right now,

grow and shrink when you need,

pay only for what you use

and wait minutes, not months, so you can be up

and running in just a few minutes rather

than waiting months as you would

if you were installing a server in your own data center.

When it comes to pricing, there are 4

different pricing options to be aware of.

So we've got On-Demand.

And this is where you pay by the hour

or second, depending on the type of instance you run.

And this is great for flexibility.

We've also got Reserved instances,

which allows you to reserve capacity

for 1 or 3 years with up to a 72% discount

on the hourly charge.

And this is great if you have known fixed requirements.

There's also spot instances, which allows you

to purchase unused capacity at a massive

discount of up to 90%.

Prices fluctuate with supply and demand,

but this is great for applications

with flexible start and end times

because as soon as the Spot price exceeds

the maximum you're willing to pay,

the instance will be terminated or stopped.

And finally we have Dedicated.

And this is where you have a physical EC2

server dedicated for your use.

And this is great if you have server bound licenses

that you want to reuse or compliance requirements,

which mean you cannot use multi-tenant hardware.

The instance type determines the hardware

of the host computer,

and each instance type offers different compute,

memory, and storage capabilities

and these are grouped in instance families.

And for the Developer Associate exam,

it is not necessary to memorize all

the different instance types.

Just be aware that you should select

an instance type based on the requirements

of your application.

On to EBS volumes.

And EBS volumes are highly available

and scalable storage volumes that

you can attach to an EC2 instance.

And if you remember, we launched an EC2 instance

and created an EBS volume,

and then we attached the EBS volume to our instance.

And there are a few different types of

EBS volumes to choose from.

And EBS of course stands for Elastic Block Store.

So there's gp2, which is suitable for boot disks

and general applications.

And there's also gp3, which is the latest generation of SSD.

And it's also suitable for boot disks

and general applications.

But the difference with this one is that

you get a baseline of 3000 IOPS for all volumes

no matter what their size,

and it is currently 20% cheaper than gp2

because they want to encourage everybody

to use the latest generation.

But both gp2 and gp3 are both limited

to 16,000 IOPS per volume.

So if your application requires greater

than 16,000 IOPS, then you should go for provisioned IOPS.

And with provisioned IOPS, we've got io1

and io2, and io2 is the latest generation.

And these are suitable for online transaction processing

and latency sensitive applications.

And you get up to 64,000 IOPS per volume.

And there's also io2 Block Express,

which is designed for the largest, most critical

and high-performance applications like SAP HANA, Oracle

and Microsoft SQL server, etc.,

with volume sizes of up to 64 terabytes

and 256,000 IOPS per volume.

And this really is SAN in the cloud performance.

And for HDD volumes or hard disk drives,

we've got st1, also known as Throughput Optimized HDD,

and these are all suitable for big data,

data warehouses, and extract transform load workloads.

The maximum throughput is 500 megs per second per volume.

And these cannot be used as a boot volume,

and you get up to 99.9% durability.

And there's also sc1.

And this is also known as Cold HDD.

You get a maximum Throughput of 250 megs per second

per volume, but suitable for less frequently access data.

They cannot be a boot volume.

They come in at the lowest cost

and you get up to 99.9% durability.

On to EBS snapshots.

And an EBS snapshot is a point in time

copy of an EBS volume.

And they are great for backing up your EBS volumes.

And you can use a snapshot to create a new EBS volume.

And if you create a new EBS volume

from an encrypted snapshot, then you will get

an encrypted volume.

And if you create a new EBS volume

from an un-encrypted snapshot,

then you will get an un-encrypted volume.

On to elastic load balancer.

Beginning with Application Load Balancers.

And these provide intelligent load balancing for HTTP

and HTTPS, and they allow you to route

requests to a specific web server based on

the request type.

And think of my example of a car dealership

website where you've got sales information,

loan applications, and repairs and servicing,

and you want to route the incoming requests

to the appropriate application server.

There's also Network Load Balances

and these provide high-performance

load balancing for TCP traffic.

So this is the low-latency option

and it's also the most expensive.

We've got the Classic Load Valancer

and this is the legacy option, which supports

basic HTTP, HTTPS, and TCP load balancing.

And although it's the legacy option,

it may still appear in the exam.

And we've also got Gateway Load Balancers

and these provide load balancing

for third-party virtual appliances,

for example, virtual appliances that

you might find on the AWS marketplace.

And if you need to find the IPv4 address

of the end user for the requests coming

into your load balancer, then look for

the X-Forwarded-For HTTP header.

And this is supported by Application

and Classic Load Balancers.

And if you see a 504 Error with the message

Gateway timeout, then that means

your application is not responding within

the time out period

and you will need to troubleshoot

your application or database server, fix the problem,

and that should clear the error.

So that's it for part 1 of our EC2 summary.

And if you're ready to continue on to part 2,

I will see you in the next lecture.

Thank you.

[SPANISH]

Hola, gurús de la nube,

y bienvenidos a esta conferencia, que es un resumen

de la sección EC2, incluidos todos mis consejos para el examen.

EC2 es como una máquina virtual,

que está alojado en AWS en lugar de su propio centro de datos.

Puede seleccionar la capacidad que necesita ahora mismo,

crecer y encogerse cuando lo necesites,

paga solo por lo que usas

y espera minutos, no meses, para que puedas estar despierto

y funcionando en solo unos minutos en lugar

que esperar meses como lo harías

si estuviera instalando un servidor en su propio centro de datos.

Cuando se trata de precios, hay 4

diferentes opciones de precios a tener en cuenta.

Así que tenemos On-Demand.

Y aquí es donde pagas por hora

o segundo, según el tipo de instancia que ejecute.

Y esto es genial para la flexibilidad.

También tenemos instancias reservadas ,

que permite reservar capacidad

por 1 o 3 años con hasta un 72% de descuento

en el cargo por hora.

Y esto es genial si conoce requisitos fijos.

También hay instancias puntuales, lo que le permite

para comprar capacidad no utilizada a un precio masivo

descuento de hasta el 90%.

Los precios fluctúan con la oferta y la demanda,

pero esto es genial para aplicaciones

con horas de inicio y finalización flexibles

porque tan pronto como el precio Spot supere

el máximo que estás dispuesto a pagar,

la instancia será terminada o detenida.

Y finalmente tenemos Dedicado .

Y aquí es donde tienes un EC2 físico

servidor dedicado para su uso.

Y esto es genial si tiene licencias vinculadas al servidor

que desea reutilizar o requisitos de cumplimiento,

lo que significa que no puede usar hardware multiusuario.

El tipo de instancia determina el hardware.

de la computadora anfitriona,

y cada tipo de instancia ofrece computación diferente,

de memoria y almacenamiento

y estos se agrupan en familias de instancias.

Y para el examen de Developer Associate,

no es necesario memorizar todo

los diferentes tipos de instancias.

Solo tenga en cuenta que debe seleccionar

un tipo de instancia basado en los requisitos

de su aplicación.

A los volúmenes de EBS.

Y los volúmenes de EBS están altamente disponibles

y volúmenes de almacenamiento escalables que

puede adjuntar a una instancia EC2.

Y si recuerdas, lanzamos una instancia EC2

y creó un volumen de EBS,

y luego adjuntamos el volumen de EBS a nuestra instancia.

Y hay algunos tipos diferentes de

Volúmenes de EBS para elegir.

Y EBS, por supuesto, significa Elastic Block Store.

Entonces está gp2, que es adecuado para discos de arranque

y aplicaciones generales.

Y también está gp3, que es la última generación de SSD.

Y también es adecuado para discos de arranque.

y aplicaciones generales.

Pero la diferencia con este es que

obtiene una línea base de 3000 IOPS para todos los volúmenes

no importa cuál sea su tamaño,

y actualmente es un 20% más barato que gp2

porque quieren animar a todos

para utilizar la última generación.

Pero tanto gp2 como gp3 son limitados.

a 16.000 IOPS por volumen.

Entonces, si su aplicación requiere mayor

de 16 000 IOPS, entonces debe elegir IOPS aprovisionadas.

Y con IOPS aprovisionadas, tenemos io1

e io2, y io2 es la última generación.

Y estos son adecuados para el procesamiento de transacciones en línea.

y aplicaciones sensibles a la latencia.

Y obtiene hasta 64 000 IOPS por volumen.

Y también está io2 Block Express,

que está diseñado para los más grandes, más críticos

y aplicaciones de alto rendimiento como SAP HANA, Oracle

y servidor Microsoft SQL, etc.,

con tamaños de volumen de hasta 64 terabytes

y 256 000 IOPS por volumen.

Y esto realmente es SAN en el rendimiento de la nube.

Y para volúmenes HDD o unidades de disco duro,

tenemos st1, también conocido como HDD de rendimiento optimizado,

y todos estos son adecuados para big data,

de datos y cargas de trabajo de carga de transformación de extracción.

El rendimiento máximo es de 500 megas por segundo por volumen.

Y estos no se pueden usar como un volumen de arranque,

y obtienes hasta un 99,9% de durabilidad.

Y también está sc1.

Y esto también se conoce como Cold HDD.

Obtiene un rendimiento máximo de 250 megas por segundo

por volumen, pero adecuado para datos de acceso menos frecuente.

No pueden ser un volumen de arranque.

Vienen al menor costo

y obtienes hasta un 99,9% de durabilidad.

En las instantáneas de EBS.

Y una instantánea de EBS es un punto en el tiempo

copia de un volumen de EBS.

Y son geniales para realizar copias de seguridad de sus volúmenes de EBS.

Y puede usar una instantánea para crear un nuevo volumen de EBS.

Y si crea un nuevo volumen de EBS

de una instantánea cifrada, obtendrá

un volumen cifrado.

Y si crea un nuevo volumen de EBS

de una instantánea sin cifrar,

entonces obtendrá un volumen sin cifrar.

En el equilibrador de carga elástico.

Comenzando con balanceadores de carga de aplicaciones.

Y estos proporcionan equilibrio de carga inteligente para HTTP

y HTTPS, y le permiten enrutar

solicitudes a un servidor web específico basado en

el tipo de solicitud.

Y piensa en mi ejemplo de un concesionario de automóviles

sitio web donde tienes información de ventas,

de préstamo , y reparaciones y servicios,

y desea enrutar las solicitudes entrantes

al servidor de aplicaciones adecuado.

También hay balances de carga de red

y estos proporcionan un alto rendimiento

de carga para el tráfico TCP.

Así que esta es la opción de baja latencia

y también es el más caro.

Tenemos el Classic Load Valancer

y esta es la opción heredada, que admite

básico de HTTP, HTTPS y TCP.

Y aunque es la opción heredada,

puede aparecer en el examen.

Y también tenemos Gateway Load Balancers

y estos proporcionan equilibrio de carga

para dispositivos virtuales de terceros,

por ejemplo, dispositivos virtuales que

puede encontrar en el mercado de AWS.

Y si necesita encontrar la dirección IPv4

del usuario final para las solicitudes que llegan

en su balanceador de carga, luego busque

el encabezado HTTP X-Forwarded-For.

Y esto es compatible con la aplicación.

y balanceadores de carga clásicos.

Y si ve un error 504 con el mensaje

Tiempo de espera de puerta de enlace, entonces eso significa

su aplicación no está respondiendo dentro

el tiempo de espera

y tendrá que solucionar problemas

su aplicación o servidor de base de datos, solucione el problema,

y eso debería borrar el error.

Así que eso es todo para la parte 1 de nuestro resumen de EC2.

Y si está listo para continuar con la parte 2,

Te veré en la próxima conferencia.

Gracias.