



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

Sprint 1

Grupo 11

Tecnologías de Procesamiento Big Data
3º Grado en Ingeniería Matemática e Inteligencia Artificial

Índice

INTRODUCCIÓN	3
METODOLOGÍA.....	3
RESULTADOS	5
CONCLUSIÓN	6

Introducción

En este primer sprint del análisis de criptomonedas, estará centrado en la recopilación, almacenamiento y estructuración de datos históricos para realizar un análisis posterior de los mismos.

El objetivo principal de este sprint está centrado en crear un repositorio de datos bien estructurado y eficiente para poder facilitar el acceso al histórico de datos y su consulta. Para ello, se realizará la recopilación de datos de diez criptomonedas clave, utilizando la herramienta “TradingView” y la implementación de una estructura de almacenamiento en Amazon S3. .

El trabajo realizado en este sprint es muy relevante, ya que disponer de información histórica bien estructurada permite optimizar la toma de decisiones financieras y desarrollar modelos predictivos más precisos.

Metodología

Los entornos de desarrollo utilizados en este sprint han sido: Python, la librería TradingViewData para la obtención de los datos, y AWS.

La solución diseñada consta de tres partes principales:

1. Extracción de los datos: Para extraer los datos se ha utilizado la API de TradingViewData, obteniendo los datos correspondientes a cada criptomoneda, especificando la frecuencia (daily) y el número de años de los que se quiere obtener el histórico (4 años).
2. Procesamiento de los datos: Para procesar los datos, se han almacenado el histórico de cada criptomoneda en un archivo csv.
3. Almacenamiento de los datos en AWS: La forma en la que se ha decidido almacenar los datos en AWS ha consistido en crear un único Bucket que contiene carpetas asociadas a cada criptomoneda. Después, dentro de cada una de estas carpetas se ha creado una subcarpeta correspondiente a cada año, y por último, dentro de esta se crea una carpeta con los datos de cada mes.
4. Automatización de la obtención de los datos: Para optimizar la adquisición de los datos, se ha diseñado una solución utilizando la librería boto3, que permite interactuar con los servicios de Amazon Web Services sin necesidad de usar la consola web. En esta implementación, se crea una instancia EC2 en AWS para ejecutar procesos de extracción y procesamiento de datos, junto con un cliente que facilita la conexión y gestión de recursos en la nube.

Para comprobar que la obtención de los datos funcionaba de forma correcta, se han realizado varias pruebas. En primer lugar, nos hemos asegurado de que el Exchange asociado a cada criptomoneda contuviese datos suficientes, observando los datos asociados a cada criptomoneda en TradingView, como se puede observar en la figura 1.



Figura 1. Precios de bitcoin en trading view

Como se puede ver en la Figura 2, además de analizar TradingView, también nos hemos asegurado de que los datos se han obtenido de forma correcta, viendo el aspecto de los datos adquiridos.

[symbol	open	high	low \
datetime				
2021-01-31 01:00:00	Bitstamp:BTCUSD	34300.19	34380.21	32193.70
2021-02-01 01:00:00	Bitstamp:BTCUSD	33110.32	34736.10	32333.01
2021-02-02 01:00:00	Bitstamp:BTCUSD	33497.16	36008.67	33431.67
2021-02-03 01:00:00	Bitstamp:BTCUSD	35496.91	37732.87	35382.45
2021-02-04 01:00:00	Bitstamp:BTCUSD	37693.00	38748.52	36197.65
...
2025-01-25 01:00:00	Bitstamp:BTCUSD	104855.00	105237.00	104126.00
2025-01-26 01:00:00	Bitstamp:BTCUSD	104741.00	105424.00	102450.00
2025-01-27 01:00:00	Bitstamp:BTCUSD	102573.00	103186.00	97750.00
2025-01-28 01:00:00	Bitstamp:BTCUSD	102065.00	103745.00	100225.00
2025-01-29 01:00:00	Bitstamp:BTCUSD	101291.00	102953.00	101288.00
	close	volume		
datetime				
2021-01-31 01:00:00	33141.38	4750.123156		
2021-02-01 01:00:00	33532.74	8374.056620		
2021-02-02 01:00:00	35522.16	7225.865020		
2021-02-03 01:00:00	37691.63	8777.332877		
2021-02-04 01:00:00	36996.32	11762.802348		
...		
2025-01-25 01:00:00	104738.00	736.357155		
2025-01-26 01:00:00	102573.00	798.341319		
2025-01-27 01:00:00	102090.00	3210.026254		
...				
2025-01-28 01:00:00	5.480993e+06			
2025-01-29 01:00:00	2.668999e+06			

Figura 2. Estructura de los datos obtenidos

Resultados

Como se puede observar en las Figura 3, 4, 5, 6 y 7, se ha conseguido llevar a cabo la organización de los datos en AWS planteada en el apartado anterior.

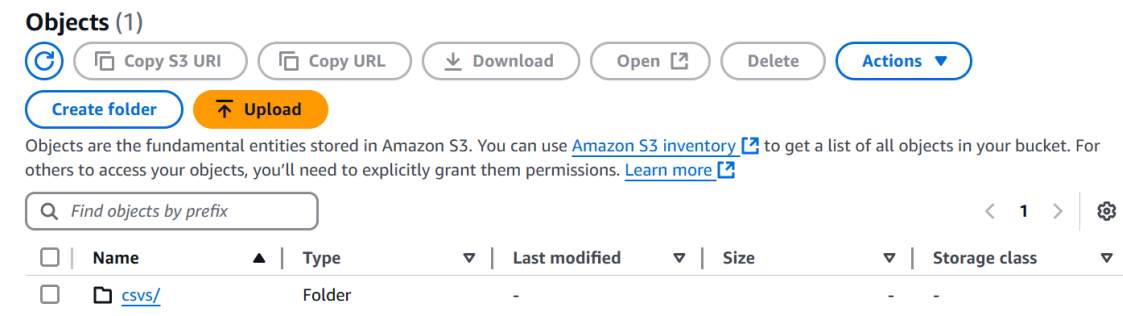


Figura 3. Captura del bucket en AWS

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	AAVEUSD/	Folder	-	-	-
<input type="checkbox"/>	ADAUSD/	Folder	-	-	-
<input type="checkbox"/>	BTCUSD/	Folder	-	-	-
<input type="checkbox"/>	DOGEUSD/	Folder	-	-	-
<input type="checkbox"/>	DOTUSD/	Folder	-	-	-
<input type="checkbox"/>	ETHUSD/	Folder	-	-	-
<input type="checkbox"/>	SHIBUSD/	Folder	-	-	-
<input type="checkbox"/>	SOLUSD/	Folder	-	-	-
<input type="checkbox"/>	XLMUSD/	Folder	-	-	-
<input type="checkbox"/>	XRPUSD/	Folder	-	-	-

Figura 4. Subcarpetas asociadas a cada criptomoneda

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	2021/	Folder	-	-	-
<input type="checkbox"/>	2022/	Folder	-	-	-
<input type="checkbox"/>	2023/	Folder	-	-	-
<input type="checkbox"/>	2024/	Folder	-	-	-

Figura 5. Subcarpetas asociadas a cada año

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	1/	Folder	-	-	-
<input type="checkbox"/>	10/	Folder	-	-	-
<input type="checkbox"/>	11/	Folder	-	-	-
<input type="checkbox"/>	12/	Folder	-	-	-
<input type="checkbox"/>	2/	Folder	-	-	-
<input type="checkbox"/>	3/	Folder	-	-	-
<input type="checkbox"/>	4/	Folder	-	-	-
<input type="checkbox"/>	5/	Folder	-	-	-
<input type="checkbox"/>	6/	Folder	-	-	-
<input type="checkbox"/>	7/	Folder	-	-	-
<input type="checkbox"/>	8/	Folder	-	-	-
<input type="checkbox"/>	9/	Folder	-	-	-

Figura 6. Subcarpetas asociadas a cada mes

<input type="checkbox"/>	AAVEUSD20211.csv	csv	January 29, 2025, 19:55:21 (UTC+01:00)	2.8 KB	Standard
--------------------------	----------------------------------	-----	----------------------------------------------	--------	----------

Figura 7. Archivo csv con los datos diarios

Conclusión

Para concluir, en este sprint se ha logrado diseñar e implementar un sistema eficiente y automatizado para la recopilación y almacenamiento de datos históricos de criptomonedas. La solución combina la API de TradingView con servicios en la nube de AWS, lo que permite extraer, estructurar y almacenar grandes volúmenes de información de manera optimizada y escalable.

El uso de TradingView ha sido fundamental para obtener datos de mercado en tiempo real y datos históricos. Para gestionar y almacenar estos datos, se ha empleado la infraestructura de AWS, aprovechando servicios como S3 para el almacenamiento de datos y EC2 para la ejecución de procesos automatizados. Esto permite que el sistema funcione de manera autónoma.

Además, se llevaron a cabo pruebas para verificar la correcta obtención, procesamiento y almacenamiento de los datos.

Como resultado, se ha desarrollado un sistema robusto que permite disponer de datos estructurados y accesibles, facilitando su utilización en análisis futuros. Esto sienta las bases para futuros sprints en los que se pueda analizar la información adquirida.