
BIG DATA

INCIDENCIAS Y EVENTOS EN AWS RECOGIDA DE DATOS CON KINESIS FIREHOSE

EDUARD LARA

INDICE

1. Amazon GuardDuty
2. Amazon Inspector
3. Kinesis data analytics
4. Kinesis data Firehose
5. Kinesis data streams
6. Kinesis video streams
7. Laboratorio Kinesis Firehose

1. AMAZON GUARDDUTY

- Es un servicio de detección de amenazas inteligentes que monitoriza de forma constante para detectar comportamientos maliciosos o no autorizados.
- Envía los hallazgos detallados de seguridad para su visibilidad y resolución.
- Los casos de uso son:
 - Detener la actividad no autorizada.
 - Análisis y monitoreo continuos
 - Simplificar análisis forense.

1. AMAZON GUARDDUTY

Detener la actividad no autorizada.

- Se evita el uso de credenciales vulnerables, el acceso inusual a datos de S3, las llamadas desde direcciones IP malintencionadas

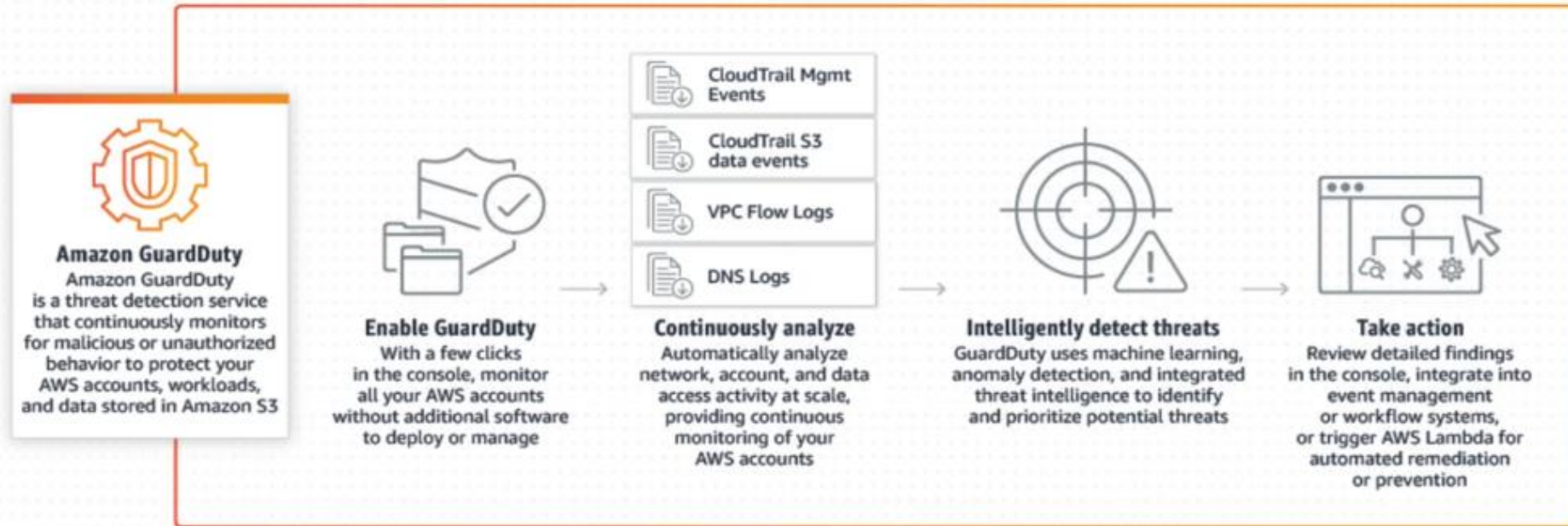
Análisis y monitoreo continuos

- Permite obtener información acerca de eventos de seguridad con hallazgos que proporcionan contexto, metadatos y detalles sobre los recursos afectados.

Simplificar análisis forense

- Sirve para determinar rápidamente la causa raíz de las actividades sospechosas con el uso de la integración en la consola de Amazon GuardDuty con Amazon Detected.

1. AMAZON GUARDDUTY



2. AMAZON INSPECTOR

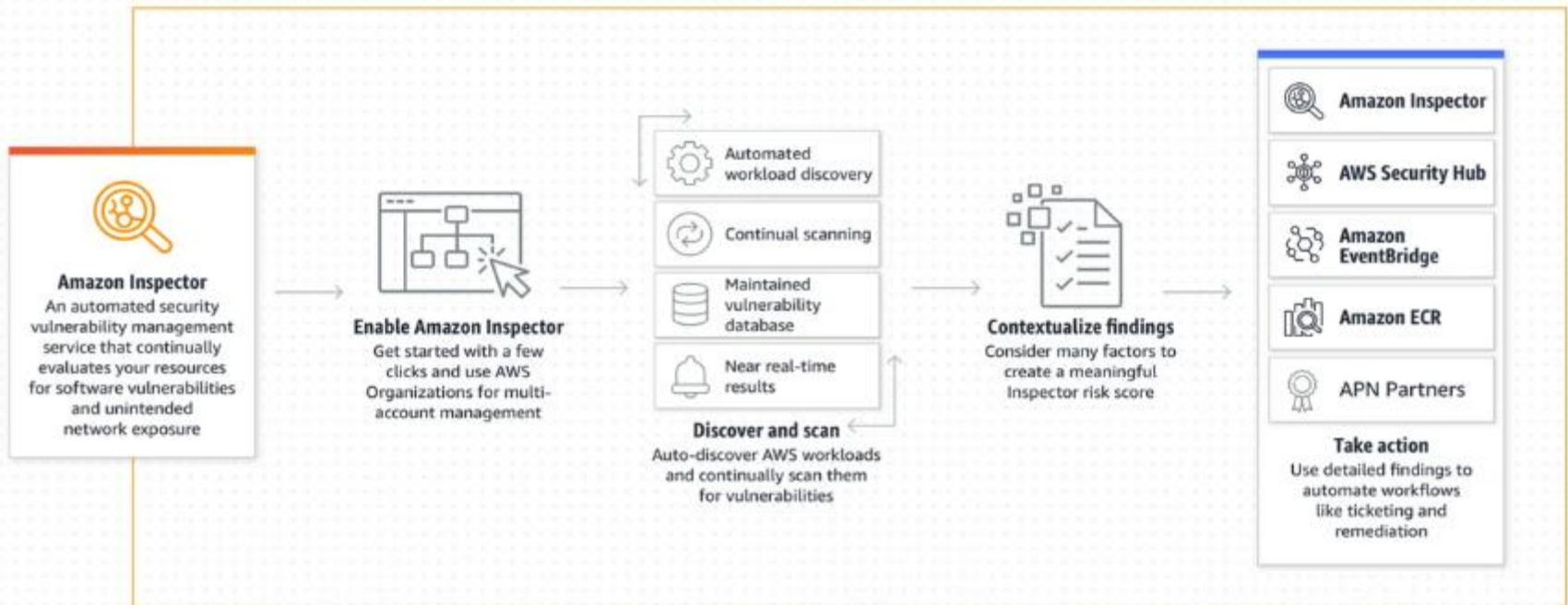
- Amazon Inspector es un servicio de administración automatizada de vulnerabilidades que analiza las cargas de trabajo de AWS en busca de vulnerabilidades de software y exposición involuntaria a la red.
- Es un servicio que examina tus aplicaciones en busca de vulnerabilidades y produce un informe con los temas de seguridad encontrados.
- Casos de uso:
 - Detección rápida de vulnerabilidades,
 - Priorización en la corrección de parches,
 - Cumplimentación de los requisitos de conformidad
 - Identificación temprana de vulnerabilidades de día cero.

2. AMAZON INSPECTOR

Beneficios de Amazon Inspector

- Identifica problemas de seguridad.
- Es conducido por API para facilitar su implementación.
- Reduce el tiempo a mostrarse los problemas de seguridad antes de que aparezcan.
- Reduce el tiempo medio de resolución de vulnerabilidades
- Ayuda a cumplir los estándares. Permite definir tus propios estándares de seguridad que desees cumplir.

2. AMAZON INSPECTOR



3. KINESIS DATA ANALYTICS

- ❑ Este servicio permite transformar y analizar datos que recibimos en 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
- ❑ Podemos crear aplicaciones de consulta de datos mediante SQL: Java, Scala y Python

3. 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)

3. KINESIS DATA ANALYTICS

- ❑ Diagrama de utilización del 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.



4. KINESIS DATA FIREHOSE

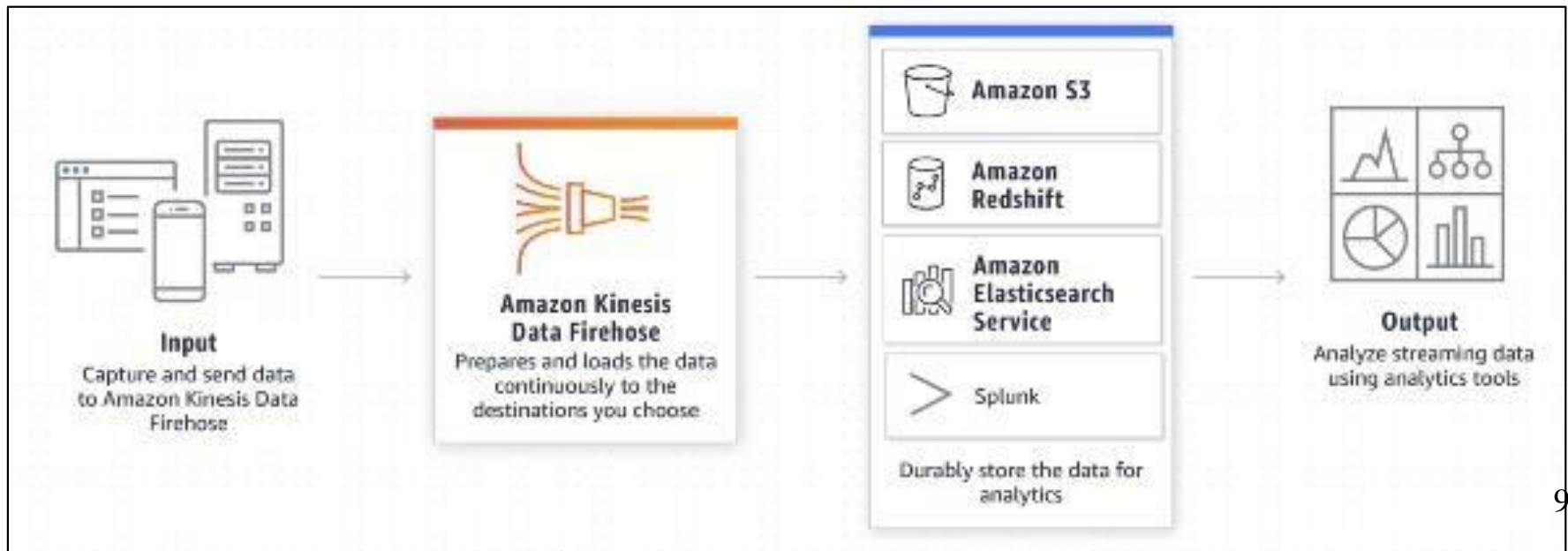
- ❖ Amazon Kinesis Data Firehose sirve para preparar y cargar transmisiones de datos generados en tiempo real en almacenes de datos y servicios de análisis
- ❖ Servicio completamente administrado para enviar datos de streaming en tiempo real a varios destinos admitidos, de manera sencilla y fiable
- ❖ Los productores de datos son configurados para que envíen datos a Kinesis Data Firehose y éste los entrega automáticamente al destino especificado
- ❖ Kinesis Data Firehose puede transformar los datos antes de entregarlos, mediante la función AWS lambda, si fuera necesario

4. KINESIS DATA FIREHOSE

- ❖ 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
- ❖ No se paga por servidores, se paga por volumen de datos transmitido
- ❖ Casos de uso:
 - Análisis de dispositivos IoT
 - Análisis de secuencia de clics en sitios web
 - Análisis de registros
 - Monitorización de la seguridad

4. KINESIS DATA FIREHOSE

- ❖ En el diagrama tenemos diferentes entradas de datos por medio de input y mediante Kinesis data Firehose podemos transformarlos con AWS Lambda o enviarlos a distintos almacenes de datos: S3, RedShift, ElasticSearch
- ❖ Podemos hacer un análisis a partir de estos datos almacenados en estos dispositivos de almacenamiento



4. 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.

4. 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.

5. KINESIS DATA STREAMS

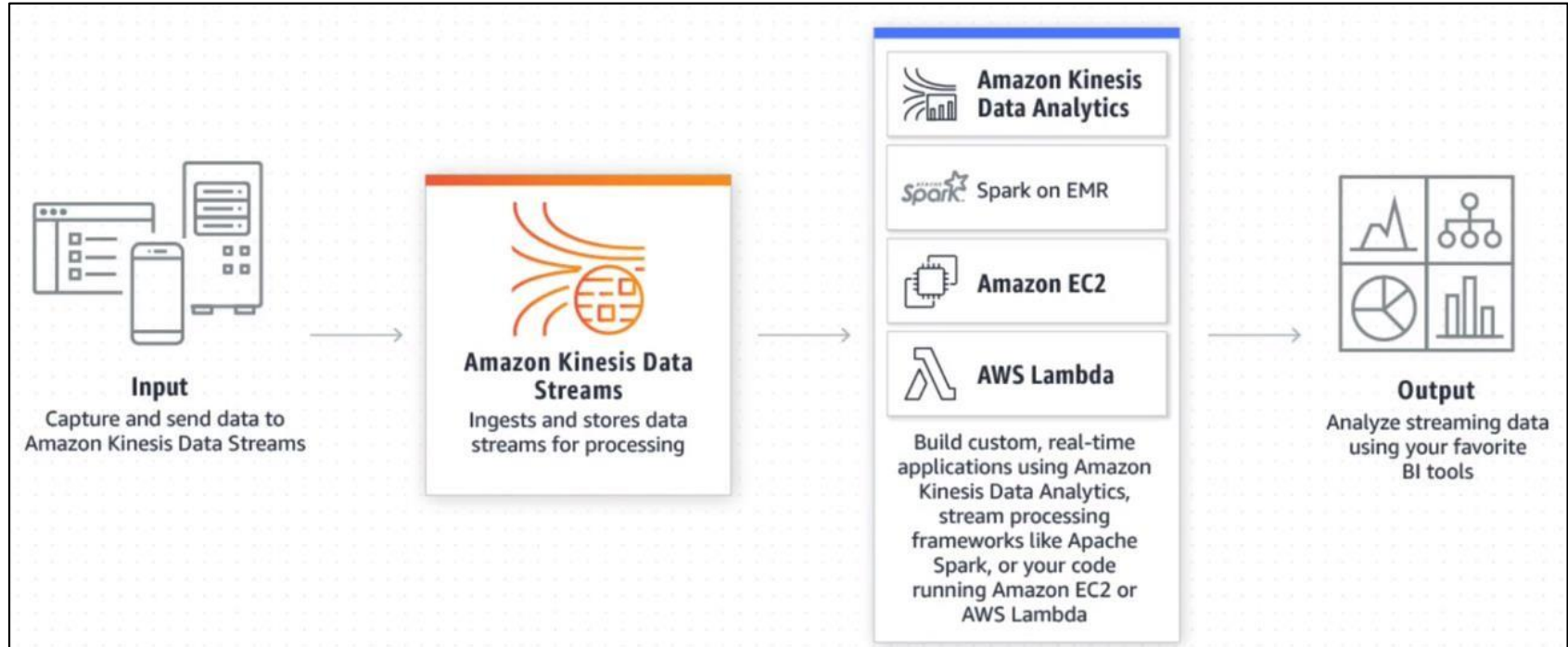
- Permite recopilar datos de streaming a gran escala para el análisis en tiempo real
- Es capaz de registrar de manera continua Gigabytes de datos por segundo de miles de orígenes:
 - Transmisiones de clics de sitios web
 - Transmisiones de eventos de bases de datos
 - Transacciones financieras
 - Fuentes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

5. 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

5. 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..



6. KINESIS VIDEO STREAMS

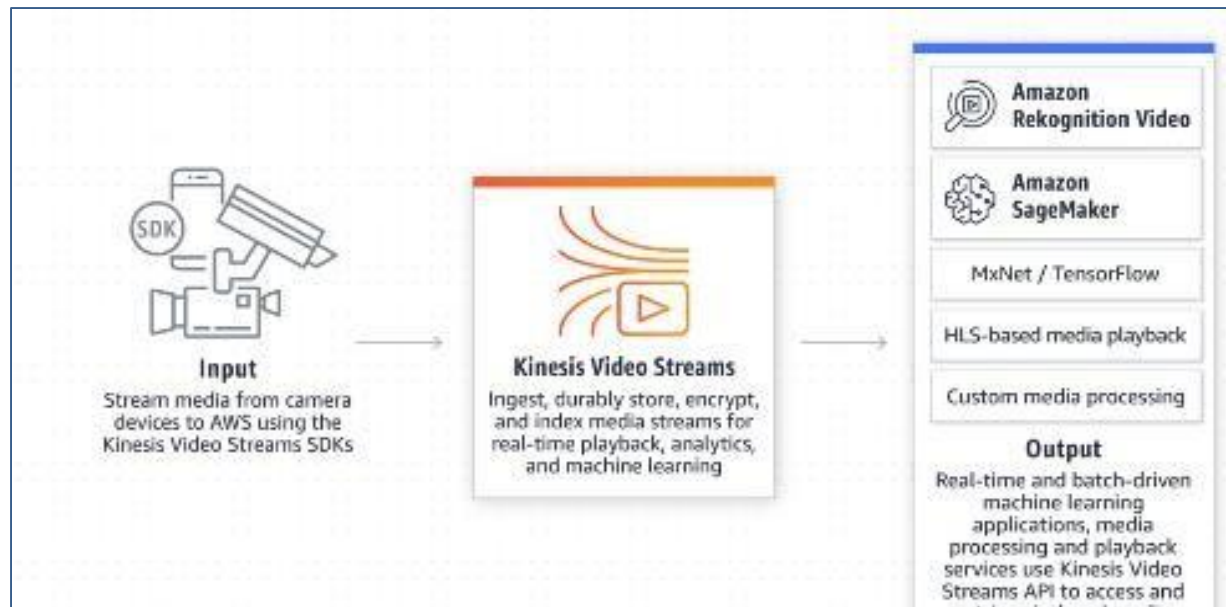
- Kinesis Video Streams permite la transmisión segura de videos desde dispositivos conectados a AWS para tareas de análisis, aprendizaje automático, reproducción y otros procesos.
- Permite transmitir videos desde muchos dispositivos que estén conectados a AWS como pueden ser: teléfonos móviles, cámaras 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

6. KINESIS VIDEO STREAMS

- Permite 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)
 - Automatización industrial

6. 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
- ❖ Ese video puede ser procesado y almacenado y utilizado en las herramientas de Amazon Rekognition Video, SageMaker



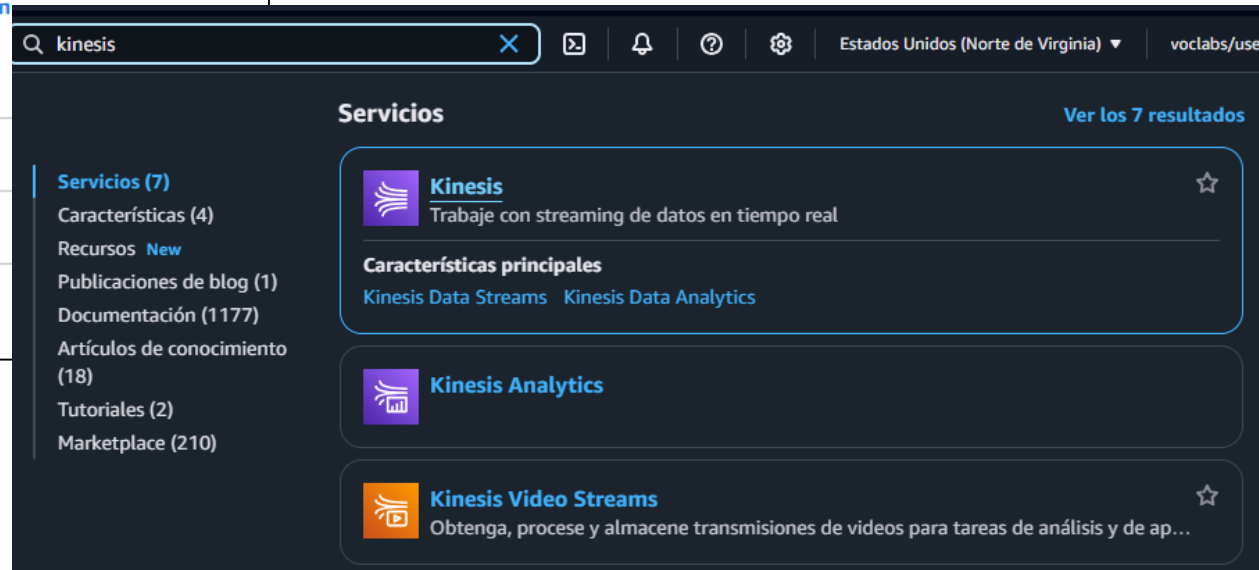
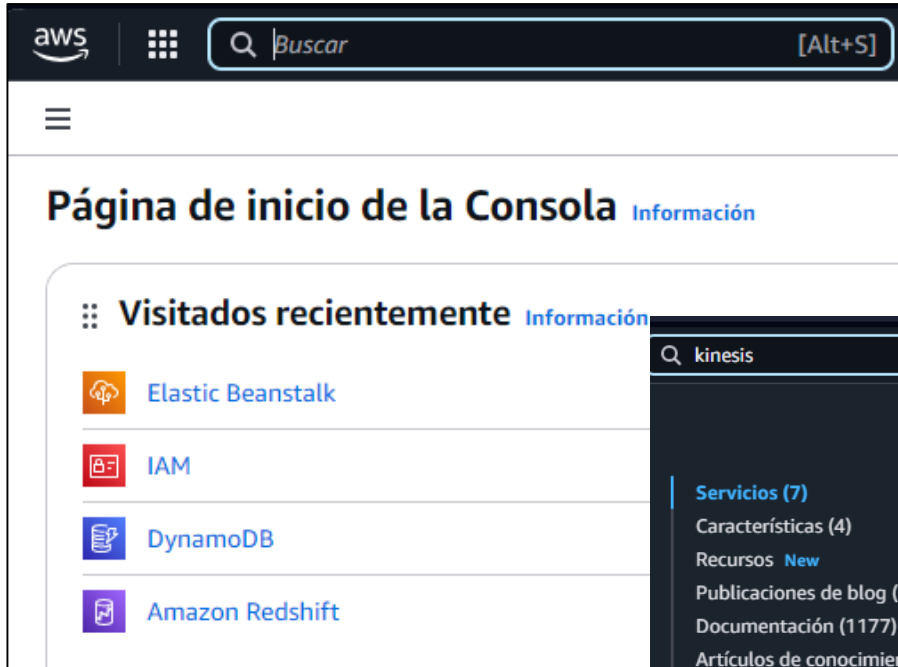
7. LABORATORIO KINESIS FIREHOSE

Paso 0. Haremos un laboratorio de Kinesis donde utilizaremos Kinesis Data Firehose para recoger datos de bolsa que generaremos de forma automática y lo guardaremos en S3 para posteriormente realizar un análisis con SQL de esos datos.



7. LABORATORIO KINESIS FIREHOSE

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



7. LABORATORIO KINESIS FIREHOSE

Paso 2. Podemos elegir los tres tipos de Kinesis que hay:

- Secuencia de datos
- Data firehose
- Data Analytics

Elegimos Data Firehose y hacemos click en crear secuencia de Firehose. Esto va a recoger datos de distintas fuentes, en este caso de una fuente automática, y lo va a almacenar en S3.

Servicios de Amazon Kinesis

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

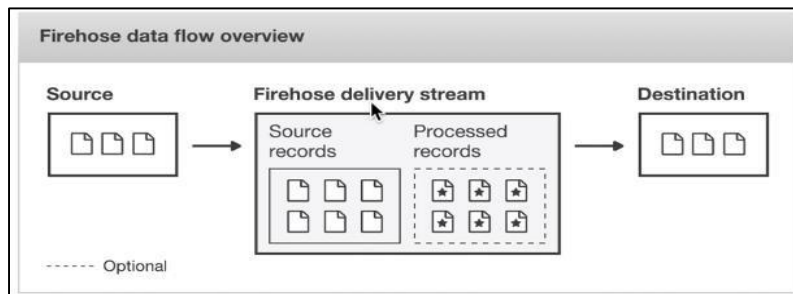
Introducción

- ☐ Kinesis Data Streams
Recopilar datos de streaming con una secuencia de datos.
- ☒ Amazon Data Firehose - *nuevo*
Antiguamente Kinesis Data Firehose
Procese y entregue datos de streaming con una secuencia de entrega de datos.
- ☐ Managed Apache Flink
Formerly Kinesis Data Analytics
Analice datos de streaming con aplicaciones de análisis de datos.

[Crear secuencia de Firehose](#)

7. LABORATORIO KINESIS FIREHOSE

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



Crear la secuencia de Firehose [Información](#)

► Amazon Data Firehose: cómo funciona

Elija un origen y un destino

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

Origen | [Información](#)

Elija un origen ▼

Destino | [Información](#)

Elija un destino ▼

[Cancelar](#)

[Crear la secuencia de Firehose](#)

7. LABORATORIO KINESIS FIREHOSE

Paso 4. Como origen de datos, marcamos la opción Direct Put, que permite crear una secuencia de Amazon Data Firehose en las que las aplicaciones productoras escriban directamente. **Vamos a hacerlo de forma automática con los datos de ejemplo que viene en la creación del Firehose.** Podríamos haber elegido entre Kinesis Data Streams, Amazon MSK, Mysql y PostgreSQL

Amazon Kinesis Data Streams
Elija esta opción si desea utilizar Kinesis Data Streams como origen de datos para la secuencia de Firehose.
Amazon MSK
Elija esta opción si desea utilizar Amazon MSK como origen de datos para la secuencia de Firehose.
Base de datos MySQL
Elija esta opción si desea utilizar bases de datos MySQL como origen de datos para la secuencia de Firehose.
Base de datos PostgreSQL
Elija esta opción si desea utilizar bases de datos PostgreSQL como origen de datos para la secuencia de Firehose.
Direct PUT
Elija esta opción para crear una secuencia de Amazon Data Firehose en la que las aplicaciones productoras escriban directamente.
Direct PUT

Crear la secuencia de Firehose [Información](#)

► Amazon Data Firehose: cómo funciona

Elija un origen y un destino

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

Origen [Información](#)

Direct PUT

Destino [Información](#)

Elija un destino

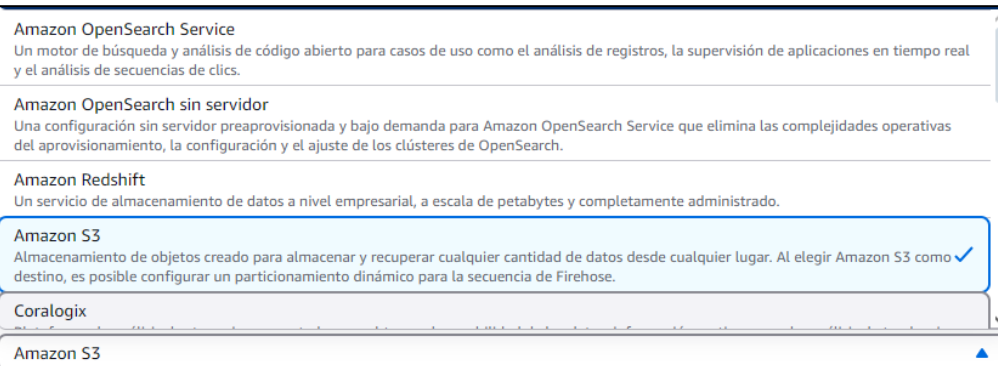
[Cancelar](#)

[Crear la secuencia de Firehose](#)

7. LABORATORIO KINESIS FIREHOSE

Paso 5. Como destino, los datos que recoge Firehose los va a guardar en un destino que puede ser: S3, RedShift, OpenSearch, Coralogix, Elastic, MongoDB Cloud, etc.

Seleccionamos S3, es decir, todo lo que recoja Firehose lo va a guardar directamente en S3 (en nuestro caso sin transformar)



Amazon OpenSearch Service
Un motor de búsqueda y análisis de código abierto para casos de uso como el análisis de registros, la supervisión de aplicaciones en tiempo real y el análisis de secuencias de clics.

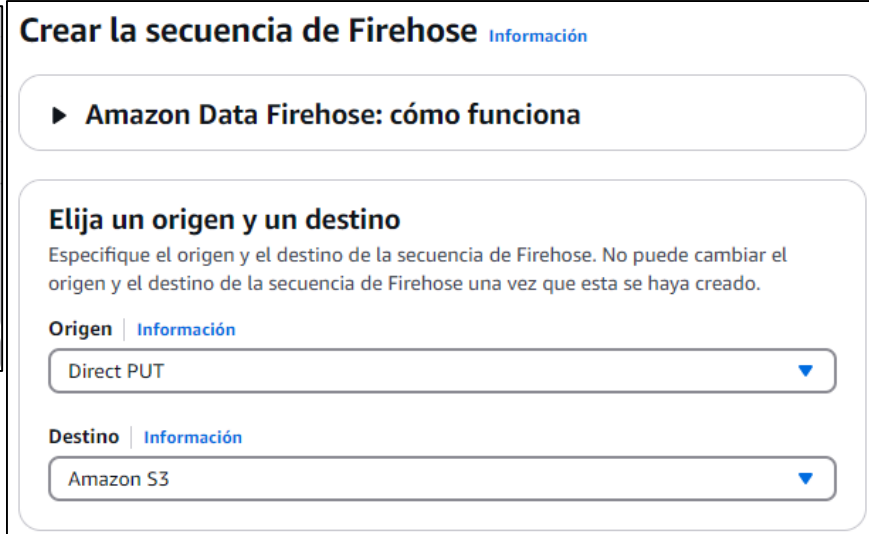
Amazon OpenSearch sin servidor
Una configuración sin servidor preaprovisionada y bajo demanda para Amazon OpenSearch Service que elimina las complejidades operativas del aprovisionamiento, la configuración y el ajuste de los clústeres de OpenSearch.

Amazon Redshift
Un servicio de almacenamiento de datos a nivel empresarial, a escala de petabytes y completamente administrado.

Amazon S3
Almacenamiento de objetos creado para almacenar y recuperar cualquier cantidad de datos desde cualquier lugar. Al elegir Amazon S3 como destino, es posible configurar un particionamiento dinámico para la secuencia de Firehose. ✓

Coralogix

Amazon S3



Crear la secuencia de Firehose [Información](#)

► **Amazon Data Firehose: cómo funciona**

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

Origen | [Información](#)
Direct PUT ▼

Destino | [Información](#)
Amazon S3 ▼

7. LABORATORIO KINESIS FIREHOSE

Paso 6. Ponemos el nombre de kinesis1 a la secuencia de Firehose. No habilitaremos la transformación de registros con Lambda ni la conversión de formatos de los registros.

Nombre de la secuencia de Firehose

Nombre de la secuencia de Firehose

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

Transformar y convertir registros: *opcional*

Configure Amazon Data Firehose para transformar y convertir los datos de registro.

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

Amazon 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 de datos para los datos de origen entrantes antes de su entrega al destino especificado.

☐ Activar 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. Amazon 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

Descomprima los registros origen de registros de Amazon CloudWatch: *nuevo* | [Información](#)

Cuando esta característica está activada, Amazon Data Firehose descomprime los registros de Amazon CloudWatch y los entrega al destino. Para esta funcionalidad habrá cargos adicionales. Para obtener más información, consulte [Precios de Amazon Data Firehose](#). Al habilitar la descompresión, tiene la opción de activar la extracción de mensajes para entregar solo el contenido de los campos de mensajes. [Más información](#).

☐ Activar la descompresión

7. LABORATORIO KINESIS FIREHOSE

Paso 7. 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](#)
Amazon 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 de datos para los datos de origen entrantes antes de su entrega al destino especificado.

☒ **Activar la transformación de datos**

Función de AWS Lambda

Versión o alias
 ▼

Navegar

Crear función

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 período durante el cual Amazon 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: 0 segundos, máximo: 900 segundos.


7. LABORATORIO KINESIS FIREHOSE

Paso 8. Dentro de la configuración de destino, tenemos que elegir un bucket. En este caso crearemos un bucket nuevo.

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

Bucket de S3

[Examinar](#)

[Crear](#) 

Formato: s3://bucket

7. LABORATORIO KINESIS FIREHOSE

Paso 9. El nombre de bucket debe ser único para todo el mundo, por ejemplo **nombre_alumno + kinesis1.bolsa**, así no coincidirá con ningún otro bucket.

Crear bucket [Información](#)

Los buckets son contenedores de datos almacenados en S3.

Configuración general

Región de AWS
EE.UU. Este (Norte de Virginia) us-east-1

Tipo de bucket [Información](#)

☒ **Uso general**
Recomendado para la mayoría de los casos de uso y patrones de acceso. Los buckets de uso general son del tipo de bucket de S3 original. Permiten una combinación de clases de almacenamiento que almacenan objetos de forma redundante en múltiples zonas de disponibilidad.

☐ **Directorio**
Recomendado para casos de uso de baja latencia. Estos buckets utilizan únicamente la clase de almacenamiento S3 Express One Zone, que proporciona un procesamiento más rápido de los datos dentro de una única zona de disponibilidad.

Nombre del bucket [Información](#)

edulara.kinesis1.bolsa

El nombre del bucket debe ser único dentro del espacio de nombres global y seguir las reglas de nomenclatura del bucket. [Consulte las reglas para la asignación de nombres de buckets](#) [↗](#)

Copiar la configuración del bucket existente: *opcional*
Solo se copia la configuración del bucket en los siguientes ajustes.

[Elegir el bucket](#)

Formato: s3://bucket/prefijo

Le ponemos bolsa porque los datos van a ser de bolsa.

7. LABORATORIO KINESIS FIREHOSE

Paso 10. Dejamos las opciones por defecto y le damos a Crear Bucket.

Tipo de cifrado | Información

☒ Cifrado del servidor con claves administradas de Amazon S3 (SSE-S3)
☐ Cifrado del servidor con claves de AWS Key Management Service (SSE-KMS)
☐ Cifrado de doble capa del servidor con claves de AWS Key Management Service (DSSE-KMS)
Proteja sus objetos con dos capas de cifrado independientes. Para obtener más información sobre los precios, consulte [DSSE-KMS pricing](#) (Precios de DSSE-KMS) en la pestaña **Storage** (Almacenamiento) de la [página de precios de Amazon S3](#). [↗](#)

Clave de bucket

El uso de una clave de bucket de S3 para SSE-KMS reduce los costos de cifrado al reducir las llamadas a AWS KMS. Las claves de bucket de S3 no son compatibles con DSSE-KMS. [Más información](#) [↗](#)

☐ Desactivar
☒ Habilitar

► Configuración avanzada

i

Después de crear el bucket, puede cargar archivos y carpetas, y configurar ajustes adicionales en él.

Cancelar

Crear bucket

7. LABORATORIO KINESIS FIREHOSE

Paso 11. Nos indica que el bucket se creo correctamente

Amazon S3 > Buckets

✓ El bucket "edulara.kinesis1.bolsa" se creó correctamente

Para cargar archivos y carpetas, o para configurar ajustes adicionales del bucket, elija [Ver detalles](#).

Ver detalles

► Instantánea de la cuenta: actualizada cada 24 horas

Todas las regiones de AWS

Ver panel de Storage Lens

Storage Lens permite visualizar el uso del almacenamiento y las tendencias de la actividad. Las métricas no incluyen los buckets de directorio. [Más información](#)

Buckets de uso general

Buckets de directorio

Buckets de uso general (2)

Información

Todas las regiones de AWS

Copiar ARN

Vaciar

Eliminar

Crear bucket

Los buckets son contenedores de datos almacenados en S3.

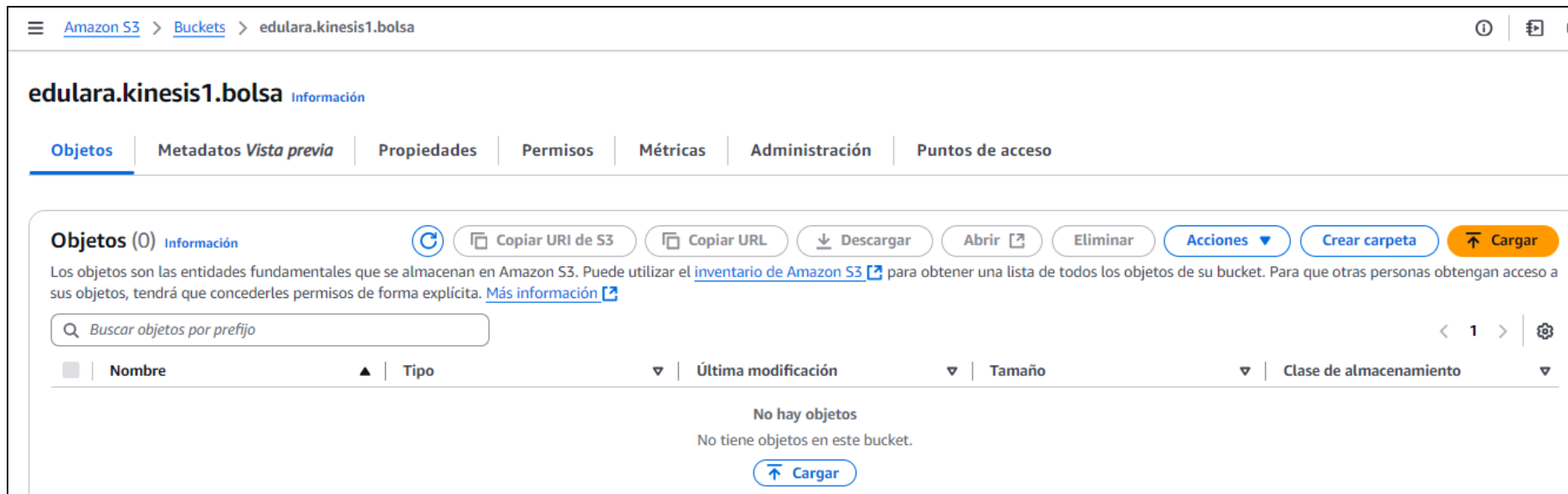
Buscar buckets por nombre

< 1 > ⚙

Nombre	Región de AWS	Analizador de acceso de IAM	Fecha de creación
<input type="radio"/> edulara.kinesis1.bolsa	EE.UU. Este (Norte de Virginia) us-east-1	Ver analizador para us-east-1	11 Jan 2025 8:44:50 PM CET
<input type="radio"/> elasticbeanstalk-us-east-1-068131188799	EE.UU. Este (Norte de Virginia) us-east-1	Ver analizador para us-east-1	24 Dec 2024 9:18:07 AM CET

7. LABORATORIO KINESIS FIREHOSE

Paso 12. Si entramos en el enlace del bucket, vemos que está vacío, porque todavía no ha recibido ningún dato. Tan pronto Firehose reciba datos, los recogerá y los enviará a este bucket.



7. LABORATORIO KINESIS FIREHOSE

Paso 13. A continuación asignamos el bucket creado anteriormente al Firehose kinesis1 que estamos creando

Elija un bucket en Amazon S3

Buckets (1/2)

Nombre

Región

☒ edulara.kinesis1.bolsa

EE.UU. Este (Norte de Virginia)

☐ elasticbeanstalk-us-east-1-068131188799

EE.UU. Este (Norte de Virginia)

Cancelar

Elegir

Configuración de destino [Información](#)

Especifique la configuración de destino de la secuencia de Firehose.

Bucket de S3

[Examinar](#)[Crear](#)

Formato: s3://bucket

Nuevo delimitador de línea

Puede configurar la secuencia de Firehose de manera que agregue un nuevo delimitador de línea entre los registros de los objetos que se entregan a Amazon S3.

- ☒ No está habilitado
- ☐ Habilitado

Particionamiento dinámico [Información](#)

El particionamiento dinámico permite crear conjuntos de datos específicos mediante la partición de los datos de streaming de S3 en función de las claves de partición. Puede particionar los datos de origen mediante el análisis sintáctico en línea o la función de AWS Lambda especificada. Puede habilitar el particionamiento dinámico únicamente al crear un nuevo flujo de Firehose. No puede habilitar el particionamiento dinámico para un flujo de Firehose que ya existe. Al habilitar el particionamiento dinámico se incurre en costos adicionales según los GiB de datos particionados. Para obtener más información, consulte [Precios de Amazon Data Firehose](#).

- ☒ No está habilitado
- ☐ Habilitado

7. LABORATORIO KINESIS FIREHOSE

Paso 14. 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

Prefijo de bucket de S3: *opcional*

De forma predeterminada, Amazon Data Firehose añade el prefijo "AAAA/MM/dd/HH" (en UTC) a los datos que predeterminado si se especifica un prefijo personalizado que incluya expresiones que se evalúan en la versión ejecutable.

datos/

Puede repetir las mismas claves en el prefijo del bucket de S3. Máximo de caracteres del prefijo de bucket de S3

Prefijo de salida de error de bucket S3: *opcional*

Puede especificar un prefijo de salida de error del bucket de S3 que se utilizará en las condiciones de error. Este Data Firehose las evalúa en la versión ejecutable.

errores/

Zona horaria del prefijo de salida de error del bucket de S3 y S3 | [Información](#)

Elegir la zona horaria que desea utilizar para la fecha y la hora en los prefijos de S3

UTC ▼

7. LABORATORIO KINESIS FIREHOSE

Paso 15. En Sugerencias del almacenamiento en buffer de S3, elegimos las condiciones del buffer, donde se prealmacenarán los datos. En el tamaño del buffer pondremos 1MB y en el intervalo del buffer 60 seg. Los datos que recoge Firehose se almacenan en este buffer durante 1 min o hasta que llega un máximo de 1 MB y entonces se vuelcan en S3.

▼ Sugerencias de almacenamiento en búfer, compresión, extensión de archivos y cifrado

Los campos que aparecen a continuación se han completado previamente con los valores predeterminados recomendados para S3. Los precios pueden variar en función de los costos del almacenamiento y de las solicitudes.

Sugerencias del almacenamiento en búfer de S3

Amazon Data Firehose almacena en búfer los registros entrantes antes de entregarlos al bucket de S3. La entrega de registros se activa una vez que se alcanza el valor de cualquiera de las sugerencias de almacenamiento en búfer especificadas.

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: 0 segundos, máximo: 900 segundos. Valor recomendado: 300 segundos.

Firehose es casi en tiempo real, porque hay un retardo de 60 seg.

7. LABORATORIO KINESIS FIREHOSE

Paso 16. S3 lo podemos comprimir o encriptar. Dejamos las opciones que están por defecto

Compresión para registros de datos

Amazon Data Firehose puede comprimir los registros antes de entregarlos al bucket de S3.

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

Formato de extensión de archivo: *opcional*

Ingrese el formato de extensión completo. Si desea agregar "extensión de archivo gzip" para los archivos "json", ingrese ".json.gz".

Ingresa una extensión de archivo (por ejemplo, .json, .zip, .json.gz)

La extensión de archivo debe comenzar con un punto y solo puede contener caracteres permitidos: 0-9a-z!-_*(). La extensión de archivo



Si especifica una extensión de archivo, anulará la extensión de archivo predeterminada que Amazon Data Firehose

Cifrado de los registros de datos

El registro se cifra en el bucket de S3 mediante una clave de AWS Key Management Service (KMS).

- ☒ Utilice la configuración de cifrado del bucket de S3
El cifrado predeterminado del bucket de S3 es el cifrado del servidor con claves administradas de Amazon S3 (SSE-S3). Consulte [S3: Protección de los datos mediante cifrado del servidor](#) para ver otras opciones de cifrado.
- ☐ Utilice SSE-KMS
Cifrado del servidor con claves de AWS Key Management Service (KMS).

7. LABORATORIO KINESIS FIREHOSE

Paso 17. En configuración avanzada dejamos habilitado el Registro de errores de logging de Amazon Cloudwatch tal y como ya estaba

▼ **Configuración avanzada**

Cifrado del servidor no está habilitada; registro de errores habilitado; rol de IAM KinesisFirehoseServiceRole-kinesis1-us-

Cifrado del servidor | [Información](#)

Puede utilizar AWS Key Management Service (KMS) para crear y administrar las claves y para controlar el uso del cifrado en una amplia gama de

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

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

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

☐ No está habilitado

☒ Habilitado

7. LABORATORIO KINESIS FIREHOSE

Paso 18. 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](#)

Amazon Data Firehose utiliza este rol de IAM para todos los permisos que la secuencia de Firehose necesita. Si desea especificar diferentes roles pa

☐ Cree o actualice el rol de IAM **KinesisFirehoseServiceRole-kinesis1-us-east-1-1736625540316**
Crea un nuevo rol o actualiza uno existente y agrega a este la política requerida, a la vez que permite que Amazon Data Firehose lo asuma.

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

i Si elige un rol de IAM existente para Amazon Data Firehose, asegúrese de que el rol de IAM elegido tenga una política de información, consulte [Control de acceso con Amazon Data Firehose](#).

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

LabRole

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

7. LABORATORIO KINESIS FIREHOSE

Paso 19. En Etiquetas Opcional podemos poner un nombre y un valor. Esto sirve después sirve para luego facturar. Por ejemplo si tenemos un departamento de desarrollo, podemos añadir un tag de clave **desarrollo.kinesis** y de valor **bolsa**. Esto nos servirá para facturar el coste de 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	
<input type="text" value="desarrollo.kinesis"/>	<input type="text" value="bolsa"/>	<input type="button" value="Eliminar"/>

Puede agregar hasta 49 etiquetas más.

7. LABORATORIO KINESIS FIREHOSE

Paso 20. Hacemos click en el botón Crear secuencia de Firehose. 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.

kinesis1 Información

✓ kinesis1 se creó correctamente.


kinesis1 Información [Eliminar la secuencia de Firehose](#)

Detalles de la secuencia de Firehose


Estado ✓ Activo	Destino Amazon S3	Transformación de datos No está habilitado	Tiempo de creación 11 de enero de 2025, 22:41 CET
Origen Direct PUT	ARN  arn:aws:firehose:us-east-1:068131188799:deliverystream/kinesis1	Particionamiento dinámico No está habilitado	Estado de los registros de errores ✓ 0 Registros de errores de destino

7. LABORATORIO KINESIS FIREHOSE



Paso 21. Una vez creado, entramos en kinesis1, y vemos su información: está activo, la dirección del bucket de S3 donde se van a guardar los datos, creación, permisos, etc

Secuencias de Firehose (1) Información  [Eliminar](#) [Crear la secuencia de Firehose](#)

Puede crear una secuencia de Firehose para configurar un origen, un destino y una transformación opcional para la entrega de datos de secuencia.




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

	Nombre	Estado	Tiempo de cr...	Origen	Transformaci...	Tipo de destino	URL de destino
<input type="radio"/>	kinesis1	 Activo	11 de enero de ...	Direct PUT	No está habilitada	Amazon S3	edulara.kinesis1...

 kinesis1 se creó correctamente. 

kinesis1 Información [Eliminar la secuencia de Firehose](#)

Detalles de la secuencia de Firehose

Estado  Activo	Destino Amazon S3	Transformación de datos No está habilitado	Tiempo de creación 11 de enero de 2025, 22:41 CET
Origen Direct PUT	ARN  <code>arn:aws:firehose:us-east-1:068131188799:deliverystream/kinesis1</code>	Particionamiento dinámico No está habilitado	Estado de los registros de errores  0 Registros de errores de destino

7. LABORATORIO KINESIS FIREHOSE

Paso 22. Si pulsamos en Kinesis1 hay un apartado que es **Probar con datos de demostración**. Esto ejecuta un script que envía datos de prueba a Firehose. Podemos construir nuestro propio programa o utilizar este test demo para enviar datos a Firehose. 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 Firehose. Se aplican los cargos estándar de Amazon Data Firehose.

Esta prueba ejecuta un script en el navegador para introducir datos de demostración en la secuencia de Amazon Data Firehose, que realiza el env

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

7. LABORATORIO KINESIS FIREHOSE

Paso 23. Le damos a enviar datos y empieza a enviarse datos de bolsa, generados de forma automática y aleatoria, a nuestro Firehose. 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 dejamos 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

7. LABORATORIO KINESIS FIREHOSE

Paso 24. En la pestaña Configuración, Configuración de destino, vemos la url del bucket de S3, donde se guarda toda la información de valores de bolsa generada de forma aleatoria en el Firehose. Hay un retardo de 1 min o de 1MB de información antes de pasarlo a S3.

Configuración de destino Información Editar

Especifique la configuración de destino de la secuencia de Firehose.

Destino de Amazon S3

Bucket de S3 edulara.kinesis1.bolsa	Prefijo de salida de error del bucket de S3 errores/	Zona horaria del prefijo de salida de error del bucket de S3 y S3 UTC
Nuevo delimitador de línea No está habilitado		

Particionamiento dinámico Información

Particionamiento dinámico No está habilitado	Desagregación de varios registros No está habilitado	Tipo de desagregación de varios registros -
Delimitador de desagregación -		Análisis sintáctico en línea para JSON No está habilitado

Prefijo de bucket de S3

datos/

7. LABORATORIO KINESIS FIREHOSE

Paso 25. Más abajo vemos el tag que hemos puesto. No hemos puesto encriptación y hemos habilitado los logs S3

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

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

Cifrado del servidor
No está habilitado

Registros de errores de destino [Información](#)

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

Registro de errores de Amazon CloudWatch
Habilitado

Acceso al servicio [Información](#)

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

Rol de IAM
[LabRole](#)

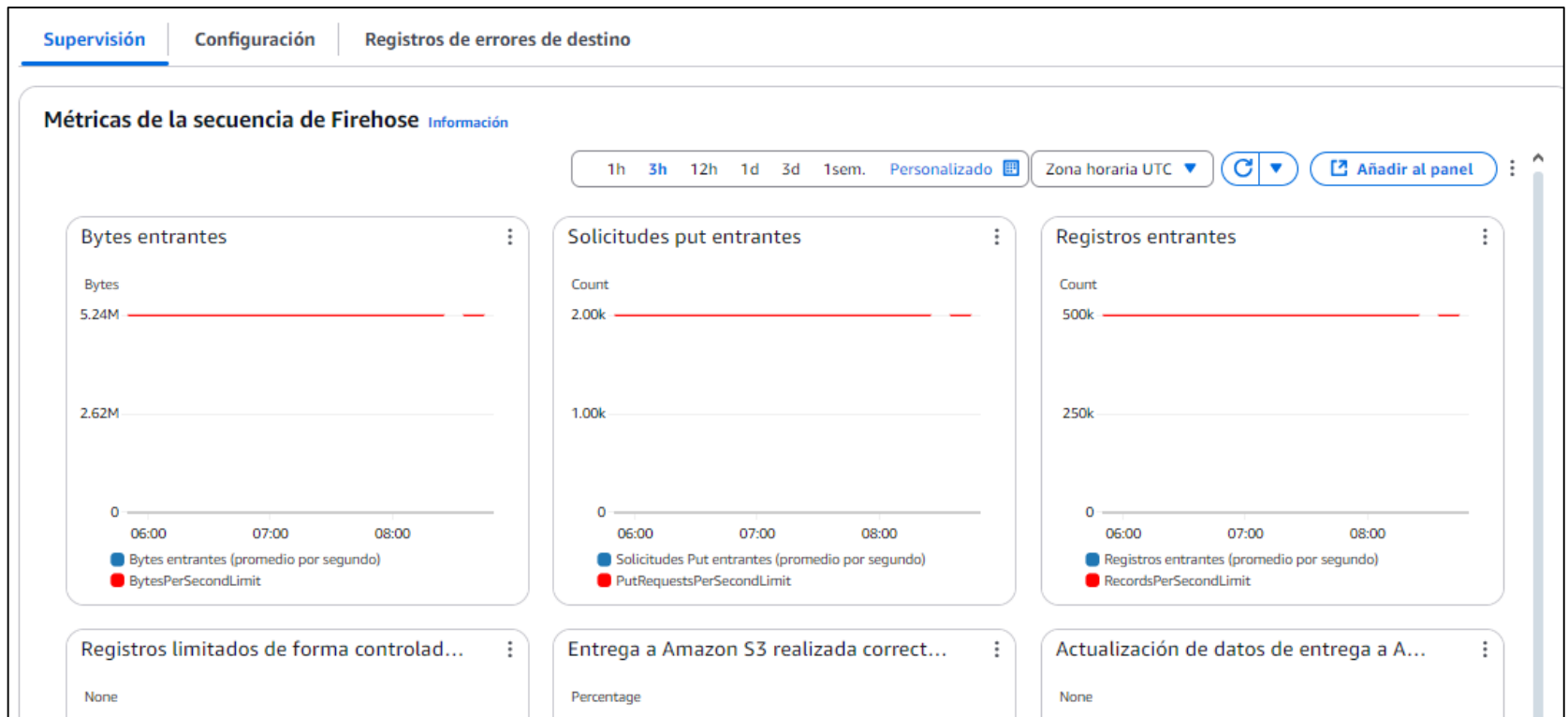
Etiquetas (1) [Información](#)

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

Clave	Valor
desarrollo.kinesis	bolsa

7. LABORATORIO KINESIS FIREHOSE

Paso 26. En la pestaña de Supervision podemos ver la monitorización. Lo que va leyendo y lo que va grabando.









7. LABORATORIO KINESIS FIREHOSE



Paso 27. Si vamos al bucket y actualizamos, vemos que ya ha grabado la carpeta datos. Esta es la carpeta creada para que se fueran metiendo los datos. 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 Información

Objetos | Metadatos Vista previa | Propiedades | Permisos | Métricas | Administración | Puntos de acceso

Objetos (1) Información

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

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](#) 

 Buscar objetos por prefijo

☐

Nombre

▲

Tipo

▼

Última modificación

▼


Tamaño

▼

Clase de almacenamiento

▼

☐

 [datos/](#)

Carpeta

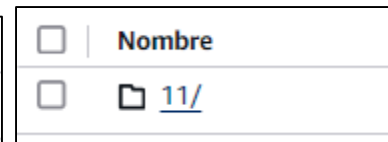
-

-

-

7. LABORATORIO KINESIS FIREHOSE

Paso 28. Si entramos en la carpeta datos, vemos que ya tiene información. La de errores, todavía no se ha creado (no hay errores). Si vamos pulsando en las carpetas vemos que va almacenando los ficheros por fecha:
→ 2025/ (año) → 01/ (mes) → 11/ (dia) → 22/ (hora)



22/ Copiar URI de S3

Objetos | Propiedades

Objetos (5) Información Copiar URI de S3 Copiar URL Descargar Abrir Eliminar Acciones Crear carpeta Cargar

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](#)

Buscar objetos por prefijo

<input type="checkbox"/>	Nombre	Tipo	Última modificación	Tamaño	Clase de almacenamiento
<input type="checkbox"/>	kinesis1-1-2025-01-11-22-01-05-ca72a757-0a19-4241-a059-cd3d1290585d	-	11 Jan 2025 11:02:06 PM CET	646.0 B	Estándar
<input type="checkbox"/>	kinesis1-1-2025-01-11-22-02-03-6718f273-e756-41f7-b300-351616da9faf	-	11 Jan 2025 11:03:07 PM CET	716.0 B	Estándar
<input type="checkbox"/>	kinesis1-1-2025-01-11-22-03-05-13136cc6-c5bd-4aff-8e5f-2c30e97a53c2	-	11 Jan 2025 11:04:10 PM CET	645.0 B	Estándar
<input type="checkbox"/>	kinesis1-1-2025-01-11-22-04-21-2e5014cc-bfcd-4c24-b1f0-3c73daa733a8	-	11 Jan 2025 11:05:26 PM CET	716.0 B	Estándar

7. LABORATORIO KINESIS FIREHOSE

Paso 29. Si pulsamos en el primer fichero, vemos que lo podemos descargar y abrir con un editor de texto normal.

kinesis1-1-2025-01-11-22-01-05-ca72a757-0a19-4241-a059-cd3d1290585d Información

Copiar URI de S3

Descargar

Abrir

Acciones de objetos ▼

Propiedades

Permisos

Versiones

Información general sobre el objeto

Propietario
awslabsc0w4002021t1647574653

Región de AWS
EE.UU. Este (Norte de Virginia) us-east-1

Última modificación
11 Jan 2025 11:02:06 PM CET

Tamaño
646.0 B

Tipo

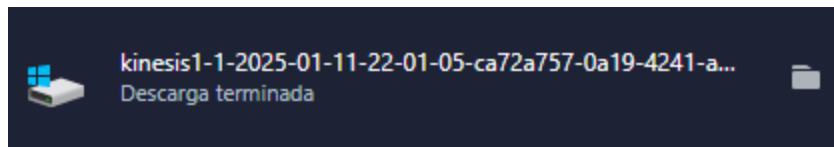
Clave
datos/2025/01/11/22/kinesis1-1-2025-01-11-22-01-05-ca72a757-0a19-4241-a059-cd3d1290585d

URI DE S3
s3://edulara.kinesis1.bolsa/datos/2025/01/11/22/kinesis1-1-2025-01-11-22-01-05-ca72a757-0a19-4241-a059-cd3d1290585d

Nombre de recurso de Amazon (ARN)
arn:aws:s3:::edulara.kinesis1.bolsa/datos/2025/01/11/22/kinesis1-1-2025-01-11-22-01-05-ca72a757-0a19-4241-a059-cd3d1290585d

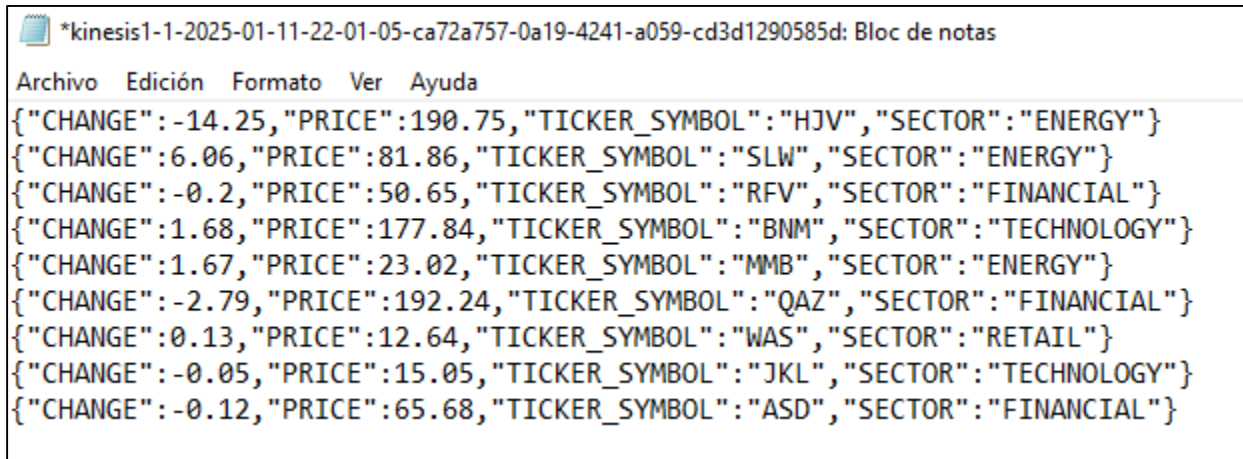
Etiqueta de entidad (Etag)
fb65ae97229b25f77afd93b2b0632922

URL del objeto
https://s3.us-east-1.amazonaws.com/edulara.kinesis1.bolsa/datos/2025/01/11/22/kinesis1-1-2025-01-11-22-01-05-ca72a757-0a19-4241-a059-cd3d1290585d



7. LABORATORIO KINESIS FIREHOSE

Paso 30. Vemos los datos que ha ido almacenando: ticket, tecnología, precio, etc.



```
*kinesis1-1-2025-01-11-22-01-05-ca72a757-0a19-4241-a059-cd3d1290585d: Bloc de notas
Archivo  Edición  Formato  Ver  Ayuda
[{"CHANGE": -14.25, "PRICE": 190.75, "TICKER_SYMBOL": "HJV", "SECTOR": "ENERGY"}]
[{"CHANGE": 6.06, "PRICE": 81.86, "TICKER_SYMBOL": "SLW", "SECTOR": "ENERGY"}]
[{"CHANGE": -0.2, "PRICE": 50.65, "TICKER_SYMBOL": "RFV", "SECTOR": "FINANCIAL"}]
[{"CHANGE": 1.68, "PRICE": 177.84, "TICKER_SYMBOL": "BNM", "SECTOR": "TECHNOLOGY"}]
[{"CHANGE": 1.67, "PRICE": 23.02, "TICKER_SYMBOL": "MMB", "SECTOR": "ENERGY"}]
[{"CHANGE": -2.79, "PRICE": 192.24, "TICKER_SYMBOL": "QAZ", "SECTOR": "FINANCIAL"}]
[{"CHANGE": 0.13, "PRICE": 12.64, "TICKER_SYMBOL": "WAS", "SECTOR": "RETAIL"}]
[{"CHANGE": -0.05, "PRICE": 15.05, "TICKER_SYMBOL": "JKL", "SECTOR": "TECHNOLOGY"}]
[{"CHANGE": -0.12, "PRICE": 65.68, "TICKER_SYMBOL": "ASD", "SECTOR": "FINANCIAL"}]
```

Hemos generado datos de forma automática que se envían al kinesis Firehose. Éste los recoge, en este caso no los transforma y simplemente los guarda en S3 en diferentes ficheros.

7. LABORATORIO KINESIS FIREHOSE

Paso 31. Si volvemos a Kinesis1, mientras no le demos a parar seguirá enviando datos de bolsa al Firehose, y éste a su vez a S3 donde se guardan en ficheros

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

Ingiera datos simulados para probar la configuración de la secuencia de Firehose.

Esta prueba ejecuta un script en el navegador para introducir datos de demostración.

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

Paso 1

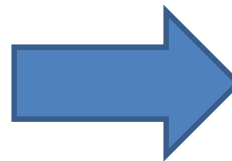
Comience a enviar datos de demostración a la secuencia de Firehose.

Enviar datos de demostración

Paso 2

Deje de enviar datos de demostración a la secuencia de Firehose una vez que haya terminado.

Dejar de enviar datos de demostración



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

Ingiera datos simulados para probar la configuración de la secuencia de Firehose.

Esta prueba ejecuta un script en el navegador para introducir datos de demostración.

```
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 Firehose. Si desea detener la prueba, haga clic en el botón de abajo.

Comience a enviar datos de demostración

Paso 2

Deje de enviar datos de demostración a la secuencia de Firehose una vez que haya terminado.

Dejar de enviar datos de demostración