

Universidad EAFIT

Escuela de Ingeniería

Departamento de Informática y Sistemas

Proyecto 3

Presentado a:

Edwin Nelson Montoya Munera

Elaborado por:

Martin Villegas Aristizabal

Daniel Gonzalez Bernal

Jorge Alfredo Villarreal Márquez

Universidad EAFIT

Medellin

2023

ARQUITECTURA DEL PROYECTO: ANÁLISIS DE PRECIOS DE CRIPTOMONEDAS EN AWS

Para la arquitectura de este proyecto se implementó un clúster EMR en AWS, entre las apps instaladas por defecto dejamos configurado Jupyter Notebooks. Los Jupyter Notebooks fueron usados como la interfaz principal para el desarrollo y ejecución de los procesos del proyecto. La obtención de datos vivos se realizó mediante la integración con la API de CoinGecko. Esta API ofrece información detallada sobre los precios de diversas criptomonedas. Debido a las limitaciones de la versión gratuita de la API, la actualización de datos se realiza solamente cada minuto y medio, pero para efectos académicos consideramos que el sistema igualmente es válido.

Los datos recopilados con la API de CoinGecko son entonces dirigidos hacia Amazon Kinesis, un servicio de transmisión de datos en tiempo real. Kinesis facilita la ingestión y procesamiento continuo de grandes volúmenes de datos, proporcionando una solución eficaz para el análisis en tiempo real de los precios de las criptomonedas. Este flujo de datos en tiempo real permite reaccionar de manera ágil a cambios en el mercado.

Finalmente, para garantizar la persistencia de los datos procesados, se almacenan en un bucket de Amazon S3. El formato seleccionado para el almacenamiento es CSV, lo que facilita la manipulación y el análisis posterior de los datos. Amazon S3 proporciona durabilidad y escalabilidad, asegurando un acceso rápido y confiable a los datos almacenados.

En resumen, la arquitectura del proyecto se estructura de manera que aprovecha al máximo los servicios gestionados de AWS. Desde la obtención de datos con Jupyter Notebooks y la API de CoinGecko, pasando por el análisis en tiempo real en Kinesis, hasta el almacenamiento persistente en S3, cada componente contribuye a una solución completa y eficiente para el análisis continuo de precios de criptomonedas en la nube de AWS.

IMPLEMENTACION

Con el fin de que el proyecto sea fácilmente reproducible, indicaremos paso a paso a modo de tutorial todas las configuraciones y pasos necesarios para poder reproducir la totalidad del proyecto.

Como primer paso crearemos el Clúster de AWS y lo configuraremos para que contenga la aplicación de Jupyter Notebooks, adicionalmente es necesario crear un bucket para almacenar los datos procesados y tener persistencia sobre los mismos. Este bucket puede ser el mismo utilizado al momento de configurar Jupyter.

Como siguiente paso crearemos una instancia de Kinesis que utilizaremos como Data Stream, esta instancia recibirá los datos vivos que se capturen por medio de un script que estará corriendo en un notebook de Jupyter.

[Amazon Kinesis](#) > [Secuencias de datos](#) > [Crear secuencia de datos](#)

Crear secuencia de datos [Información](#)

Configuración de la secuencia de datos

Nombre de la secuencia de datos

Los caracteres aceptados son letras mayúsculas y minúsculas, números, guiones bajos, guiones y puntos.

Capacidad de secuencia de datos [Información](#)

Modo de capacidad

☐ **Bajo demanda**

Utilice este modo cuando los requisitos de rendimiento del flujo de datos sean impredecibles y variables. Con el modo en diferido, la capacidad del flujo de datos escala automáticamente.

☒ **Aprovisionado**

Utilice el modo aprovisionado cuando pueda estimar de forma confiable los requisitos de rendimiento del flujo de datos. Con el modo aprovisionado, la capacidad del flujo de datos es fija.

Particiones aprovisionadas

La capacidad total de un flujo es la suma de las capacidades de sus particiones. Ingrese el número de particiones aprovisionadas para ver la capacidad total del flujo de datos.

Calculador de particiones

Mínimo: 1; máximo disponible: 500; límite de cuota de la cuenta: 500. [Solicitar aumento de cuota de particiones](#)

Capacidad total de secuencia de datos

La capacidad de particiones está determinada por el número de particiones aprovisionadas. Cada partición captura hasta 1 MiB por segundo y 1000 registros por segundo y emite hasta 2 MiB por segundo. Si las escrituras y lecturas superan la capacidad, la aplicación quedará sujeta a limitaciones controladas.

Capacidad de escritura	Capacidad de lectura
Máximo	Máximo
1 MiB/segundo y 1000 registros/segundo	2 MiB/segundo

El modo aprovisionado tiene un modelo de precios de rendimiento fijo.

[Ver Precios de Kinesis para modo aprovisionado](#)

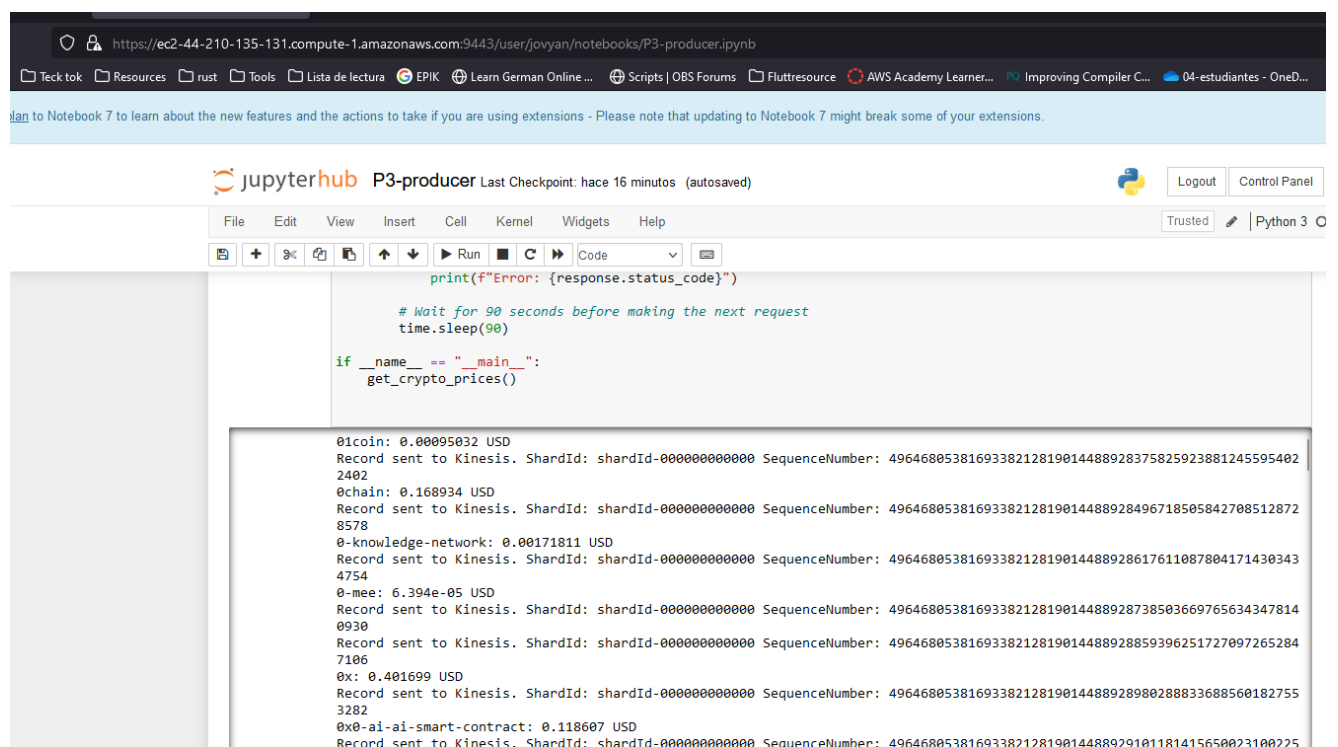
Ahora bien, ya creado nuestro Data Stream procederemos a entrar a Jupyter Notebooks, con el fin de ejecutar el script que nos permitirá obtener los datos vivos es necesario correr los siguientes códigos para las importaciones de las librerías necesarias:

```
!pip install requests
!pip install boto3
!pip install pandas
```

Ahora correremos el script, el código lo pueden encontrar con el nombre de “P3-producer.ipynb” en nuestro GitHub:

<https://github.com/jovillarrealm/jovillarrealm-st0263/blob/main/p3/>

Una vez ejecutado el código veremos algo como esto:



The screenshot shows a Jupyter Notebook titled "P3-producer" running on an AWS EC2 instance. The notebook contains a Python script that prints error messages and then sleeps for 90 seconds. The output shows the script running successfully and sending data to Kinesis.

```
print(f"Error: {response.status_code}")

# Wait for 90 seconds before making the next request
time.sleep(90)

if __name__ == "__main__":
    get_crypto_prices()
```

Output:

```
01coin: 0.00095032 USD
Record sent to Kinesis. ShardId: shardId-000000000000 SequenceNumber: 4964680538169338212819014488928375825923881245595402
2402
0chain: 0.168934 USD
Record sent to Kinesis. ShardId: shardId-000000000000 SequenceNumber: 4964680538169338212819014488928496718505842708512872
8578
0-knowledge-network: 0.00171811 USD
Record sent to Kinesis. ShardId: shardId-000000000000 SequenceNumber: 4964680538169338212819014488928617611087804171430343
4754
0-mee: 6.394e-05 USD
Record sent to Kinesis. ShardId: shardId-000000000000 SequenceNumber: 4964680538169338212819014488928738503669765634347814
0930
Record sent to Kinesis. ShardId: shardId-000000000000 SequenceNumber: 4964680538169338212819014488928859396251727097265284
7106
0x: 0.401699 USD
Record sent to Kinesis. ShardId: shardId-000000000000 SequenceNumber: 4964680538169338212819014488928900288833688560182755
3282
0x0-ai-ai-smart-contract: 0.118607 USD
Record sent to Kinesis. ShardId: shardId-000000000000 SequenceNumber: 4964680538169338212819014488929101181415650023100225
```

Los prints nos muestran una base de la información recopilada y nos indica que todo está corriendo perfectamente y se están capturando los datos vivos para ser enviados a Kinesis y generando su persistencia en el bucket S3.

Con el fin de validar que la información esta llegando correctamente a Kinesis, verificaremos en el panel de control que los datos estén, se debe ver algo como lo siguiente:

← → ↻ https://us-east-1.console.aws.amazon.com/kinesis/home?region=us-east-1#/streams/details/gecko/dataViewer

Importar marcadores... Tech tok Resources rust Tools Lista de lectura EPK Learn German Online... Scripts | OBS Forums | Fluttersource AWS Academy Learner... Improving Compiler C... 04-estudiantes - Ono3...

OWS Servicios Buscar [Alt+S]

Kinesis EMR Amazon Redshift

es flujo de datosgecko se ha creado correctamente.

Amazon Kinesis

Panel

Secuencias de datos

Data Firehose

Apache Flink administrado

Recursos

Plantillas de CloudFormation

Registro de esquemas de AWS

Glue

gecko

Eliminar

Resumen del flujo de datos

Estado Activo	Modo de capacidad Aprovisionado	ARN arnaws:kinesis:us-east-1:170407211578:stream/gecko	Hora de creación 26 de noviembre de 2023, 15:33 GMT-5
Periodo de retención de datos 1 día			

Aplicaciones | Monitoreo | Configuración | **Visor de datos** | Distribución ramificada mejorada (0) | Uso compartido de secuencias de datos: nuevo | Canalizaciones EventBridge

Partición: shardId-000000000000 Posición inicial: Información

Recortar horizonte Obtener registros

Registros (14)

Partición: shardId-000000000000 Posición inicial: Recortar horizonte

Buscar registros

Próximos registros

Clave de pa...	Dato	Marca de tiempo de llegada aproximada	Número de secuencia
shardId-00000...	{"crypto": "01coin", "price": 0.00095083}	26 de noviembre de 2023, 15:43:52 GMT-5	4964680538169338212819014488890657340351844967552...
shardId-00000...	{"crypto": "0chain", "price": 0.166512}	26 de noviembre de 2023, 15:43:52 GMT-5	4964680538169338212819014488890778232933806430469...
shardId-00000...	{"crypto": "0-knowledge-network", "price": 0.00164085}	26 de noviembre de 2023, 15:43:52 GMT-5	4964680538169338212819014488890899125515767893387...

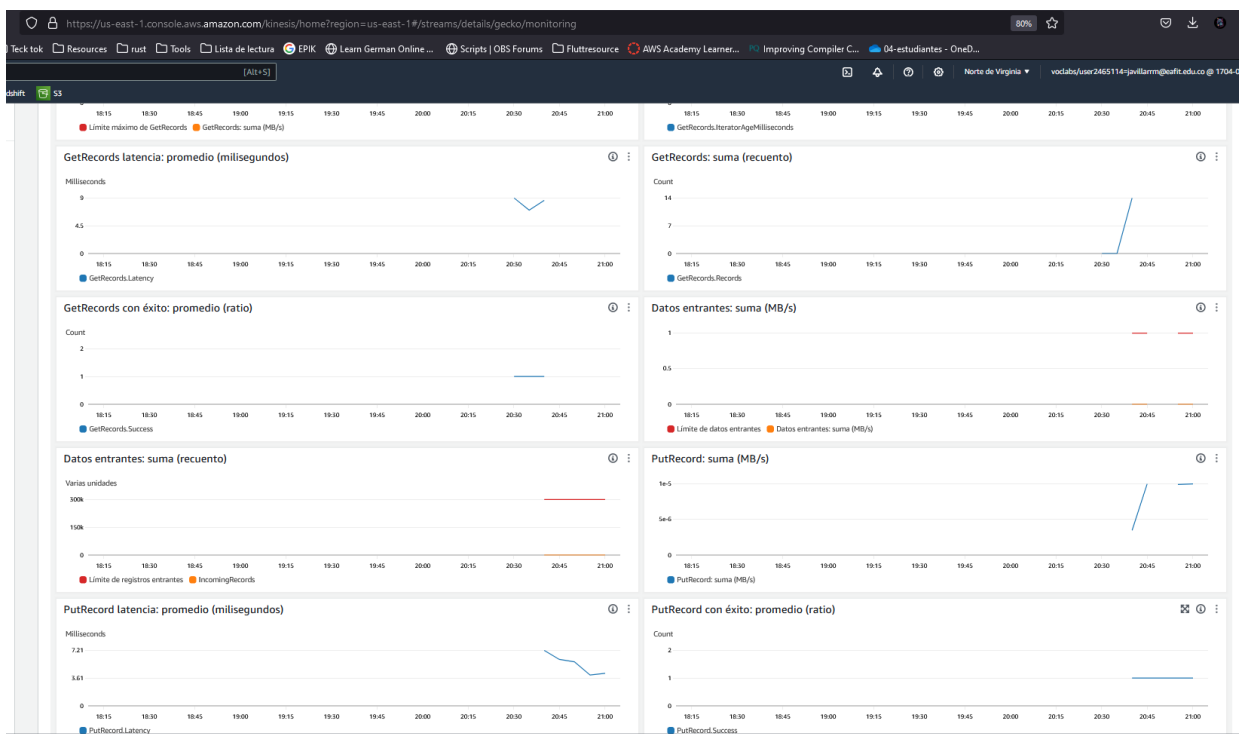
Registros (50)

Partición: shardId-000000000000 Posición inicial: Recortar horizonte

Buscar registros

Clave de pa... ▼	Dato	Marca de tiempo de llegada aproximada
shardId-00000...	{"crypto": "01coin", "price": 0.00095083}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0chain", "price": 0.166512}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0-knowledge-network", "price": 0.00164085}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0-mee", "price": 6.231e-05}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0-mee", "price": 6.231e-05}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0x", "price": 0.401526}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0x0-ai-ai-smart-contract", "price": 0.121079}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0x1-tools-ai-multi-tool", "price": 0.00583905}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0xaiswap", "price": 0.362019}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "0xanon", "price": 0.02436332}	26 de noviembre de 2023, 15:43:52 GMT-5
shardId-00000...	{"crypto": "01coin", "price": 0.00095093}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0chain", "price": 0.16661}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0-knowledge-network", "price": 0.00164085}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0-mee", "price": 6.231e-05}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0-mee", "price": 6.231e-05}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0x", "price": 0.401764}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0x0-ai-ai-smart-contract", "price": 0.121119}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0x1-tools-ai-multi-tool", "price": 0.00583905}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0xaiswap", "price": 0.362019}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "0xanon", "price": 0.02436332}	26 de noviembre de 2023, 15:45:22 GMT-5
shardId-00000...	{"crypto": "01coin", "price": 0.00095112}	26 de noviembre de 2023, 15:46:53 GMT-5
shardId-00000...	{"crypto": "0chain", "price": 0.166863}	26 de noviembre de 2023, 15:46:53 GMT-5

Adicionalmente, utilizando el panel de control de Kinesis, podremos validar por medio de graficas múltiples variables de los datos obtenidos.



Por último, para estar seguros de que los datos se están almacenando de forma persistente, validaremos en el bucket que se estén generando los archivos CSV con los datos obtenidos por medio de la API.

Amazon S3 > Buckets > jorge-bigdatalake > raw/ > csv/

csv/

Objetos | Propiedades

Objetos (2)

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.

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

Buscar objetos por prefijo

<input type="checkbox"/>	Nombre	Tipo	Última modificación	Tamaño	Clase de almacenamiento
<input type="checkbox"/>	4964680538169338212819014488934299562 4399929285514625026.csv	csv	26 Nov 2023 6:09:02 PM -05	981.0 B	Estándar
<input type="checkbox"/>	4964680538169338212819014488940344191 5380697646608941058.csv	csv	26 Nov 2023 6:17:58 PM -05	981.0 B	Estándar