

# Brainstorm

- *El proyecto consta de una web de consulta, que permita escoger que supermercado se adapta a los productos que necesito y me ofrece el mejor precio. Comparando 3 supermercados principales.*

## Diseño del DAaaS

### Definición la estrategia del DAaaS

- ***Se va a crear un servicio de comparación de precios entre varios supermercados, Este te va a permitir seleccionar productos para simular tu compra, mediante distribución de productos por categorías, o un buscador. A medida que añadas los productos, la web comprobará el precio en todos los supermercados, e irá completando tu carrito de compra, simulando la compra en cada uno de los supers., para que una vez finalices, te muestre el coste total de tu cesta en las diversas opciones. Gracias a eso, conseguiremos tener el conocimiento de, en que supermercado me saldrá más económico hacer mi compra.***

*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 analíticas y otros.*

### Arquitectura DAaaS

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

- API de productos de supermercado, necesitaremos una para cada supermercado a analizar, "Mercadona, Carrefour España y DIA" tienen API disponible.
- Mediante una cloud function programada con Python, descargaremos la información de cada supermercado para que se ejecute 1 vez al día, y así actualizar precios y crear o eliminar productos antiguos.
- Bucket de Google Storage donde almacenaremos las consultas a la API que nos envía la cloud function.
- Generamos un HDFS con un Cluster para Hadoop que coja la información siguiente:
  - La imagen de producto
  - El nombre de producto.

- El código EAN.
  - Precio total.
  - Precio por KG.
  - Categoría de producto.
  - Supermercado que vende producto.
- Una tarea de Elastic que procese la información de Hadoop cuando descargamos datos de la API. Buscará los EAN en los distintos supermercados.
  - Extraemos los datos del Cluster mediante un Script de Python, devolviéndolo así al bucket, con todos los productos precalculados. Para que así puedan ser trabajados desde el bucket
  - Las imágenes de los productos
  - Desde el bucket hacemos Push a la web para presentar los productos.

## 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.*

- Ejecución programada cada 24h de la Cloud Function programada con Python para extraer datos de las API's
- Guardamos resultados en el Bucket.
- Con un Hadoop, vamos a cargar la información del Bucket para usarla con el Elastic.
- Elastic distribuirá los productos de cada supermercado en categorías, y los relacionará entre sí para poder luego trabajar con ellos.
- 

## Desarrollo de la plataforma DAaaS.

- Elastic prepara los productos que vende cada super, organizándolos por categorías.
- Elastic revisa los nombres de las categorías de productos de cada super, si hay repetidos o parecidos, los junta y solo muestra 1 de ellos,
- Elastic incluso si relaciona los productos de los diferentes supermercados mediante el código EAN (PK), cuando comprueba si hay productos repetidos con nombres similares y que coinciden en formato filtra y devuelve al scrip de python solo 1 item de los repetidos, el más barato.

Link a Diagrama: