[ENGLISH]

Hello, Cloud Gurus, and welcome to this lesson,

which is going to introduce Kinesis.

So, what is Kinesis?

Well, it's a family of services

that enables you to collect, process,

and analyze streaming data in real time.

And it allows you to build custom applications

for your own business needs,

and make decisions or take action based on the data

and the analysis of that data that you are streaming.

And when I first came across Kinesis,

I wondered, "why did they choose that name?"

Well, Kinesis is originally a Greek word,

and it means movement or motion.

And it's actually a really good name for this service,

because Kinesis deals with data that is in motion.

So in other words, with streaming data

rather than data that is static.

For instance, stored in S3 or Elastic Block Store

or in a database.

So Kinesis is all about collecting

and analyzing streaming data.

But what do I mean by streaming data?

Well, in this context,

I mean data that is generated continuously

by thousands of data sources,

which typically send the data records simultaneously

and in small sizes.

So think in the order of kilobytes.

And to give you a few examples,

it's things like a stream of financial transactions.

It could be stock prices,

in-game data as the gamer plays,

social media feeds,

location tracking data

(for example, Uber or Google Maps),

IoT sensors

(So think about sensors inside a factory,

maybe detecting air temperature or air quality),

clickstream data,

and log files.

So this is all fairly small amounts of data.

It's often unstructured data.

And you might want to collect

and provide a constant supply,

or stream, of this data into an application

for analysis and processing.

And this is all about high-throughput information.

And Kinesis allows you to collect

and analyze this kind of data in real time,

so that you can make business decisions

and take action based on your analysis.

Now, Kinesis isn't just one service.

It's actually a family of services,

and each offering is slightly different.

But the one thing that they have in common

is that they are all dealing with streaming data.

So first of all, we've got Kinesis Streams,

and this enables you to stream data or video

to allow you to build custom applications

which process that data in real time.

And there are actually 2 different options

with Kinesis Streams.

So we've got Kinesis Data Streams,

which deals with data.

And Kinesis Video Streams,

which is optimized for video data.

We've got Kinesis Data Firehose,

and this allows you to capture, transform,

and load data streams into AWS data stores

to enable near real-time analytics

using business intelligence tools.

So Firehose is all about capturing, transforming,

and loading the data.

So you can transform your raw data into another format,

for instance, Apache Parquet format,

and then load it into S3, Redshift, or OpenSearch

directly from Kinesis Data Firehose.

And then, finally, we have Kinesis Data Analytics.

And with this, you can analyze, query,

and transform your streamed data in real time

using standard SQL,

and then you can store the results in an AWS data store.

So with Kinesis Data Analytics,

you can interactively query your streaming data

as it comes in,

and then store the results in an AWS service,

like S3, Redshift, or Amazon OpenSearch.

So Kinesis Data Analytics is all about real-time analytics.

Now, in the exam,

they do sometimes give you a scenario

and then ask you to select

the most appropriate service to use.

So you really need to know the differences

between the different Kinesis options.

So let's take a look at each of these

in a little bit more detail with some examples,

beginning with Kinesis Data Streams.

So here's an example architecture.

And down on the left,

we've got our data producers.

So these are all the devices that are producing the data.

So that could be EC2 instances,

mobile devices,

laptops,

and IoT devices.

And all of these devices are producing data

and sending it over to Kinesis Data Streams.

Now, Kinesis Streams retains the data

by default for 24 hours,

with a maximum of 365 days retention.

And the data is stored in these things called shards.

And a shard is basically a sequence of data records.

Each data record has a unique sequence number,

and a Kinesis stream is made up of 1 or more shards.

But don't worry too much about that right now,

because we're going to look at shards

in more detail in the next slide.

So the data is stored in these shards,

and it's retained for 24 hours by default.

And once our data is in Kinesis,

if we want to do anything with it,

we'll need to introduce some data consumers.

And in this example, we've got some EC2 instances

and Lambda functions as well,

which are going to consume the data from Kinesis.

And the consumers basically take the data from the shard

and process it.

For example, they might be running some algorithm

on stock prices,

running sentiment analysis on a social media feed,

or analyzing clickstream data

to generate product recommendations.

And then, once the data consumers

have completed their calculations,

they can then save the data to permanent storage.

For example, to DynamoDB, S3,

Elastic MapReduce, or Redshift.

And these are just a few examples,

because you can save the data wherever you want to.

And Kinesis Data Streams

is basically facilitating this whole architecture

by making it really easy to collect and stream the data,

so that your application can then consume it

and process it however you like

and then go on to store it in a permanent place.

So now let's circle back to Kinesis shards

and take a look at where they fit in.

Now, shards only apply to Kinesis Streams,

so they don't appear in Kinesis Data Analytics

or Kinesis Firehose.

So if you see anything in the exam relating to shards,

then you know they are referring to Kinesis Streams.

So Kinesis Streams are made up of shards,

and each shard is a sequence of 1 or more data records.

And each shard provides a fixed unit of capacity.

For example, with each shard,

you get 5 reads per second

with a maximum total read rate of 2 megabytes per second.

And you also get 1,000 writes per second

with a maximum total rate

of 1 megabyte per second for writes.

And don't worry,

you don't need to remember these numbers for the exam.

And the main thing I want you to remember

is that the data capacity of a Kinesis stream

is determined by the number of shards it has.

And as the data rate increases,

you can increase capacity on your stream

by increasing the number of shards,

allowing your stream to consume more data.

And then finally, with Kinesis Data Streams,

each record has its own unique sequence number

that identifies it.

So the order of records is always maintained,

and you can process the data in sequence order.

So, what about Kinesis Video Streams?

Well, Kinesis Video Streams allows you

to securely stream video from connected devices to AWS.

And these videos can be used

for analytics and machine learning.

And the main thing to remember

is that if you see a scenario in the exam

related to streaming video content and analyzing the videos,

then I want you to think Kinesis Video Streams.

And they might try to confuse you with Kinesis Data Streams

or Kinesis Data Analytics.

But if you're dealing with video data

that needs to be streamed for analysis by your application,

then think Kinesis Video Streams.

Moving on to Kinesis Data Firehose.

Here are the data producers.

And they're sending data to Kinesis Data Firehose.

And with Firehose, there are no shards,

and all the capacity in sizing is automated for you.

And you don't even need consumers

to consume the data, either.

You can optionally use a Lambda function

to process or analyze the data in real time

before sending the data to be stored in S3.

And with Firehose, there is no data retention at all.

So when the data is received by Firehose,

it's either picked up by Lambda and processed,

or it's going to be saved directly to AWS storage.

For instance, you can save it to S3

and then load it into Redshift from S3.

Or it could be saved directly to Amazon OpenSearch,

where you can analyze the data

using your own business intelligence tools.

And Firehose is very much an automated setup.

So you're using Lambda for any processing,

and then you're saving the data

straight to permanent storage,

with no data consumers involved.

So unlike Kinesis Data Streams,

there are no shards and no consumers.

And, after you've saved your data to S3,

to Redshift, or OpenSearch,

you can then analyze it

using your own business intelligence tools.

So moving on to Kinesis Data Analytics.

And this can sometimes come up in the exam

just as a high-level concept,

so it's worth knowing about just from the high level.

So we've got our data producers on the left,

and they're producing data,

which is being sent to Kinesis Streams

and Kinesis Data Firehose.

And then this is where Kinesis Data Analytics fits in.

So Kinesis Data Analytics is going to allow you

to run SQL queries on your data

as it comes in from Kinesis Data Firehose

and Kinesis Data Streams.

And then you can store the results of your query

in S3,

or Redshift,

or Amazon OpenSearch.

So it's a way of analyzing the data inside Kinesis

using standard SQL queries.

And it works with Kinesis Streams,

as well as Firehose.

And this is all about real-time analytics.

So if you see anything in the exam

relating to real-time analytics with streaming data

and particularly, running standard SQL queries on your data

as it comes in from Firehose or Kinesis Streams,

then think Kinesis Data Analytics.

So onto my exam tips.

And the main thing I would like you

to take away from this lesson

is the differences between these services.

So we've got Kinesis Data Streams and Video Streams,

which enable you to capture

and store streaming video and data.

And consumer applications can then process

and analyze the data in real time.

We've then got Kinesis Data Firehose,

which allows you to capture, transform,

and load data continuously into AWS data stores.

And your existing business intelligence applications

and tools can then be used

for near real-time analytics on the stored data.

And then, finally, there's Kinesis Data Analytics,

which provides real-time analytics

using standard SQL queries

on data received by Kinesis Data Streams

and Kinesis Data Firehose.

And it goes on to store the processed data

in an AWS data store,

like S3, Redshift, or OpenSearch.

So that is the end of this lesson.

If you have any questions, please let me know.

Otherwise, feel free to move on to the next lesson.

Thank you.

[SPANISH]

Hola, Cloud Gurus, y bienvenidos a esta lección.

que va a presentar Kinesis.

Entonces, ¿qué es Kinesis?

Bueno, es una familia de servicios.

que le permite recopilar, procesar,

y analice la transmisión de datos en tiempo real.

Y te permite crear aplicaciones personalizadas.

para sus propias necesidades comerciales,

y tomar decisiones o emprender acciones basadas en los datos

y el análisis de esos datos que estás transmitiendo.

Y cuando me encontré por primera vez con Kinesis,

Me pregunté, "¿por qué eligieron ese nombre?"

Bueno, Kinesis es originalmente una palabra griega,

y quiere decir movimiento o movimiento.

Y en realidad es un muy buen nombre para este servicio,

porque Kinesis trata con datos que están en movimiento.

En otras palabras, con la transmisión de datos

en lugar de datos que son estáticos.

Por ejemplo, almacenado en S3 o Elastic Block Store

o en una base de datos.

Así que Kinesis se trata de coleccionar

y analizar datos de transmisión.

Pero, ¿qué quiero decir con transmisión de datos?

Bueno, en este contexto,

Me refiero a datos que se generan continuamente.

por miles de fuentes de datos,

que normalmente envían los registros de datos simultáneamente

y en tallas pequeñas.

Así que piensa en el orden de los kilobytes.

Y para darle algunos ejemplos,

son cosas como un flujo de transacciones financieras.

Podrían ser los precios de las acciones,

datos del juego mientras el jugador juega,

fuentes de redes sociales,

datos de seguimiento de ubicación

(por ejemplo, Uber o Google Maps),

sensores IoT

(Así que piense en los sensores dentro de una fábrica,

tal vez detectando la temperatura del aire o la calidad del aire),

datos de flujo de clics,

y archivos de registro.

Así que todo esto son cantidades bastante pequeñas de datos.

A menudo son datos no estructurados.

Y es posible que desee recopilar

y proporcionar un suministro constante,

o flujo, de estos datos en una aplicación

para análisis y procesamiento.

Y todo esto se trata de información de alto rendimiento.

Y Kinesis le permite recopilar

y analizar este tipo de datos en tiempo real,

para que puedas tomar decisiones de negocio

y tome medidas basadas en su análisis.

Ahora, Kinesis no es solo un servicio.

En realidad es una familia de servicios,

y cada oferta es ligeramente diferente.

Pero lo único que tienen en común

es que todos se ocupan de la transmisión de datos.

En primer lugar, tenemos Kinesis Streams,

y esto le permite transmitir datos o video

para permitirle crear aplicaciones personalizadas

que procesan esos datos en tiempo real.

Y en realidad hay 2 opciones diferentes.

con flujos de Kinesis.

Así que tenemos Kinesis Data Streams,

que se ocupa de los datos.

y secuencias de vídeo de Kinesis,

que está optimizado para datos de video.

Tenemos Kinesis Data Firehose,

y esto te permite capturar, transformar,

y cargue flujos de datos en almacenes de datos de AWS

para permitir análisis casi en tiempo real

usando herramientas de inteligencia de negocios.

Firehose se trata de capturar, transformar,

y cargando los datos.

Para que pueda transformar sus datos sin procesar en otro formato,

por ejemplo, formato Apache Parquet,

y luego cárguelo en S3, Redshift u OpenSearch

directamente desde Kinesis Data Firehose.

Y luego, finalmente, tenemos Kinesis Data Analytics.

Y con esto puedes analizar, consultar,

y transforme sus datos transmitidos en tiempo real

utilizando SQL estándar,

y luego puede almacenar los resultados en un almacén de datos de AWS.

Entonces, con Kinesis Data Analytics,

puede consultar interactivamente sus datos de transmisión

como entra,

y luego almacenar los resultados en un servicio de AWS,

como S3, Redshift o Amazon OpenSearch.

Entonces, Kinesis Data Analytics tiene que ver con el análisis en tiempo real.

Ahora, en el examen,

a veces te dan un escenario

y luego pedirle que seleccione

el servicio más adecuado a utilizar.

Así que realmente necesitas saber las diferencias.

entre las diferentes opciones de Kinesis.

Así que echemos un vistazo a cada uno de estos

con un poco más de detalle con algunos ejemplos,

comenzando con Kinesis Data Streams.

Así que aquí hay una arquitectura de ejemplo.

Y abajo a la izquierda,

tenemos nuestros productores de datos.

Estos son todos los dispositivos que están produciendo los datos.

Podrían ser instancias EC2,

dispositivos móviles,

portátiles,

y dispositivos IoT.

Y todos estos dispositivos están produciendo datos.

y enviarlo a Kinesis Data Streams.

Ahora, Kinesis Streams retiene los datos

por defecto durante 24 horas,

con un máximo de 365 días de retención.

Y los datos se almacenan en estas cosas llamadas fragmentos.

Y un fragmento es básicamente una secuencia de registros de datos.

Cada registro de datos tiene un número de secuencia único,

y un flujo de Kinesis se compone de 1 o más fragmentos.

Pero no te preocupes demasiado por eso en este momento,

porque vamos a mirar fragmentos

con más detalle en la siguiente diapositiva.

Así que los datos se almacenan en estos fragmentos,

y se conserva durante 24 horas de forma predeterminada.

Y una vez que nuestros datos están en Kinesis,

si queremos hacer algo con eso,

necesitaremos introducir algunos consumidores de datos.

Y en este ejemplo, tenemos algunas instancias EC2

y funciones Lambda también,

que van a consumir los datos de Kinesis.

Y los consumidores básicamente toman los datos del fragmento

y procesarlo.

Por ejemplo, podrían estar ejecutando algún algoritmo

sobre los precios de las acciones,

ejecutar un análisis de sentimiento en un feed de redes sociales,

o analizando los datos del flujo de clics

para generar recomendaciones de productos.

Y luego, una vez que los consumidores de datos

han completado sus cálculos,

luego pueden guardar los datos en un almacenamiento permanente.

Por ejemplo, a DynamoDB, S3,

Elástico MapReduce o Redshift.

Y estos son solo algunos ejemplos,

porque puedes guardar los datos donde quieras.

y flujos de datos de Kinesis

básicamente está facilitando toda esta arquitectura

al hacer que sea realmente fácil recopilar y transmitir los datos,

para que su aplicación luego pueda consumirlo

y procesalo como quieras

y luego pasar a guardarlo en un lugar permanente.

Así que ahora volvamos a los fragmentos de Kinesis

y echa un vistazo a dónde encajan.

Ahora, los fragmentos solo se aplican a Kinesis Streams,

para que no aparezcan en Kinesis Data Analytics

o Kinesis Firehose.

Entonces, si ve algo en el examen relacionado con fragmentos,

entonces sabrá que se refieren a Kinesis Streams.

Kinesis Streams se compone de fragmentos,

y cada fragmento es una secuencia de 1 o más registros de datos.

Y cada fragmento proporciona una unidad fija de capacidad.

Por ejemplo, con cada fragmento,

obtienes 5 lecturas por segundo

con una tasa máxima de lectura total de 2 megabytes por segundo.

Y también obtienes 1,000 escrituras por segundo

con una tasa total máxima

de 1 megabyte por segundo para escrituras.

Y no te preocupes,

no necesita recordar estos números para el examen.

Y lo principal que quiero que recuerdes

es que la capacidad de datos de un flujo de Kinesis

está determinado por el número de fragmentos que tiene.

Y a medida que aumenta la tasa de datos,

puedes aumentar la capacidad en tu transmisión

aumentando el número de fragmentos,

permitiendo que su transmisión consuma más datos.

Y finalmente, con Kinesis Data Streams,

cada registro tiene su propio número de secuencia único

que lo identifica.

Así se mantiene siempre el orden de los registros,

y puede procesar los datos en orden secuencial.

Entonces, ¿qué pasa con Kinesis Video Streams?

Bien, Kinesis Video Streams le permite

para transmitir videos de forma segura desde dispositivos conectados a AWS.

Y estos videos se pueden usar

para análisis y aprendizaje automático.

Y lo principal para recordar.

es que si ves un escenario en el examen

relacionados con la transmisión de contenido de video y el análisis de los videos,

entonces quiero que piense en Kinesis Video Streams.

Y es posible que intenten confundirlo con Kinesis Data Streams

o Kinesis Data Analytics.

Pero si se trata de datos de video

que debe transmitirse para su análisis por parte de su aplicación,

entonces piense en Kinesis Video Streams.

Pasando a Kinesis Data Firehose.

Aquí están los productores de datos.

Y están enviando datos a Kinesis Data Firehose.

Y con Firehose, no hay fragmentos,

y toda la capacidad en dimensionamiento está automatizada para ti.

Y ni siquiera necesitas consumidores

para consumir los datos, tampoco.

Opcionalmente, puede usar una función Lambda

para procesar o analizar los datos en tiempo real

antes de enviar los datos para ser almacenados en S3.

Y con Firehose, no hay ninguna retención de datos.

Entonces, cuando Firehose recibe los datos,

es recogido por Lambda y procesado,

o se guardará directamente en el almacenamiento de AWS.

Por ejemplo, puede guardarlo en S3

y luego cárguelo en Redshift desde S3.

O podría guardarse directamente en Amazon OpenSearch,

donde puedes analizar los datos

usando sus propias herramientas de inteligencia de negocios.

Y Firehose es en gran medida una configuración automatizada.

Entonces, está usando Lambda para cualquier procesamiento,

y luego estás guardando los datos

directamente al almacenamiento permanente,

sin consumidores de datos involucrados.

Entonces, a diferencia de Kinesis Data Streams,

no hay fragmentos ni consumidores.

Y, después de haber guardado sus datos en S3,

a Redshift, o OpenSearch,

entonces puedes analizarlo

usando sus propias herramientas de inteligencia de negocios.

Entonces, pasemos a Kinesis Data Analytics.

Y esto a veces puede surgir en el examen.

como un concepto de alto nivel,

por lo que vale la pena conocerlo solo desde el nivel alto.

Así que tenemos a nuestros productores de datos a la izquierda,

y están produciendo datos,

que se envía a Kinesis Streams

y Kinesis Data Firehose.

Y aquí es donde encaja Kinesis Data Analytics.

Kinesis Data Analytics le permitirá

para ejecutar consultas SQL en sus datos

como viene de Kinesis Data Firehose

y flujos de datos de Kinesis.

Y luego puede almacenar los resultados de su consulta

en S3,

o corrimiento al rojo,

o Amazon OpenSearch.

Es una forma de analizar los datos dentro de Kinesis.

mediante consultas SQL estándar.

Y funciona con Kinesis Streams,

así como Firehose.

Y todo esto se trata de análisis en tiempo real.

Así que si ves algo en el examen

relacionados con el análisis en tiempo real con transmisión de datos

y, en particular, ejecutar consultas SQL estándar en sus datos

como viene de Firehose o Kinesis Streams,

entonces piense en Kinesis Data Analytics.

Así que en mis consejos de examen.

Y lo principal te quisiera

para llevar de esta lección

son las diferencias entre estos servicios.

Así que tenemos Kinesis Data Streams y Video Streams,

que le permiten capturar

y almacenar transmisión de video y datos.

Y las aplicaciones del consumidor pueden procesar

y analizar los datos en tiempo real.

Luego tenemos Kinesis Data Firehose,

que te permite capturar, transformar,

y cargue datos continuamente en almacenes de datos de AWS.

Y sus aplicaciones de inteligencia comercial existentes

y las herramientas se pueden utilizar

para análisis casi en tiempo real de los datos almacenados.

Y luego, finalmente, está Kinesis Data Analytics,

que proporciona análisis en tiempo real

utilizando consultas SQL estándar

sobre los datos recibidos por Kinesis Data Streams

y Kinesis Data Firehose.

Y continúa almacenando los datos procesados.

en un almacén de datos de AWS,

como S3, Redshift u OpenSearch.

Así que ese es el final de esta lección.

Si tiene alguna pregunta, por favor hágamelo saber.

De lo contrario, siéntase libre de pasar a la siguiente lección.

Gracias.