Brainstorm  
  
Se va a montar un entorno utilizando **Google Cloud**, para poder realizar el estudio de precios de mercado de la energía en el mundo, conociendo precios Históricos, precios Forecast y precios Forward de forma **automatizada**, para **mejorar** el sistema actual de la empresa, que descarga, organiza y prepara millones de datos de forma manual, siendo un trabajo muy tedioso.

Diseño del DAaaS

Definición la estrategia del DAaaS

Para este proyecto, vamos a realizar la siguiente estrategia:

* Crear un software, mediante **Google Cloud Functions**, que conectará a la API de datos de mercados energéticos del mundo (marco Data Source), y los descargará en “raw”, sin tratar, todos los ficheros siempre vendrán en formato CSV y se almacenarán en una instancia **Google Cloud Storage**, y dicho **Google Cloud Storage** estará dentro de un marco SIT (Storage, Ingest, Transform).

Para ello se hará uso de las API´s de Consultores especializados en el negocio, como pueden ser entre otros:

* + Aurora EOS (<https://api.auroraer.com>)
  + Standard & Poor’s (<https://api.connect.ihsmarkit.com>)
* Una vez se ejecute dicho software y ya los datos estén en el **Google Cloud Storage**, mediante herramientas como **Google Data Fusion** y ayudándose de **Google Data Catalog** (como servicio de administración de metadatos) se podrán realizar todas las funciones del proceso ETL para tratar los datos “raw”.
* Cuando ya se han tratado los datos, se volverán a llevar a una nueva instancia de **Google Cloud Storage**
* Por otro lado, y aun dentro del mismo marco SIT, se procederá a la carga de dichos datos tratados a un **Snowflake Data Cloud**
* Finalmente, y ya en un nuevo marco Visualize, mediante la herramienta **Tableau**, se construirán dashboards “Ad-Hoc” para que los puedan utilizar y visualizar los analistas de mercado energético.

Arquitectura DAaaS

* Fuentes de datos:
  + API´s de consultores de mercados energético del mundo
* Componentes:
  + Google Cloud Functions
  + Google Cloud Storage
  + Google Data Fusion
  + Google Data Catalog
  + SnowFlake Data Cloud
  + Tableau

DAaaS Operating Model Design and Rollout

1. Delimitar que se van a tener 3 marcos
   1. El primero, con las fuentes de origen, marco Data Source
   2. A continuación, donde estará la ingesta, tratamiento y almacenamiento de los datos, marco SIT (Storage, Ingest, Transform)
   3. Finalmente, para poder visualizar todos esos datos de forma, limpia, ordenada y personalizada, tendremos el marco Visualize
2. Crear y preparar el entorno cloud con todas las instancias necesarias, como:
   1. Google Cloud Functions
   2. Google Cloud Storage
   3. Google Data Fusion
   4. Google Data Catalog
   5. SnowFlake Data Cloud
   6. Tableau
3. Con un crontab en nuestro sistema, que se ejecute una vez al día, se disparará un trigger que arrancará todas las instancias del sistema completo que está en el marco SIT, este sistema por defecto, permanecerá apagado para evitar consumos innecesarios de recursos.
4. El software desarrollado, consultará las API´s en busca de nuevos datos, omitiendo los ya cargados anteriormente.
5. Si el sistema detecta nuevos datos, los descargará almacenándolos en la instancia Google Cloud Storage para ficheros “raw”, los tratará haciendo uso de Google Data Fusion y Google Data Catalog, y finalmente, los dejará ya tratados en la instancia Google Cloud Storage de tratados, que mediante otro trigger se encargará de lanzar un proceso para cargar los nuevos datos en SnowFlake Data Cloud.
6. Tras el paso anterior, si ha habido nuevos datos, se notificará al usuario mediante correo electrónico personalizado (indicando mercado, valores, etc), que existen nuevos datos. Para ello se hará uso del servicio de retransmisión de SMTP del Google Workspace dentro del software desarrollado.

Link a Diagrama:

Adjunto fichero drawio en mi repositorio de la práctica:

https://github.com/manuestevez/big\_data\_architecture/blob/main/Pr%C3%A1ctica%20Big%20Data%20Architecture%20-%20Manuel%20Est%C3%A9vez%20Simonet.drawio