Brainstorm

* Mercado Agrícola:
  + Predecir a diferentes plazos los precios de los cereales que cotizan en lonjas provinciales.
  + Predecir los costes energéticos relacionados con el cultivo (gasoleo B y electricidad) para la próxima campaña.
  + Estimar valoración actual de un stock en diferentes ubicaciones del país, previa aporte de los datos de stocks y ubicación del usuario.
  + Ofrecer un ranking de mejor valoración nacional del stock.
  + Ofrecer una gráfica con el histórico de precios.
  + Ofrecer un gráfico de evolución de los precios del gasóleo B y del kw/h eléctrico.
  + Ofrecer en paralelo información de las medias mensuales pluviométricas y de temperatura por provincias.

Diseño del DAaaS

Definición de la estrategia del DAaaS

Definir el catálogo de servicios que proporcionará la plataforma DAaaS, que incluye incorporación de datos, limpieza de datos, transformación de datos, datapedias, bibliotecas de herramientas de analítica y otros.

Datos a incorporar:

* Último precio de los cereales en cada lonja.
* Histórico de precios por cereal en cada provincia, y un par de referencias internacionales.
* Histórico del clima por provincias.
* Histórico del precio del gasóleo B.
* Histórico del precio Kw/h de la electricidad.
* Noticias en páginas especializadas sobre el cereal nacional.
* Noticias internacionales que afecten a los principales graneros mundiales que influyan en el mercado nacional (Rusia, Ucrania, Brasil, Francia, Balcanes y Argentina) especialmente relacionadas con sequias, inundaciones y conflictos geopolíticos
* Noticias nacionales e internacionales que afecten al precio del gasoil y la electricidad/gas.

Arquitectura DAaaS

Definir la selección de componentes, la definición de procesos de ingeniería y el diseño de interfaces deusuario. Diseño y ejecución de Proof-of.Concept (PoC) para demostrar la viabilidad del enfoque DAaaS.

Listado de componentes:

* [VM con python] Crawler de páginas web, para extraer la información de precios del cereal, valores climatológicos y costes energéticos.
* [VM con python] Scraping de noticias sobre cereal, clima y conflictos para anticipar reacciones en los precios.
* [VM con python] Scraping de noticias sobre energía, productores de petróleo y de gas para anticipar reacciones en los precios.
* Google Cloud Storage para hacer el staging de los datos.
* [Dataproc] Hadoop para hacer el procesado de los datos y generar las previsiones
* [Cloud function] Actualizar los datos deL SQL y encender las maquinas.
* [VM con SQL] Base de datos para los resultados del análisis de datos.
* [DBeaver] Para recibir los datos y sincronizar con la SQL cloud.
* [VM con python] Server para el API (máquina virtual con Python y un servidor web).

DAaaS Operating Model Design and Rollout

Personalizar los modelos operativos DAaaS para cumplir con los procesos, la estructura organizacional, las reglas y el gobierno de los clientes individuales. Realizar seguimiento de consumo y mecanismos de informe.

1. Cloud Scheduler arranca todos los días a las 6:00 la máquina virtual de Compute Engine que tiene los arranques programados de los crawler a través de Scrapy:
   1. Crawler de los precios de las lonjas.
   2. Crawler de los precios del gasoil.
   3. Crawler de los precios de la electricidad.

Programo la máquina virtual para que una vez terminado el Crawler esta se apague.

1. Cloud scheduler arranca todos los días a las 6:00 la máquina virtual de Compute Engine que tiene los arranques programados de los scraping de noticias:
   1. Scraping noticias cereal
   2. Scraping noticias clima (sequias/inundaciones)
   3. Scraping noticias conflictos geopolíticos en áreas de influencia
   4. Scraping noticias de energía

Programo la máquina virtual para que una vez terminado el Scraping esta se apague.

1. Los diferentes Crawler y Scraping se programan a través de Scrapy para que guarde los datos obtenidos en un csv que a través de Python se descargarán a diario a DBeaver
2. Programamos a través de una Cloud Function la subida de este csv a Google Cloud Storage, automatizándolo a que se ejecute todos los días con Cloud Scheduler.
3. Staging de los datos en Google Cloud Storage, creando un bucket.
4. Conectamos el bucket a SQL cargando la base de datos.
5. Dataproc para la creación de nuestro Hadoop en la nube y lo conectamos con SQL para que procese los datos.
6. Procesados los datos automatizamos la descarga a DBeaver.
7. Conectamos DBeaver a Tableau y programamos su actualización diaria a final del día.

Desarrollo de la plataforma DAaaS.

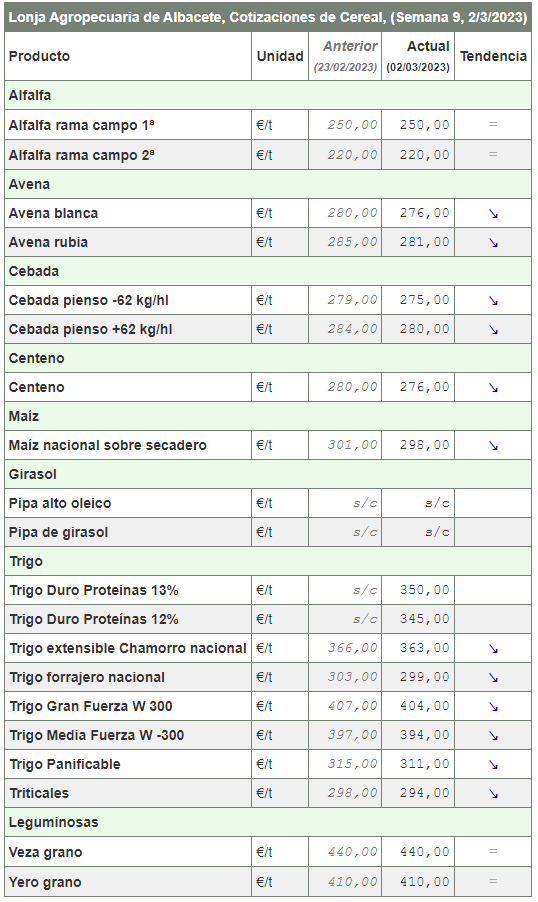
Construcción iterativa de todas las capacidades de la plataforma, incluido el diseño e integración, pruebas, carga de datos, metadatos y población de catálogos y despliegue.

* Crawler de las páginas web:

COTIZACIÓN CEREALES

* + - <https://www.bolsadecereales.com/precios-internacionales>
  + - <https://www.asaja.com/sectoriales/cereales_3>
  + - <https://www.fedeto.es/lonja/cereales.htm>
  + - <http://m.cerealessaiz.com/lonja_de_cuenca.html>
  + - <https://www.feriasymercados.net/index.php/lonja/demo/22>
  + -<https://www.oviespana.com/Articulos/316043-Lonja-de-Cereales-de-Albacete-Cotizaciones.html>

Ejemplo web Lonja Albacete



* + -<https://www.bolsadecereales.com/precios-internacionales>

Ejemplo web cotización cereal en Chicago



PRECIOS ENERGIA

Gasoil

* + <https://www.cetm.es/evolucion-precios-gasoleo/>

Twitter precios gasoil por gasolinera

* + <https://twitter.com/GPortalMinetur>

Electricidad

* + <https://www.somenergia.coop/es/tarifas-de-electricidad/historico-de-tarifas-de-electricidad/>



NOTICIAS AGRICOLAS

* + <https://efeagro.com/agricultura/>
  + <https://www.agroclm.com/>
  + <https://es.euronews.com/tag/agricultura>
  + <https://www.europapress.es/temas/agricultura/>

Link a Diagrama:

<https://drive.google.com/file/d/1UFqdgHnZkClBa0MZdDVE6fbfE2px_Dh5/view?usp=share_link>