

## Arquitectura Real Time de datos de la bolsa de valores.

(Real time stock market Data Architecture)

Integrantes	
Nombres	Sebastián Guerra - Juan David Prieto - Juan David Echeverri - Santiago Puerta
Correos	jsguerrah@eafit.edu.co - jdprietom@eafit.edu.co - jdecheverv@eafit.edu.co – <a href="mailto:spuertaf@eafit.edu.co">spuertaf@eafit.edu.co</a>
Profesor	Edwin Montoya
Curso	ST0263

### A. Descripción general del proyecto

El proyecto se centra en el desarrollo y puesta en producción de una arquitectura en Streaming para el monitoreo de algunas etiquetas de cotización de la bolsa.

### B. Sobre el despliegue y servicios

El despliegue de esta arquitectura se hará en la plataforma GCP y constará de los siguientes servicios:

- *Apache Kafka*: Usado para el almacenamiento de la data cruda proveniente del TradingStreaming.
- *DataBricks (sujeto a cambios)*: Plataforma que contiene el cluster de Apache Spark, en la misma, se realizara todo el pipeline ETL que consume de Apache Kafka, transforma y carga en Google Cloud Storage. Además del despliegue del modelo ML streaming.
- *Google Cloud Storage*: Usado para almacenar la data procesada.
- *Apache Spark*: Usado para el consumo y procesamiento de la data disponible en Apache Kafka.
- *Finhub API*: API que contiene información en tiempo real de la bolsa.
- *TradingStreaming*: Servicio propio desarrollado para la conexión con Finhub API, con el fin de obtener los datos de las etiquetas de cotización de la bolsa; además, este servicio se encarga también de enviar dicha data a Apache Kafka. Este sistema funcionara como un Streaming Pipeline ELT.
- *Grafana (sujeto a cambios)*: Plataforma que se usara con el fin de graficar los datos presentes en Google Cloud Storage.

### C. Referencias

Realmente la idea del proyecto no la tomamos de ningún lado más, principalmente miramos Twitter, pero al ver las limitantes de la API en su versión gratis se nos ocurrió que otra posible fuente de datos vivos seria la bolsa.

- FinHub API docs: <https://finnhub.io/docs/api/introduction>

En el caso de Machine Learning en Streaming tomamos la idea de este artículo:

- <https://avi-soori.medium.com/databricks-automl-sql-streaming-ml-on-the-lakehouse-6ea1aa1fcd18>