

Brainstorm

- Cada vez que tengo que hacer la compra para prepararme el menú semanal, siempre me pregunto si en alguna otra tienda será más barato y mejor.
- Hoy en día debe ser fácil scrapear las páginas de las grandes superficies y comparar los ingredientes de mi lista de la compra para ver a qué supermercado o supermercado me saldría mejor la compra.
- <https://tienda.mercadona.es/> Tiene API (aquí se explota <https://medium.com/@ablancodev/trasteando-la-api-del-mercadona-cff067abc00>)
- <https://www.carrefour.es/> También tiene (<https://helpcenter.channable.com/hc/es/articles/360011191379--C%C3%B3mo-configurar-una-API-de-Carrefour-Espa%C3%B1a>)

Diseño del DAaaS

Definición la estrategia del DAaaS

Voy a crear una aplicación que en función de la lista de ingredientes de la compra que se seleccione recomiende a sus usuarios donde debería ir a realizar la compra para que su cesta de la compra tenga el mejor precio.

Arquitectura DAaaS

- React para crear la aplicación que usarán los usuarios
- API Gateway usaré KONG.
- OracleDB para almacenar los datos.
- Script de Python para automatizar la carga de datos en la BBDD.
- Biblioteca Pandas en Python la normalización de datos. (Semiautomático el proceso de unir todos los csvs en un único (datos_unificados_fecha.csv) para su ingesta en la BBDD, sería necesario un pequeño equipo para su normalización y mantenimiento)
- Script de Python para extracción de los ficheros con formato supermercado_fecha.csv, ie, mercadona_20240513.
- Utilizar las propias APIs de estas grandes superficies para la extracción de datos (manual en un principio).

DAaaS Operating Model Design and Rollout

- Cada martes por la mañana se ejecutará el proceso de extracción de datos.
- Si el martes por la mañana da tiempo a cargar todos los datos se procede a su normalización por la tarde, supervisado por una persona.
- El jueves se debe realizar la carga de los datos a Oracle DB para luego poder alimentar React a través de KONG y que los datos estén actualizados para el viernes y el fin de semana que es cuando se suele comprar.

Desarrollo de la plataforma DAaaS. (ligera descripción del desarrollo)

Desarrollo de trabajo manual

1. Comenzamos realizando las diversas extracciones de los csvs para ver las columnas que los componen.
2. Teniendo estandarizado el formato de los supermercado_fecha.csv procedemos a la creación de un proceso (que pueda ser reutilizado) en Jupyter utilizando la biblioteca de Pandas en Python para la normalización y unión de todas las tablas en único .csv (datos_unificados_fecha.csv).
3. Se crea el script de Python que cargará el datos_unificados_fecha.csv al OracleDB.
4. Se conecta el OracleDB a KONG.
5. Pruebas de seguridad, permisos, conexiones...
6. Comprobar si somos capaces de extraer datos a través de KONG para explotarlos en React.
7. Crear aplicación para el usuario en React. Subcontratación de equipo.

Desarrollo de automatizaciones una vez el proyecto es funcional

1. Una vez tenemos la base de cómo queremos que sean estas tablas se valora la posibilidad de optimizar su extracción, extractor.py y realizarla de forma automática con una periodicidad de una semana.
2. Se deberá crear una automatización de python, loader.py para una vez extraídos los datos se ingestan en OracleDB.
3. Se puede revisar el hecho de generar un trigger, launcher.py, que avise si todos los ficheros supermercado_fecha.csv se hayan generado correctamente y entonces se mande un correo de aviso / se ejecute la normalización automáticamente.
4. Automatización de la normalización, unificacion_normalizacion.py (Siempre deberá ser supervisada y se deberá autorizar la ingesta de datos_unificados_fecha.csv al OracleDB).

5. Automatización de la carga de datos a OracleDB, loader_unificados.py.

Link a Diagrama:

