
BIG DATA







RECOGIDA DATOS EN AWS

EDUARD LARA








1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Análisis	Análisis interactivo	 Amazon Athena
	Procesamiento de big data	 Amazon EMR
	Almacenamiento de datos	 Amazon Redshift
	Análisis interactivo	 Amazon Kinesis
	Análisis operacionales	 Amazon OpenSearch Service
	Paneles y visualizaciones	 Amazon QuickSight
	Preparación de datos visuales	 AWS Glue DataBrew

1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Migración de datos	Movimiento de datos en tiempo real	 AWS Glue  Amazon Managed Streaming for Apache Kafka (MSK)  Amazon Kinesis Data Streams  Amazon Kinesis Data Firehose  Amazon Kinesis Video Streams  AWS Database Migration Service
---------------------------	------------------------------------	---

1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Lago de datos	Almacenamiento de objetos	 Amazon S3  AWS Lake Formation
	Copias de seguridad y archivo	 Amazon S3 Glacier  AWS Backup
	Catálogo de datos	 AWS Glue  AWS Lake Formation
	Datos de terceros	 AWS Data Exchange

1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Análisis predictivo y aprendizaje automático	Marcos e interfaces	 AMI de aprendizaje profundo de AWS
	Servicios de la plataforma	 Amazon SageMaker

1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Colección de datos

- ☐ Kinesis Data Streams
- ☐ Kinesis Data Firehose
- ☐ Kinesis Video Streams
- ☐ Kinesis Data Analytics
- ☐ Laboratorio de kinesis Firehouse
- ☐ SQS - Gestión de colas de mensajes
- ☐ IOT - Internet de las cosas
- ☐ Data pipeline
- ☐ Formas de introducir datos en AWS

Almacenamiento de datos

- ☐ Glacier
- ☐ DynamoDB - Introducción
- ☐ DynamoDB - Particiones
- ☐ DynamoDB - Indices secundarios
- ☐ DynamoDB - Streams
- ☐ DynamoDB - Rendimiento
- ☐ DynamoDB - Crear tablas y elementos
- ☐ Laboratorio de DynamoDB

1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Procesamiento de datos

- ☐ EMR - Introducción
- ☐ EMR - Apache Hadoop
- ☐ EMR - Arquitectura
- ☐ EMR - Operaciones
- ☐ EMR - Aplicaciones
- ☐ EMR - Spark
- ☐ EMR - Almacenamiento de ficheros
- ☐ Funciones Lambda para entornos big data
- ☐ Hcatalog y Glue
- ☐ Lab creación cluster EMR de forma rápida
- ☐ Lab creación cluster EMR con opciones avanzadas

1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Análisis de datos

- ☐ Redshift - Introducción
- ☐ Redshift - Arquitectura
- ☐ Redshift - Integración
- ☐ Redshift - Bases de datos en columnas
- ☐ Redshift - Diseño de tablas
- ☐ Redshift - Gestión de carga de trabajo
- ☐ Redshift - Carga de datos
- ☐ Redshift - Mantenimiento
- ☐ Laboratorio de Redshift

1. INTRODUCCIÓN SERVICIOS BIGDATA AWS

Visualización de datos

- ☐ Introducción
- ☐ Laboratorio de Quicksight

2. KINESIS DATA STREAMS

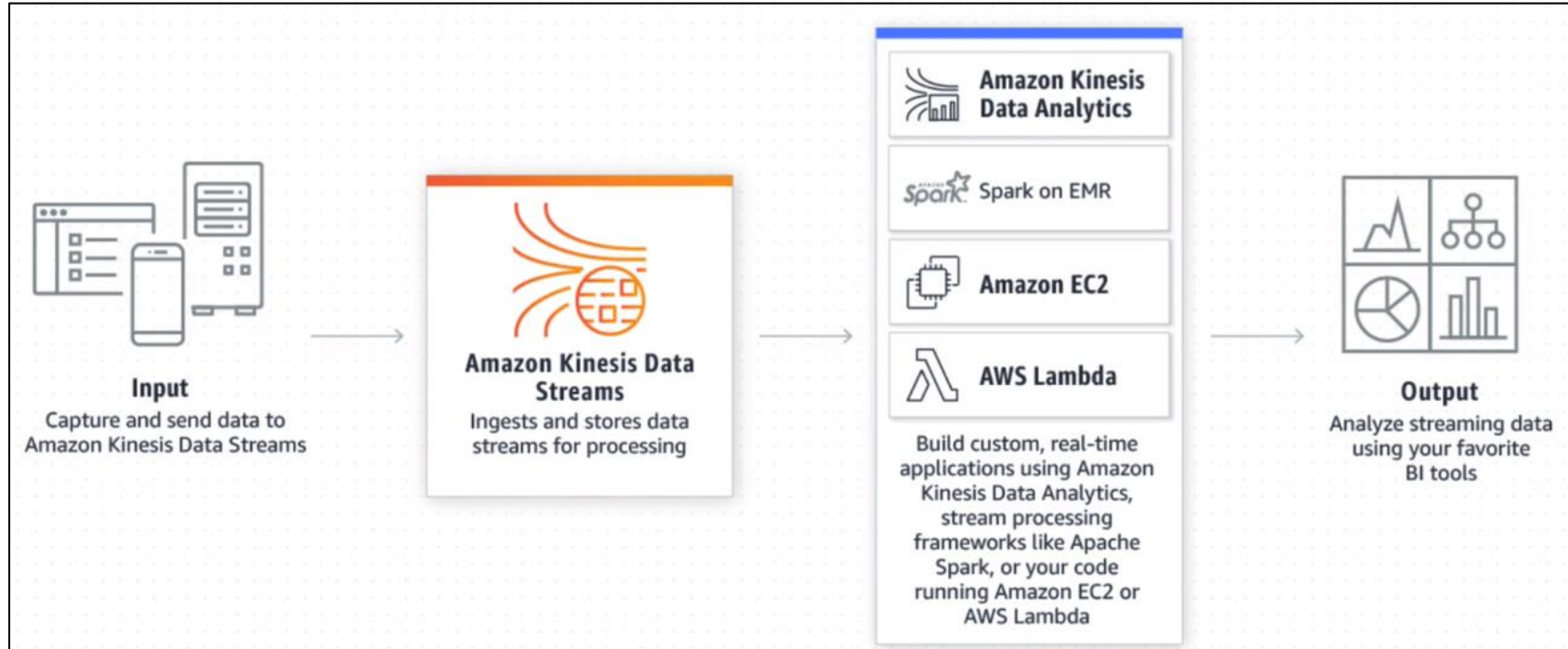
- Permite recopilar datos de streaming a gran escala para el análisis en tiempo real
- Es capaz de registrar de manera continua Gigabytes de datos por segundo de miles de orígenes:
 - Transmisiones de clics de sitios web
 - Transmisiones de eventos de bases de datos
 - Transacciones financieras
 - Fuentes de Redes sociales
 - Registros TIC
 - Eventos de Seguimientos de ubicaciones

2. KINESIS DATA STREAMS

- Los datos recopilados se encuentran disponibles en milisegundos para posibilitar los casos de uso de análisis en tiempo real, como pueden ser:
 - paneles en tiempo real,
 - detección de anomalías en tiempo real
 - precios dinámicos
- Los casos de uso pueden ser:
 - Recopilación de datos de eventos y registros
 - Análisis en tiempo real
 - Registros de datos móviles
 - Fuentes de datos para videojuegos

2. KINESIS DATA STREAMS

- ❖ En el diagrama podemos ver una entrada múltiple de datos, aparece Amazon Kinesis Data Streams
- ❖ Los procesadores pueden ser Amazon EC2, lambda, etc
- ❖ Hay disponibilidad de sacar los datos mediante análisis de datos con gráficos, etc..



3. KINESIS DATA FIREHOSE

- ❖ Sirve para preparar y cargar transmisiones de datos generados en tiempo real en almacenes de datos y servicios de análisis
- ❖ Amazon Kinesis Data Firehose es un servicio completamente administrado para enviar datos de streaming en tiempo real a varios destinos admitidos.
- ❖ Con Kinesis Data Firehose, se pueden configurar los productores de datos para que envíen datos a Kinesis Data Firehose y este entregue automáticamente los datos al destino especificado
- ❖ También puede configurar Kinesis Data Firehose para transformar los datos antes de entregarlos.

3. KINESIS DATA FIREHOSE

Conceptos clave:

- ❑ **Flujo de entrega:** la entidad subyacente de Kinesis Data Firehose. Kinesis Data Firehose se utiliza mediante la creación de un flujo de entrega de Kinesis Data Firehose y el posterior envío de datos al mismo.
- ❑ **Registro:** los datos de interés que el productor de datos envía a un flujo de entrega de Kinesis Data Firehose. Cada registro puede pesar hasta 1 000 KB.

3. KINESIS DATA FIREHOSE

- ❑ **Productor de datos:** los productores envían registros a los flujos de entrega de Kinesis Data Firehose. Por ejemplo, un servidor web que envía datos de registro a un flujo de entrega es un productor de datos. También puede configurar el flujo de entrega de Kinesis Data Firehose para que lea automáticamente los datos de un flujo de datos de Kinesis existente y los cargue en los destinos.
- ❑ **Tamaño e intervalo del búfer:** Kinesis Data Firehose almacena en el búfer los datos de streaming entrantes hasta alcanzar un determinado tamaño o un determinado periodo de tiempo antes de entregarlos en los destinos. El tamaño del búfer se mide en MiB y el intervalo en segundos.

3. KINESIS DATA FIREHOSE

- ❖ Permite utilizar la función AWS lambda para hacer transformaciones de datos si fuera necesario
- ❖ No se paga por servidores, se paga por volumen de datos transmitido
- ❖ Permite registrar, transformar y entregar datos de streaming de manera fiable en almacenes de datos:
 - Amazon S3
 - Amazon Redshift
 - Amazon Elasticsearch
 - Puntos de enlaces HTTP genéricos
 - Proveedores de servicios Datadog, New Relic, MongoDB y Splunk

3. KINESIS DATA FIREHOSE

❖ Casos de uso:

- Análisis de dispositivos IoT
- Análisis de secuencia de clics en sitios web
- Análisis de registros
- Monitorización de la seguridad

3. KINESIS DATA FIREHOSE

- ❖ Diferentes entradas de datos son direccionadas hacia Kinesis data Firehose, que puede transformar los datos mediante AWS Lambda o enviarlos directamente a distintos almacenes de datos: Amazon, S3, RedShift.
- ❖ Podemos hacer un análisis a partir de estos datos almacenados en estos dispositivos de almacenamiento



4. KINESIS VIDEO STREAMS

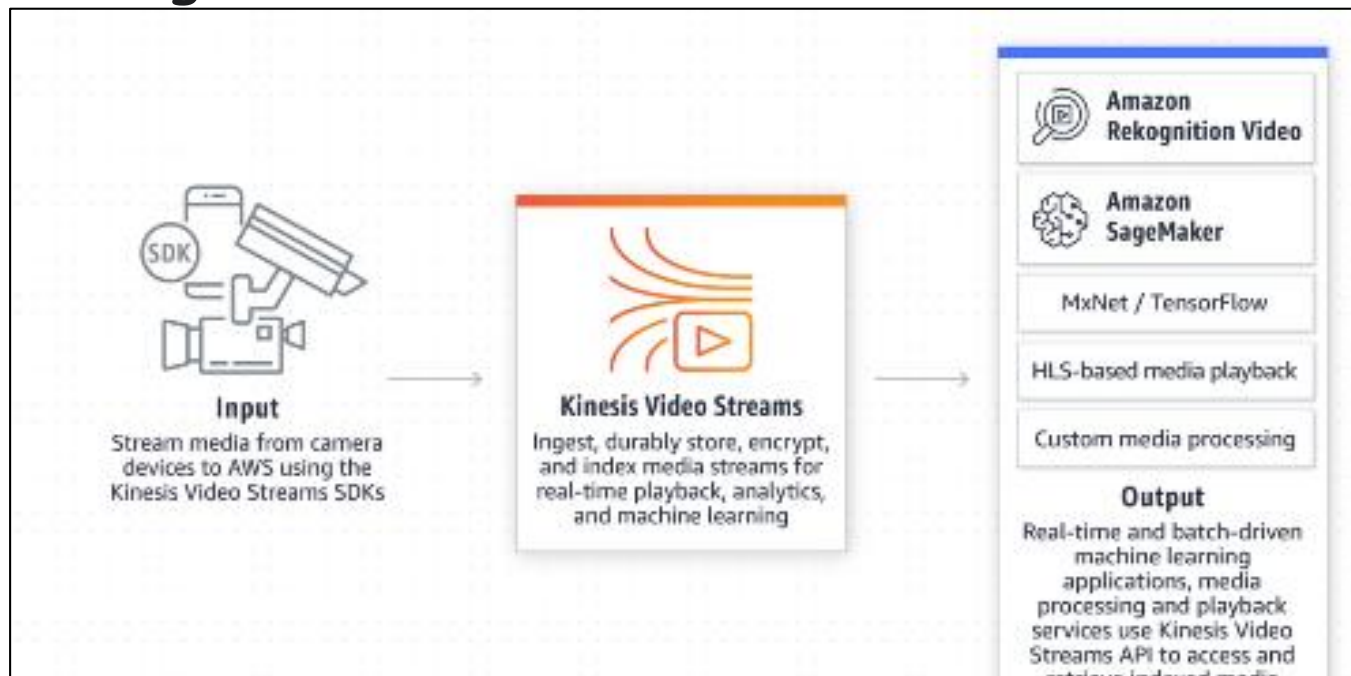
- Permite la transmisión segura de videos desde dispositivos conectados a AWS para tareas de análisis, aprendizaje automático, reproducción y otros procesos.
- Permite transmitir videos desde muchos dispositivos que estén conectados a AWS como pueden ser: teléfonos móviles, cámaras de seguridad, drones, satélites, sensores, etc.
- Permite crear aplicaciones de visión artificial con la capacidad de reconocimiento de video en tiempo real mediante la integración con Amazon Rekognition Video
- Permite la reproducción de transmisiones de videos grabados y en directo

4. KINESIS VIDEO STREAMS

- Crear aplicaciones con transmisiones de medios bidireccionales y en tiempo real como videollamadas
- Casos de uso:
 - Hogar inteligente (aplicaciones inteligentes para el control domótica de su casa como timbre, cámaras de vigilancia, iluminación)
 - Ciudad inteligente (control de las cámaras de semáforos, centros comerciales, espacios públicos)
 - Autorización industrial

4. KINESIS VIDEO STREAMS

- ❖ El servicio de kinesis video streaming para recoger los inputs que pueden ser de una entrada de una cámara que esté conectada a AWS
- ❖ todo ese video puede ser procesado y almacenado y utilizado en las herramientas de Amazon Rekognition Video, SageMaker



5. KINESIS DATA ANALYTICS

- ❑ Permite transformar y analiza datos de streaming en tiempo real
- ❑ Facilita la integración con Apache Flink
- ❑ Permite escalar automáticamente según el volumen de los datos de entrada que tengamos
- ❑ En un servicio sin servidores, no tenemos que preocuparnos de administrar servidores
- ❑ Solo cobra por los recursos utilizados en el streaming de datos
- ❑ Puede crear aplicaciones de consulta de datos mediante SQL: Java, Scala y Python

5. KINESIS DATA ANALYTICS

Casos de uso:

- Streaming de procesos ETL (Extracción, Transformación y Carga)
- Análisis en tiempo real (Monitoreo de registros y análisis web)
- Procesamiento de eventos (Detección de anomalías)

5. KINESIS DATA ANALYTICS

- ❑ Diagrama de utilización de este servicio de Data Analytics donde recibimos datos de diferentes fuentes como pueden ser Kinesis Data Streams.
- ❑ Permite utilizar esta herramienta Kinesis Data Analytics, realizando consultas, analizando estos datos de streaming y permitiendo una salida que puede ser procesada por otras herramientas de análisis.



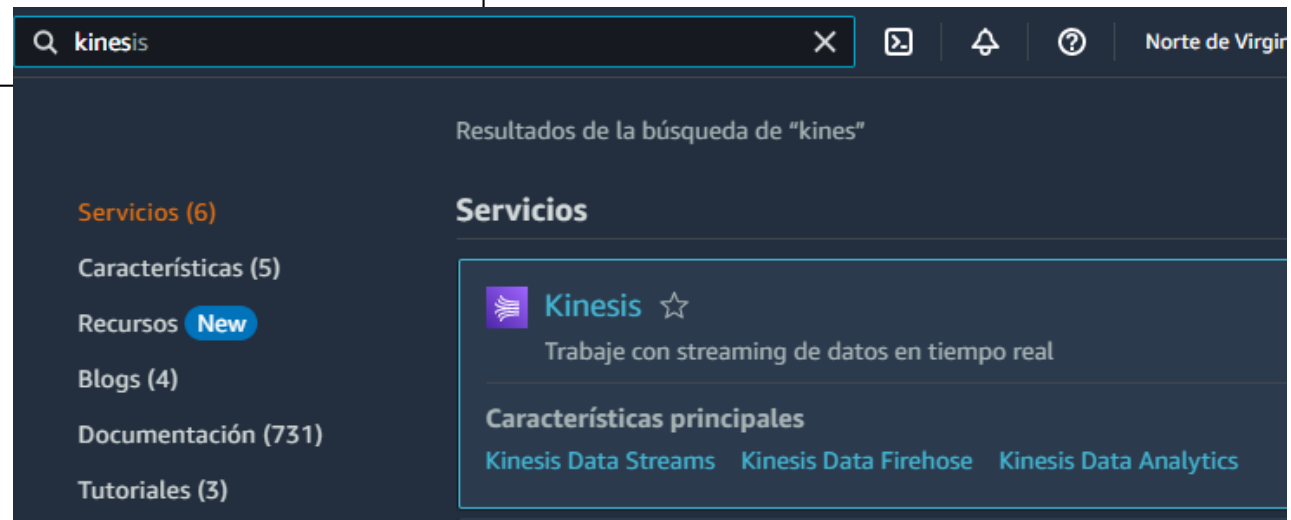
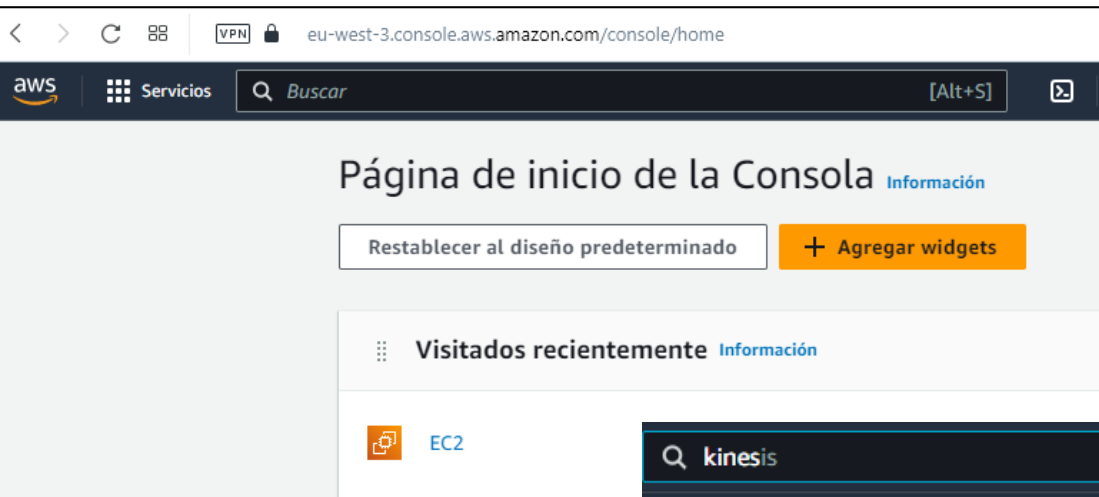
6. LABORATORIO KINESIS FIREHOSE

- ❑ Haremos un laboratorio de Kinesis donde utilizaremos Kinesis Data Firehose para recoger datos que generaremos mediante un programa Python
- ❑ Con Kinesis Data Firehose recuperaremos los datos que genera este fichero en Python y lo guardaremos en S3 para posteriormente realizar un análisis con SQL de los datos.



6. LABORATORIO KINESIS FIREHOSE

Paso 1. Vamos a la pagina de inicio de consola de AWS. Buscamos el servicio de Kinesis en la barra de AWS, y seleccionamos este servicio



6. LABORATORIO KINESIS FIREHOSE

Paso 2. Podemos elegir los tres tipos que hay:

- Secuencia de datos
- Data firehose
- Data Analytics

Elegimos Data Firehose y hacemos click en crear secuencia de entrega. Esto va a recoger datos de una fuente automática y lo va a almacenar en S3.

Analytics

Servicios de Amazon Kinesis

Recopile, procese y analice fácilmente secuencias de datos en tiempo real.

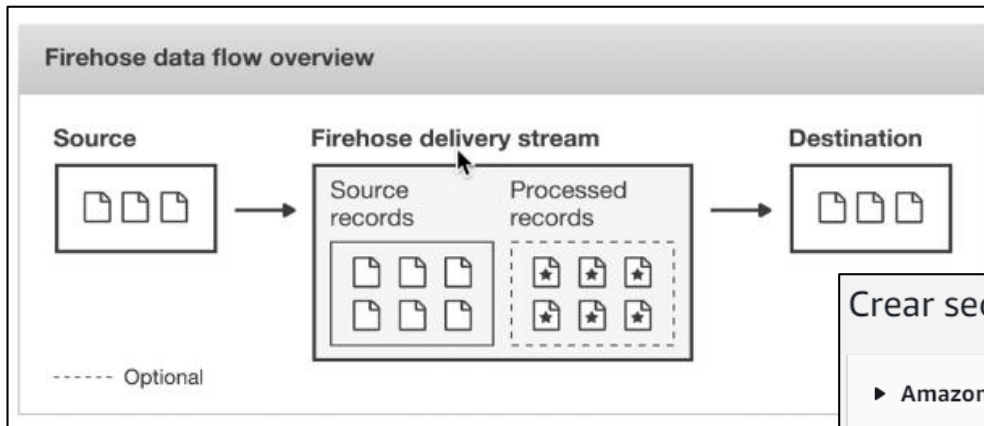
Introducción

- ☒ Kinesis Data Streams
Recopilar datos de streaming con una secuencia de datos.
- ☐ Kinesis Data Firehose
Procese y entregue datos de streaming con una secuencia de entrega de datos.
- ☐ Kinesis Data Analytics
Analice datos de streaming con aplicaciones de análisis de datos.

Crear secuencia de datos

6. LABORATORIO KINESIS FIREHOSE

Paso 3. Los datos que se recogen de una fuente se pueden procesar (si queremos) o transformar y luego se envía al destino. En nuestro caso simplemente lo vamos a pasar por el Firehose y lo vamos a enviar directamente al destino sin transformar.



Crear secuencia de entrega [Información](#)

► Amazon Kinesis Data Firehose: cómo funciona

Elija un origen y un destino

Especifique el origen y el destino de la secuencia de entrega. No puede cambiar el origen y el destino de la secuencia de entrega una vez que esta se haya creado.

Origen [Información](#)

Elija un origen

Destino [Información](#)

Elija un destino

Cancelar

Crear secuencia de entrega

6. LABORATORIO KINESIS FIREHOSE

Paso 4. Como origen de datos marcamos la opción de Direct Put, como origen de datos para la secuencia de entrega. En este caso vamos a hacerlo de forma automática con los datos de ejemplo que vienen en la creación de este Firehose

Como destino, los datos que recoge Firehose los va a guardar en un destino que puede ser Amazon S3, RedShift, Elasticsearch, Http Endpoint o un servicio de terceros, etc. Seleccionamos S3, todo lo que recojamos no lo vamos a transformar y lo guardaremos en S3.

Elija un origen y un destino
Especifique el origen y el destino de la secuencia de entrega. No puede cambiar el origen y el destino de la secuencia de entrega una vez que esta se haya creado.

Origen | Información

Direct PUT ▼

Destino | Información

Amazon S3 ▼

6. LABORATORIO KINESIS FIREHOSE

Paso 5. Ponemos el nombre de kinesis1 a la secuencia de entrega, el nombre de la fuente de datos. El resto de opciones no se habilitan (la transformación de registros con Lambda ni la conversión de formatos de los registros)

Nombre de la secuencia de entrega

Nombre de la secuencia de entrega

kinesis1

Se aceptan letras mayúsculas y minúsculas, números, guiones bajos, guiones y puntos.

Transformar y convertir registros: *opcional*
Configure Kinesis Data Firehose para transformar y convertir los datos de registro.

Transforme los registros de origen con AWS Lambda | [Información](#)

Kinesis Data Firehose puede invocar una función de AWS Lambda para transformar, filtrar, descomprimir, convertir y procesar los registros de datos de origen. La función de AWS Lambda especificada también se puede utilizar para proporcionar claves de particionamiento dinámico para los datos de origen entrantes antes de su entrega al destino especificado.

☐ Habilitar la transformación de datos

Convertir el formato de los registros | [Información](#)

Normalmente, es más eficiente consultar los datos en formato Apache Parquet o Apache ORC que en JSON. Kinesis Data Firehose puede convertir los registros de origen en formato JSON mediante un esquema de una tabla que se haya definido en [AWS Glue](#). En el caso de los registros que no están en formato JSON, cree una función de Lambda que los convierta a JSON en la sección Transformar los registros de origen con AWS Lambda que aparece más arriba.

☐ Habilitar la conversión de formatos de los registros

6. LABORATORIO KINESIS FIREHOSE

Paso 6. En este ejemplo no lo vamos a hacer, pero si quisiéramos podemos habilitar la transformación de datos, eligiendo una función lambda que transforme los datos de entrada en otro formato diferente o con otros datos añadidos. Lo dejamos como estaba

Transforme los registros de origen con AWS Lambda

[Información](#)

Kinesis Data Firehose puede invocar una función de AWS Lambda para transformar, filtrar, descomprimir, convertir y procesar los registros de datos de origen. La función de AWS Lambda especificada también se puede utilizar para proporcionar claves de particionamiento dinámico para los datos de origen entrantes antes de su entrega al destino especificado.

☒ **Habilitar la transformación de datos**

Función de AWS Lambda

Versión o alias

Formato: arn:aws:lambda:[Region]:[AccountId]:function:[FunctionName]

Tamaño del búfer

La función de AWS Lambda tiene una cuota de carga de invocación de 6 MB. El tamaño de los datos puede aumentar después de que la función de AWS Lambda los procese. Un tamaño de almacenamiento en búfer más pequeño permite disponer de más espacio en caso de que los datos se amplíen después del procesamiento.

MB

Mínimo: 0.2 MB, máximo: 3 MB.

Intervalo de almacenamiento en búfer

El periodo durante el cual Kinesis Data Firehose almacena en búfer los datos entrantes antes de invocar la función de AWS Lambda. Esta última se invoca una vez que se alcanza el valor del tamaño del almacenamiento en búfer o el intervalo de almacenamiento en búfer.

segundos

Mínimo: 60 segundos, máximo: 900 segundos.

6. LABORATORIO KINESIS FIREHOSE

Paso 7. Dentro de la configuración de destino, tenemos que elegir un bucket. En este caso creamos un bucket nuevo.

Bucket de S3

Browse

Create

Format: s3://bucket

El nombre de bucket debe ser único para todo el mundo, por ejemplo **vuestro_nombre+ kinesis.bolsa.1** ya que van a ser datos de bolsa. Le damos a crear Bucket con todas las opciones por defecto

Crear bucket [Info](#)

Los buckets son contenedores de datos almacenados en S3. [Más información](#)

Configuración general

Nombre del bucket

El nombre del bucket debe ser único en todo el mundo y no debe contener espacios ni letras mayúsculas. [Consulte las reglas para la denominación de los buckets](#)

Región de AWS

Buckets (1) [Info](#)

Los buckets son contenedores de datos almacenados en S3. [Más información](#)

Nombre

Región de AWS

Acceso

Fecha de creación

<input type="radio"/>	edulara.kinesis1.bolsa.1	EE. UU. Este (Norte de Virginia) us-east-1	Bucket y objetos que no son públicos	25 Apr 2023 8:10:01 PM CEST
-----------------------	--------------------------	--	--------------------------------------	-----------------------------

6. LABORATORIO KINESIS FIREHOSE

Paso 8. Si le damos al enlace del bucket, lo podemos ver en otra pestaña. En principio este bucket estará vacío, todavía no ha recibido ningún dato. Tan pronto reciba datos, el Firehose lo recogerá y los enviará a ese bucket.

edulara.kinesis1.bolsa.1 [Info](#)

[Objetos](#) | [Propiedades](#) | [Permisos](#) | [Métricas](#) | [Administración](#) | [Puntos de acceso](#)

Objetos (0)

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

  Copiar URI de S3  Copiar URL  Descargar  Abrir  Eliminar [Acciones ▼](#) [Crear carpeta](#)

 Cargar

[<](#) **1** [>](#) 

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
No hay objetos					
No tiene objetos en este bucket.					

 Cargar

34

6. LABORATORIO KINESIS FIREHOSE

Paso 9. A continuación asignaremos el bucket creado anteriormente a S3 y añadimos dos prefijos o carpetas a dicho bucket de S3. Una que contendrá los datos, por ejemplo datos/ y otra los errores, errores/. Si hay algún error en la recogida de datos, los meterá en esta carpeta

Bucket de S3	
s3://edulara.kinesis1.bolsa.1	Prefijo de bucket de S3: <i>opcional</i> De forma predeterminada, Kinesis Data Firehose añade el prefijo "AAAA/MM/dd/HH" (en UTC) a los datos que entrega a Amazon S3. Es posible anular este valor predeterminado si se especifica un prefijo personalizado que incluya expresiones que se evalúan en tiempo de ejecución.
Format: s3://bucket	<div>datos/</div> <p>Puede repetir las mismas claves en el prefijo del bucket de S3. Máximo de caracteres del prefijo de bucket de S3: 1024.</p> <p>Prefijo de salida de error de bucket S3: <i>opcional</i> Puede especificar un prefijo de salida de error del bucket de S3 que se utilizará en las condiciones de error. Este prefijo puede incluir expresiones para que Kinesis Data Firehose las evalúe en tiempo de ejecución.</p> <div>errores/</div>

6. LABORATORIO KINESIS FIREHOSE

Paso 10. En Sugerencias del almacenamiento en buffer de S3, elegimos las condiciones del buffer, donde se almacenan los datos. En el tamaño del buffer pondremos el tamaño mínimo de 1MB. En el intervalo del buffer interval, que es el tiempo en segundos que va a retener la información el Firehose (puede ser entre 60 y 900 seg) pondremos el mínimo 60 seg. Firehose no es en tiempo real, porque hay un retardo de 60 seg. Los datos que se recogen se almacenan en este buffer durante 1 min o hasta que llega un máximo de 1 MB y se vuelcan luego en S3.

Tamaño del búfer

El tamaño de búfer más alto puede resultar menos costoso, pero la latencia es mayor. El tamaño de búfer más bajo permitirá una entrega más rápida, con un costo más alto y una menor latencia.

MiB

Mínimo: 1 MiB, máximo: 128 MiB. Valor recomendado: 5 MiB.

Intervalo de almacenamiento en búfer

El intervalo más alto permite más tiempo para recopilar datos y el tamaño de los datos puede ser mayor. El intervalo más bajo envía los datos con mayor frecuencia y puede resultar más conveniente cuando se trata de ciclos más cortos de actividad de datos.

segundos

Mínimo: 60 segundos, máximo: 900 segundos. Valor recomendado: 300 segundos.

6. LABORATORIO KINESIS FIREHOSE

Paso 11. S3 lo podemos comprimir o encriptar. No lo vamos a hacer. En configuración avanzada dejamos habilitado el Registro de errores tal y como esta por defecto

Compresión y cifrado de S3

Kinesis Data Firehose puede comprimir los registros antes de entregarlos al bucket de S3. Los registros se cifran en el bucket de S3 mediante una clave maestra de AWS Key Management Service (KMS).

Compresión para registros de datos

Kinesis Data Firehose puede comprimir los registros antes de entregarlos al bucket

- ☒ No está habilitado
- ☐ GZIP
- ☐ Snappy
- ☐ Zip
- ☐ Snappy compatible con Hadoop

Cifrado de los registros de datos

El registro comprimido se cifra en el bucket de S3 con una clave maestra de KMS.

- ☒ No está habilitado
- ☐ Habilitado

▼ Configuración avanzada

Cifrado del servidor no está habilitada; registro de errores habilitado; rol de IAM KinesisFirehoseServiceRole-kinesis1-us-east-1-1682447209587; no hay etiquetas.

Cifrado del servidor [Información](#)

Puede utilizar AWS Key Management Service (KMS) para crear y administrar las claves maestras de los clientes (CMK) y para controlar el uso del cifrado en una amplia gama de servicios de AWS para las aplicaciones.

- ☐ Habilitar el cifrado del servidor para los registros de origen en la secuencia de entrega

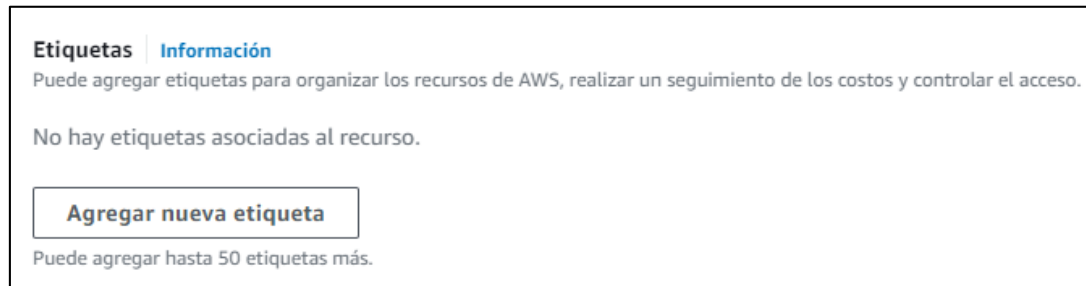
Registro de errores de Amazon CloudWatch [Información](#)

Elija la opción Habilitado si desea que Kinesis Data Firehose registre los errores de entrega de registros en CloudWatch Logs.

- ☐ No está habilitado
- ☒ Habilitado

6. LABORATORIO KINESIS FIREHOSE

Paso 12. En Etiquetas Opcional podemos poner un nombre y un valor. Esto sirve para luego facturar.



Etiquetas | Información

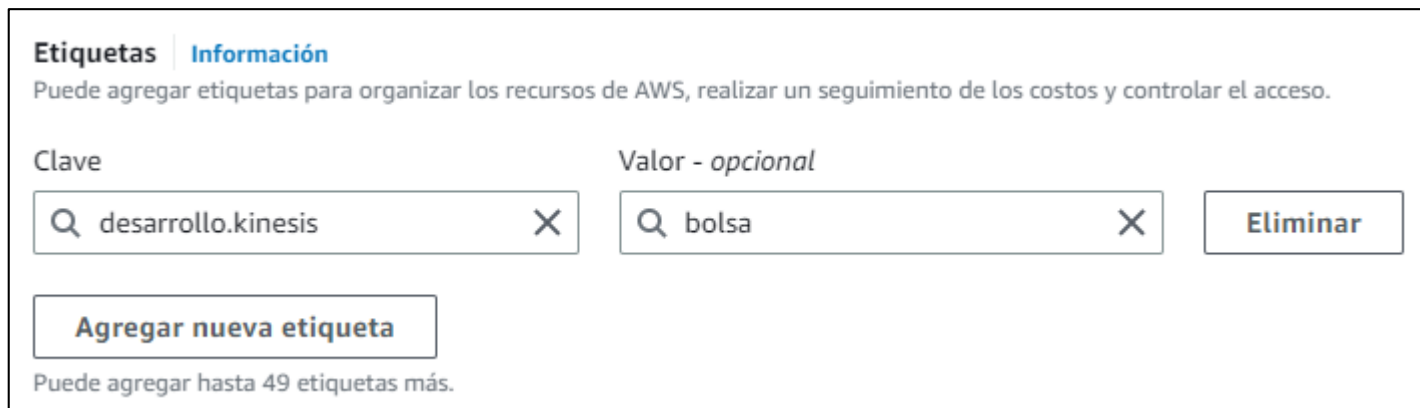
Puede agregar etiquetas para organizar los recursos de AWS, realizar un seguimiento de los costos y controlar el acceso.

No hay etiquetas asociadas al recurso.

[Agregar nueva etiqueta](#)

Puede agregar hasta 50 etiquetas más.

Si tenemos un departamento por ejemplo de desarrollo, podemos añadir un tag de clave desarrollo_kinesis y de valor bolsa, útil si queremos facturar estos elementos de kinesis a un departamento concreto



Etiquetas | Información

Puede agregar etiquetas para organizar los recursos de AWS, realizar un seguimiento de los costos y controlar el acceso.

Clave Valor - *opcional*

[Agregar nueva etiqueta](#)

Puede agregar hasta 49 etiquetas más.

6. LABORATORIO KINESIS FIREHOSE

Paso 13. Para el acceso al servicio debemos elegir el rol de IAM existente LabRole. Tiene los permisos necesarios para que se pueda acceder a S3.

Acceso al servicio | **Información**

Kinesis Data Firehose utiliza este rol de IAM para todos los permisos que la secuencia de entrega necesita. Si desea especificar diferentes roles para los distintos permisos, utilice la API o la CLI.

☐ Cree o actualice el rol de IAM **KinesisFirehoseServiceRole-kinesis1-us-east-1-1682447209587**
Crea un nuevo rol o actualiza uno existente y agrega a este la política requerida, a la vez que permite que Kinesis Data Firehose lo asuma.

☒ Elija un rol de IAM existente
El rol que elija debe tener políticas que incluyan los permisos que Kinesis Data Firehose necesita.

Roles de IAM existentes
Únicamente se pueden seleccionar los roles de IAM con [política de confianza requerida](#).





Ver [LabRole](#) en la consola de IAM.

kinesis1 es el nombre de delivery stream que va a coger los datos directamente y se los va a mandar a S3.

los datos y estos serían los roles y el tamaño del papel 1 60 segundos y lo demás que hemos puesto y nuestro tag y también

6. LABORATORIO KINESIS FIREHOSE

Paso 14. Hacemos click en el botón Crear secuencia de entrega. Se creará nuestro primer Kinesis Firehose con los datos indicados en la configuración. Esto puede tardar unos minutos en estar disponible.

 **Creando kinesis1**
La actualización del estado puede tardar hasta 5 minutos.

[Amazon Kinesis](#) > [Secuencias de entrega](#) > [kinesis1](#)

 kinesis1 se creó correctamente.

[Amazon Kinesis](#) > [Secuencias de entrega](#) > [kinesis1](#)

kinesis1 [Información](#)


[Eliminar la secuencia de entrega](#)

Detalles de la secuencia de entrega

Estado
 **Activo**

Origen
Direct PUT


Destino
Amazon S3

ARN
 `arn:aws:firehose:us-east-1:307729659735:deliverystream/kinesis1`

Transformación de datos
No está habilitado

Particionamiento dinámico
No está habilitado

Tiempo de creación
25 de abril de 2023, 20:53 CEST

Estado de los registros de errores
 **0 Registros de errores de destino**
0

6. LABORATORIO KINESIS FIREHOSE

Paso 15. Si volvemos a entrar vemos que kinesis1, el firehose que acabamos de crear está activo con la dirección de S3 donde se van a guardar los datos

Amazon Kinesis [Información](#)

Amazon Kinesis facilita la recopilación, el procesamiento y el análisis de secuencias de datos en tiempo real para poder obtener información oportuna y reaccionar rápidamente ante información nueva.

Secuencias de datos [Información](#)

Total de secuencias de datos

0

[Crear secuencia de datos](#)

Data Firehose [Información](#)

Total de secuencias de entrega

1

[Crear secuencia de entrega](#)

Data Analytics [Información](#)

Total de aplicaciones de análisis

0

[Crear aplicación](#)

[Amazon Kinesis](#) > [Secuencias de entrega](#)

Secuencias de entrega (1)

[Actualizar](#) [Eliminar](#) [Crear secuencia de entrega](#)

< 1 >

	Nombre ▲	Estado ▼	Tiempo... ▼	Origen ▼	Transfo... ▼	Tipo de... ▼	Destino ▼
<input type="radio"/>	kinesis1	✓ Activo	23 de abril...	Direct PUT	No está ha...	Amazon S3	edula.kinesis...

6. LABORATORIO KINESIS FIREHOSE

Paso 16. Si pulsamos en Kinesis1 hay un apartado que es Probar con datos de demostración. Esto ejecuta un script para poner enviar datos de prueba a Firehose Podemos construir nuestro propio programa o nuestro propio data stream que envíe datos a Firehose o utilizar este test demo que va a enviar datos automáticamente al kinesis recién creado. Los datos que se van a crear son informaciones sobre valores en bolsa: precio del valor en bolsa, el símbolo, el sector, el cambio y el precio actual.

▼ Probar con datos de demostración [Información](#)

Ingiera datos simulados para probar la configuración de la secuencia de entrega. Se aplican los cargos estándar de Amazon Kinesis Data Firehose.

Esta prueba ejecuta un script en el navegador para introducir datos de demostración en la secuencia de entrega de Kinesis Data Firehose, que realiza el envío al destino de Amazon S3.

```
1 {  
2   "TICKER_SYMBOL": "QXZ",  
3   "SECTOR": "HEALTHCARE",  
4   "CHANGE": -0.05,  
5   "PRICE": 84.51  
6 }
```

Paso 1

Comience a enviar datos de demostración a la secuencia de entrega. Si ya se transmiten datos a este destino, los datos de demostración se envían junto con los registros de origen.

Comience a enviar datos de demostración

6. LABORATORIO KINESIS FIREHOSE

Paso 17. Le vamos a enviar datos y lo que hace esto es empezar a enviar datos de la bolsa a nuestro Firehose generado de forma automática, de forma aleatoria. Vemos como está empezando a arrancar los datos. Luego si no nos interesa seguir, podemos darle al stop con lo cual pararla, pero de momento lo vamos a dejar que vaya enviando

Paso 1

Comience a enviar datos de demostración a la secuencia de entrega. Si ya se transmiten datos a este destino, los datos de demostración se envían junto con los registros de origen.

Enviar datos de demostración

Paso 2

Deje de enviar datos de demostración a la secuencia de entrega una vez concluida la prueba para dejar de incurrir en gastos de uso.

Dejar de enviar datos de demostración

6. LABORATORIO KINESIS FIREHOSE

Paso 18. En la pestaña Configuración vemos la Configuración de destino, donde toda la información de valores de bolsa que se está generando de forma automática, la está recuperando el Firehose. Hay un retardo de 1 min o de 1 MB de información antes de pasarlo a S3.

Configuración de destino [Información](#)
Especifique la configuración de destino de la secuencia de entrega.

Destino de Amazon S3

Bucket de S3
edula.kinesis.bolsa.1 [🔗](#)

Prefijo de salida de error del bucket de S3
errores/

Particionamiento dinámico [Información](#)

Particionamiento dinámico
No está habilitado

Desagregación de varios registros
No está habilitado

Tipo de desagregación de varios registros
-

Delimitador de desagregación
-

Nuevo delimitador de línea
No está habilitado

Análisis sintáctico en línea para JSON
No está habilitado

Prefijo de bucket de S3
datos/

6. LABORATORIO KINESIS FIREHOSE







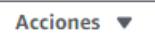
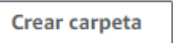

Paso 19. Si vamos al bucket, vemos que no esta vacío y ya tiene información. Hay que esperar al menos 1 min para que toda la información automática que vamos metiendo desde el Test pase al bucket (no es en tiempo real)..


edulara.kinesis1.bolsa.1 [Info](#)

Objetos | Propiedades | Permisos | Métricas | Administración | Puntos de acceso

Objetos (1)

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

  Copiar URI de S3  Copiar URL  Descargar  Abrir  Eliminar  Acciones  Crear carpeta  Cargar

<input type="checkbox"/>	Nombre	Tipo	Última modificación	Tamaño	Clase de almacenamiento
<input type="checkbox"/>	 datos/	Carpeta	-	-	-

6. LABORATORIO KINESIS FIREHOSE

Paso 20. En la pestaña de Configuración vemos el tag que hemos puesto. No hemos puesto encriptación y hemos habilitado los logs S3

Etiquetas (1) [Información](#)

Administrar etiquetas

Puede agregar etiquetas para organizar los recursos de AWS, realizar un seguimiento de los costos y controlar el acceso.

< 1 >

Clave

desarrollo.kinesis

Cifrado del servidor (SSE) [Información](#)

Puede utilizar AWS Key Management Service (KMS) para crear y administrar las claves maestras de los clientes (CMK) y para controlar el uso del cifrado en una amplia gama de servicios de AWS para las aplicaciones.

Editar

Cifrado del servidor

No está habilitado

Registros de errores de destino [Información](#)

Elija la opción Habilitado si desea que Kinesis Data Firehose registre los errores de entrega de registros en CloudWatch Logs.

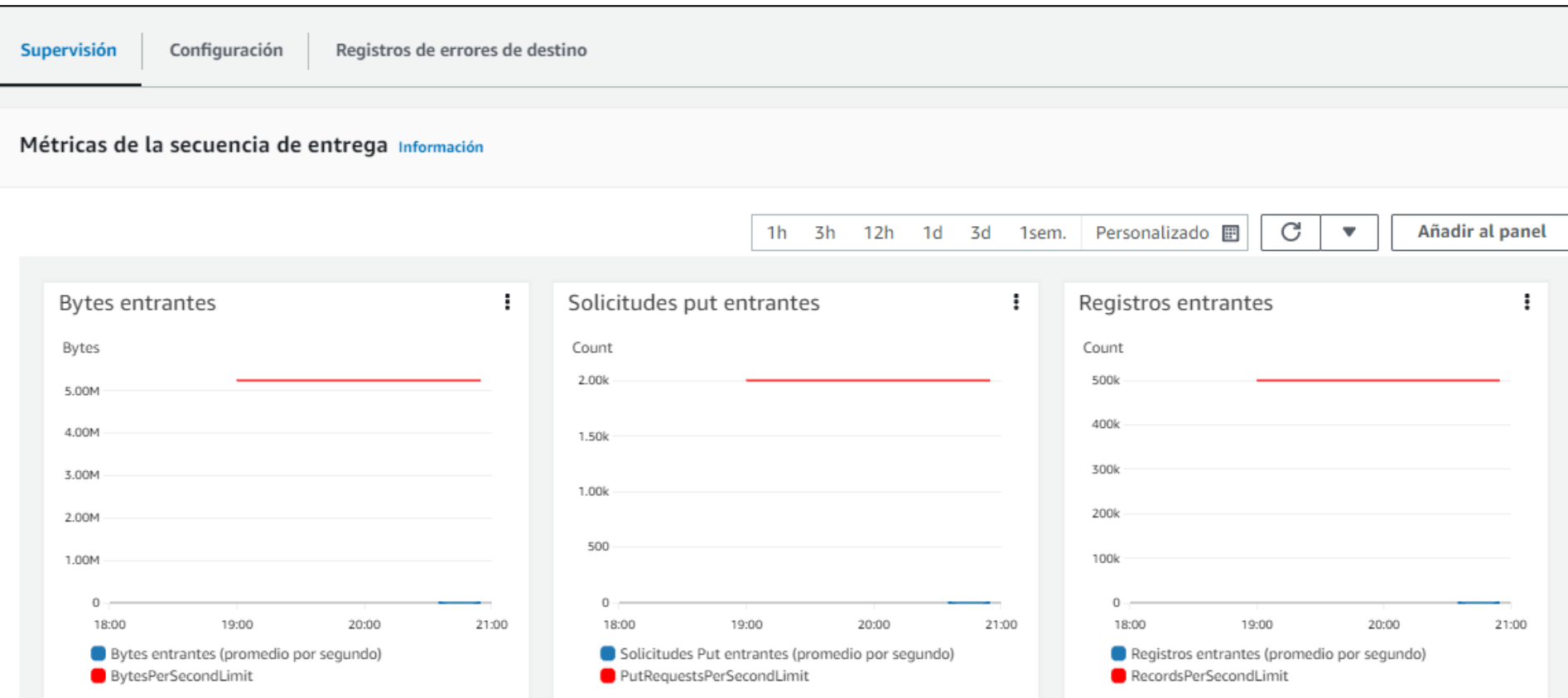
Editar

Registro de errores de Amazon CloudWatch

Habilitado


6. LABORATORIO KINESIS FIREHOSE


Paso 21. En la pestaña de supervisión podemos ver la monitorización. Lo que va leyendo y lo que va grabando.






6. LABORATORIO KINESIS FIREHOSE

Paso 22. Si volvemos al bucket y actualizamos, vemos que ha grabado la carpeta datos, creada para meter los datos. La de errores, como no hay ningún error, todavía no ha creado. Si pulsamos en la carpeta de datos vemos que va almacenando por fecha: → 2023/ → 04/ → 25 → 20 → Va almacenando ficheros.

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	 2023/	Carpeta	-	-	-

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	 04/	Carpeta	-	-	-

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	 25/	Carpeta	-	-	-

<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	 20/	Carpeta	-	-	-
<input type="checkbox"/>	 21/	Carpeta	-	-	-





<input type="checkbox"/>	Nombre ▲	Tipo ▼	Última modificación ▼	Tamaño ▼	Clase de almacenamiento ▼
<input type="checkbox"/>	 kinesis1-1-2023-04-25-20-35-48-e9014d33-036e-4ec7-8031-9b52abeefdae	-	25 Apr 2023 10:36:52 PM CEST	791.0 B	Estándar
<input type="checkbox"/>	 kinesis1-1-2023-04-25-20-36-55-1478c89c-e287-4784-b2f6-0b68ac9067be	-	25 Apr 2023 10:37:57 PM CEST	786.0 B	Estándar
<input type="checkbox"/>	 kinesis1-1-2023-04-25-20-38-00-c5cbd1ee-d535-4090-a611-c3a7470b82b4	-	25 Apr 2023 10:39:01 PM CEST	707.0 B	Estándar
<input type="checkbox"/>	 kinesis1-1-2023-04-25-20-39-01-a8687711-f4a0-463d-b5d6-e1162da0236e	-	25 Apr 2023 10:40:03 PM CEST	791.0 B	Estándar

6. LABORATORIO KINESIS FIREHOSE

Paso 23. Si pulsamos en el primer fichero, podemos descargarlo y abrirlo con un editor de texto normal. Vemos los datos que ha ido almacenando: ticket, tecnología, precio, etc

Amazon S3 > Buckets > edulara.kinesis1.bolsa.1 > datos/ > 2023/ > 04/ > 25/ > 20/ > kinesis1-1-2023-04-25-20-35-48-e9014d33-036e-4ec7-8031-9b52abeefdae

kinesis1-1-2023-04-25-20-35-48-e9014d33-036e-4ec7-8031-9b52abeefdae [Info](#)

 Copiar URI de S3  Descargar  Abrir  Acciones de objetos ▼

```
{ "CHANGE":1.32,"PRICE":67.12,"TICKER_SYMBOL":"ASD","SECTOR":"FINANCIAL" }
{ "CHANGE":0.34,"PRICE":19.72,"TICKER_SYMBOL":"PLM","SECTOR":"FINANCIAL" }
{ "CHANGE":0.35,"PRICE":25.13,"TICKER_SYMBOL":"ABC","SECTOR":"RETAIL" }
{ "CHANGE":0.05,"PRICE":5.14,"TICKER_SYMBOL":"NGC","SECTOR":"HEALTHCARE" }
{ "CHANGE":-1.03,"PRICE":29.25,"TICKER_SYMBOL":"PPL","SECTOR":"HEALTHCARE" }
{ "CHANGE":0.38,"PRICE":110.91,"TICKER_SYMBOL":"WSB","SECTOR":"FINANCIAL" }
{ "CHANGE":0.64,"PRICE":218.68,"TICKER_SYMBOL":"QWE","SECTOR":"TECHNOLOGY" }
{ "CHANGE":0.8,"PRICE":16.07,"TICKER_SYMBOL":"VVS","SECTOR":"ENERGY" }
{ "CHANGE":-0.74,"PRICE":75.06,"TICKER_SYMBOL":"SLW","SECTOR":"ENERGY" }
{ "CHANGE":0.51,"PRICE":219.42,"TICKER_SYMBOL":"QXZ","SECTOR":"FINANCIAL" }
{ "CHANGE":-0.13,"PRICE":14.97,"TICKER_SYMBOL":"JKL","SECTOR":"TECHNOLOGY" }
```

Hemos generado datos de forma automática que se envían al kinesis Firehose, este los recoge, en este caso no los transforma y simplemente los guarda en S3 en diferentes ficheros. Mientras no le demos a parar seguirá enviando,

7. AWS SQS

- ❑ El servicio de SQS es un servicio de colas de mensajes
- ❑ Donde se permite enviar, almacenar y recibir mensajes
- ❑ Permite aislar aplicaciones que envían mensajes (productores) de las aplicaciones que los consumen (consumidores)
- ❑ Sirve de buffer de información
- ❑ Sirve para mover mensajes entre distintos componentes o servicios de AWS, sin necesidad de que estén disponibles todo el tiempo: Un servicio puede enviar un mensaje a una cola, dejarlo ahí y al día siguiente otro servicio lo puede ir a buscar

7. AWS SQS

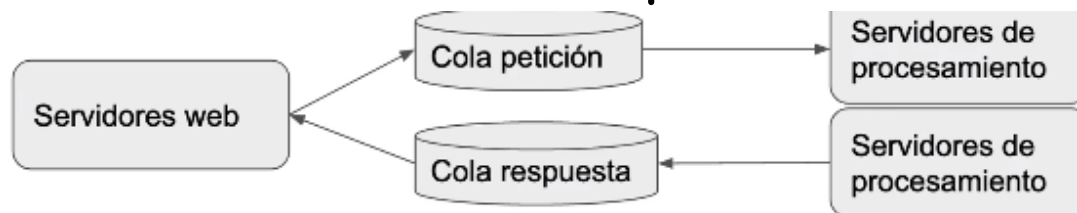
- ❑ El tamaño máximo mensajes: 256 Kbytes
- ❑ Si se quiere ampliar el tamaño de los mensajes se puede utilizar el SQS Extended Client: Procesa mensajes mayores de 256 Kbytes
- ❑ Pueden haber múltiples productores y consumidores simultáneos en la misma cola
- ❑ Ofrece 2 tipos de colas:
 - ❑ Colas estándar. Mas optimizadas, mayor rendimiento pero no garantizan el orden de los mensajes
 - ❑ Colas FIFO. Primero entra, primero sale

7. AWS SQS

- ❑ Los mensajes pueden ser retenidos durante 14 días: Una aplicación o servicio puede guardar un mensaje en una cola. Si nadie lee de esa cola, tiene 14 días para estar almacenado
- ❑ Existen mecanismos de Long Poll: Una aplicación va a buscar un mensaje y si no lo encuentra, se espera 20 segundos para volver a mirar en la cola si hay mensajes. No está continuamente mirando la cola.

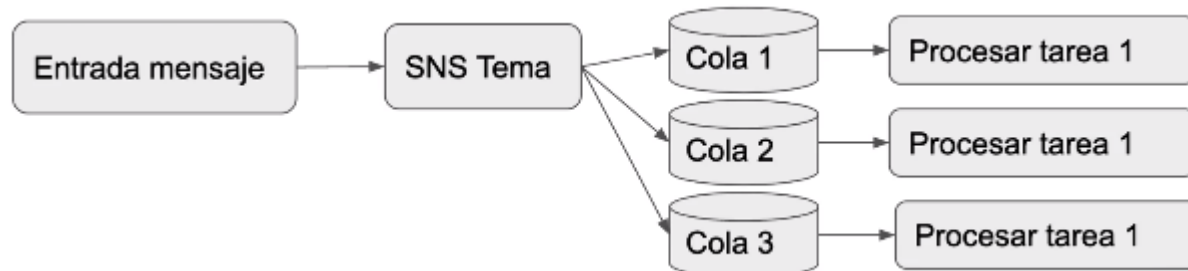
7. AWS SQS

- ❑ Permite crear sistemas asíncronos, las peticiones y las respuestas van por caminos separados e independientes.
- ❑ Un servidor web manda un mensaje de petición a una cola de petición. Una serie de servicios procesan estos mensajes de petición y generan unas respuestas que se envían por una cola de respuesta diferente. Y el servidor web recuperara las respuestas de esa cola.
- ❑ Cuando manda la petición no espera la respuesta, hay un proceso que envía y otro que recibe la respuesta
- ❑ A nivel de web, el servicio parece síncrono, pero por detrás el backend es de tipo asíncrono



7. AWS SQS

- ❑ La gestión de prioridades en los mensajes se realiza con distintas colas según prioridad. A una cola de más prioridad podemos darle más recursos procesamiento y a otra de menos prioridad, le daremos menos recursos.
- ❑ Permite realizar múltiples tareas sobre una única entrada, utilizando el fanout o despliegue. Permite bifurcar o duplicar un mensaje en diferentes colas para hacer distintos procesamiento de forma independiente.



7. AWS SQS

- ❑ Por ejemplo si tenemos una entrada que es una imagen y queremos realizar varias tareas sobre esa imagen, publicamos ese mensaje en un tema de SNS para que reparta ese mensaje por distintas colas
- ❑ En la 1º cola → un proceso va a coger esa imagen y va a hacer una miniatura de la imagen
- ❑ En la 2º cola → otro proceso va a hacer un reconocimiento facial con caras que tenemos en nuestra base de datos.
- ❑ En la 3º cola → habrá un servicio que lea la imagen para hacer una búsqueda de sus metadatos

7. AWS SQS

Casos de uso para el AWS SQS

- ☐ Procesamiento ordenado de datos, por ejemplo el proceso de reserva, en servicios web, etc
- ☐ En procesamiento de imágenes
- ☐ En sistemas asíncronas donde queremos separar la entrada de la salida.

8. AWS IOT INTERNET DE LAS COSAS

- ❑ Son dispositivos del mundo real, como pueden ser cámaras, sensores, que envían información que podemos procesar y almacenar en servicios de Amazon AWS: DynamoDB, Kinesis, funciones lambda, servicios de Big Data y servicios de Machine Learning.
- ❑ Es una plataforma en la nube completamente administrada, donde los dispositivos pueden intercambiar información con otros servicios de AWS y también con otros dispositivos de IoT.
- ❑ Los mensajes pueden ser enrutados a servicios de AWS y también pueden ser enrutados a otros dispositivos IoT.

8. AWS IOT INTERNET DE LAS COSAS

- ❑ Los servicios de AWS IoT puede soportar billones de dispositivos y trillones de mensajes
- ❑ Los dispositivos de IoT generan gran cantidad de datos que podemos analizar, procesar y almacenar en tiempo real, en los servicios de AWS para IoT y Big Data, sin importar el volumen que ocupa, gracias al auto-escalado
- ❑ Los servicios AWS IoT permiten leer de los dispositivos de IoT, y grabar la información en servicios de AWS como ElasticSearch, Firehose, Data stream, DinamoDB,
- ❑ AWS IoT Device SDK es un conjunto de librerías que utilizan los dispositivos de IoT para conectarse, autenticarse e intercambiar mensajes con nuestros servicios de AWS IoT

8. AWS IOT INTERNET DE LAS COSAS

- Autenticación y autorización:
 - En el servicio AWS IoT hay que crear un certificado que debe ser instalado en cada dispositivo IoT transmisor de información que queramos almacenar en nuestros sistemas AWS
 - Políticas IAM, para gestionar los permisos a los que puede acceder cada dispositivo o grupo dispositivos.
 - Amazon Cognito para autenticación desde aplicación móviles y así poder utilizar los servicios AWS según las políticas que definamos en IAM
 - Amazon Cognito Identity permite darle una entidad propia a cada dispositivo.

8. AWS IOT INTERNET DE LAS COSAS

- ❑ También podemos loginarnos en AWS a través de otros proveedores como Google, Facebook, Twitter, etc.
- ❑ Usuarios → aplicación móvil
 - Login en Google
 - Token/credenciales desde Cognito
 - Acceso al servicio AWS IoT
 - Acceso a otros servicios AWS según políticas IAM
- ❑ Un usuario con dispositivo móvil, puede logarse en su cuenta de Google o Facebook y con el token de acceso que le ha dado Google o Facebook, puede llamar a Amazon Cognito que le da credenciales para acceder a los servicios de AWS IoT, y desde ahí conectarse al resto de servicios de Amazon AWS.

8. AWS IOT INTERNET DE LAS COSAS

- ❑ Pasarela de dispositivos que podemos tener en AWS IoT
 - ❑ Esta pasarela mantiene sesiones para todos los dispositivos IoT conectados a nuestra red AWS
 - ❑ Les permite comunicaciones seguras 1 a 1 y 1 a N, utilizando para ello el SNS o suscripción de temas.
 - ❑ Protocolos afectados: MQTT (Message Queue Telemetric Transport) típico de los dispositivos IoT, WebSockets o el HTTP
 - ❑ Permite auto escalado automático para soportar esta gran cantidad de dispositivos.

8. AWS IOT INTERNET DE LAS COSAS

- ❑ Registro de dispositivos
 - ❑ Sirve para registrar cosas o dispositivos IoT
 - ❑ Una cosa es un dispositivo físico que puede ser un sensor o una cámara. Pero también podría ser una identidad lógica, como una aplicación móvil que me permite controlar la temperatura de mi casa.
 - ❑ Lo que llamamos internet de las cosas, estas cosas pueden ser o dispositivos físicos o identidades lógicas

8. AWS IOT INTERNET DE LAS COSAS

- ❑ Sombra del dispositivo
 - ❑ Es un documento JSON que permite almacenar el estado actual de cada cosa, dispositivo físico o lógico.
 - ❑ Existe una sombra de dispositivo por cada cosa conectada a IoT,
 - ❑ Permite actuar como canal de envío de mensajes, tanto desde dispositivo hacia AWS y viceversa

8. AWS IOT INTERNET DE LAS COSAS

- ❑ Máquinas de reglas.
 - ❑ Permiten a las cosas interactuar con el servicio AWS IoT y con el resto de servicios AWS
 - ❑ Permiten transformar y enrutar los mensajes hacia varios servicios de AWS .
 - ❑ Los mensajes se pueden transformar basándose en la sintaxis de SQL
 - ❑ Basándose en las reglas podemos activar acciones y provocar, por ejemplo, el envío de un mensaje a alguno de nuestros servicios AWS

9. DATA PIPELINE

- ❑ Es un servicio para procesar y mover datos entre servicios AWS e incluso con fuentes de datos locales
- ❑ Permite crear un trabajo ETL de extracción, transformación y carga para automatizar el procesamiento y el movimiento de los datos
- ❑ Permite copiar datos entre diferentes regiones AWS (por ejemplo para copias de seguridad)
- ❑ Permite planificar trabajos, ejecutarlos a mano, o con un temporizador para su ejecución cada x tiempo
- ❑ Internamente un Data Pipeline se ejecuta en un EC2. Al finalizar el Pipeline se eliminan los recursos utilizados
- ❑ Se puede ejecutar en local con el paquete Task Runner

9. DATA PIPELINE

- ❑ Un Pipeline contiene:
 - ❑ Nodos de datos: especifican el destino final de los datos, que puede ser DynamoDB, Mysql, RedShift, S3
 - ❑ Actividades: acciones que se van a realizar dentro del Pipeline: actividades de EMR, Hive, etc
 - ❑ Precondiciones: comprobaciones previas a la ejecución tanto en los datos como en las actividades a realizar: Verificar que existe una tabla, que tiene datos,
 - ❑ Planificadores: Especifican cuándo se ejecuta nuestro pipeline, su frecuencia, etc. Algunas las tareas podríamos ejecutarlas mediante funciones lambda: Llega un fichero a S3, se puede ejecutar una función lambda que mueva ese fichero a otro servicio AWS

10. FORMAS INTRODUCIR DATOS EN AWS

- ❑ **Conexión directa:** permite conectar recursos de tus servidores locales con recursos en AWS. Pretende mejorar el rendimiento en la transferencia de datos.
- ❑ **Snowball:** Permite exportar los datos locales a estos dispositivos físicos llamados snowball para posteriormente enviarlos a AWS y que se carguen en nuestra cuenta, para utilizar mas adelante
- ❑ Indicado para mover grandes cantidades de datos (petaBytes), Como pasarlos a través de una línea de datos de Internet puede tardar muchos días debido al gran volumen, interesa exportarlos a estos dispositivos físicos y enviarlos a Amazon para que los cargue

10. FORMAS INTRODUCIR DATOS EN AWS

- ❑ **Pasarela de almacenamiento:** Permite conectar nuestras aplicaciones locales para que puedan utilizar servicios de Amazon AWS como S3, Glacier, etc.. Tiene una transferencia de datos optimizada incluso con una caché local para los datos más utilizados.
- ❑ **Transferencia acelerada de S3:** transferencia rápida o acelerada de datos en S3. Sirve para copiar datos entre diferentes regiones. Utiliza por tanto el servicio de Cloud Front y los datos son enrutados por localizaciones próximas hasta S3 para ser más eficientes.

10. FORMAS INTRODUCIR DATOS EN AWS

- ❑ **Servicio de Migración de Bases de datos o DMS**
Database Migration Service, para traspasar datos entre diferentes bases de datos.
 - ❑ Soporta la mayoría de bases de datos comerciales.
 - ❑ La migración puede ser homogénea, entre las mismas bases de datos (de Oracle a Oracle o de MySql a MySql) o heterogéneas (de Oracle a MySql, etc)
 - ❑ La migración puede hacerse todo a la vez o mediante la replicación continua.
 - ❑ Recomendado para tamaños bases de datos < 5 TB
 - ❑ Si tiene más de 5 terabytes, se debería usar una combinación entre el dispositivo físico Snowball con este sistema de migración de bases de datos DMS