

Diseño del DAaaS

Definición la estrategia del DAaaS

Creación de una página web que ofrece diferentes tipos de análisis del mercado financiero en función del tipo de suscripción del usuario:

- Visualización de la evolución del precio de los diferentes activos financieros más mencionado en las redes.
- Visualización en gráficas del sentimiento mostrado en RRSS de los diferentes activos, tanto histórico como cada 6 horas y su posible relación con la variación del precio.
- Visualización de los análisis realizados de las opiniones y noticias del mercado cada 6 horas e históricos derivados de más de 50 importantes medios de noticias financieras de todo el mundo, que cubren acciones, criptomonedas, divisas... y su posible relación con el precio.

Arquitectura DAaaS

1. Máquina Virtual Google Cloud Obtención de datos mediante API's en:
 - Facebook
 - Instagram
 - Reddit
 - Telegram
 - Twitter
 - *Alphavantage*
 - *Interactive Broker*
2. Bucket de Google storage para almacenar los datos
3. Dataproc para realizar el tratamiento de los datos mediante un job.
4. Máquina virtual para realizar IA y ML/DL
5. Cloud function(start/stop instancias, activación job)
6. Cloud Schedule (programación de las diferentes cloud functions)
7. Instancia de Postgres para almacenar datos de la web(usuarios, posts, pagos...)
8. Servidor web
9. Página web

DAaaS Operating Model Design and Rollout

1. Realizar la descarga de datos cada 6 horas mediante API en MV
La MV se levanta de forma automática mediante cloud function y realiza las diferentes descargas programadas y su posterior guardado en bucket.
Una vez finalizadas las descargas se realiza el apagado automático de la MV

2. Tras la descarga se levanta el Dataproc
Encendido programado de forma automática mediante una cloud function programada con cloud Schedule
3. En el nodo principal del clúster hay un script hosteado que realiza la carga de los archivos del bucket al HDFS
4. Un Job ejecutado mediante una cloud function realiza el tratamiento de los datos obtenidos en RRSS y noticias para categorizar en base al activo financiero y posterior envío al bucket de los resultados.
Una vez realizado el Job se realiza el apagado del dataproc de forma automática.
5. La VM-AI/ML se levanta de forma automática al finalizar la carga de los archivos en el bucket por parte del Dataproc y se aplican los diferentes algoritmos de predicción y su posterior carga al bucket
Una vez finalizada la cloud function se apaga de forma automática la MV
A no ser que desde el departamento de ML desactive el apagado.
6. Se realiza la descarga de los resultados en la pagina web para su visualización
7. El usuario podrá acceder a los diferentes análisis en base a su suscripción
8. Toda la información que genere la página web, usuarios, gestión de pagos, métricas... se almacenará en una instancia de postgres en Google cloud

Desarrollo de la plataforma DAaaS.

1. *El encendido de las MV se realiza mediante una cloud function suscrita a un tema pub/sub creado para activar la función que contiene código Node.js. A través de cloud Schedule será invocado a las horas programadas*
2. *La descarga de datos se llevará a cabo mediante un script de python hosteado en la MV que realiza la descarga mediante API:*

RRSS – Facebook, Instagram, Reddit, Telegram. Usaremos la librería de python *snsrape* ya que es muy versátil y nos permite la descarga de información de todas las RRSS de nuestro interés. Nos devolverá archivos .CSV

Noticias fundamentales Alphavantage mediante código python se realiza la conexión con la API

Activos financieros Interactive Brokers, realizaremos la conexión mediante código Python y una interface de conexión llamada IB Getaway

3. *La carga de los archivos en el Dataproc se realiza mediante un script de python hosteado en el nodo principal.*
4. *El job del dataproc, realiza la categorización de los datos obtenidos en RRSS y noticias en base al activo financiero. Extrae únicamente las columnas relevantes para el análisis y crea un csv por activo financiero*

Por ultimo realiza la carga en el bucket en un archivo con la fecha y hora, todo realizado mediante código Python y la librería pandas. Se activará mediante cloud function programada mediante cloud Schedule.

Diagrama:

