Comandos para hacer el scafold.

docker-compose build

docker compose up -d

1.1 Archivos

create\_bucket.py

# -> Crea el bucket de trabajo en S3

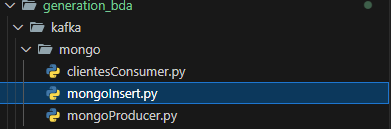
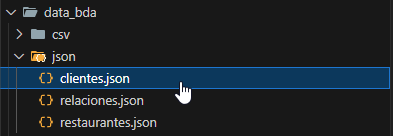
Meterse en el cluster

docker exec -it b17b7945d0d028dfc3b28a0892fb5ebbfd57e36d63ce2077ccf0b8f2e52dce65 /bin/bash

root@b17b7945d0d0:/opt/spark-apps/DataIntegration# python data\_integration.py

# -> Sube directamente los archivos al bucket habitaciones\_data.csv y hoteles.csv

2 MongoDB

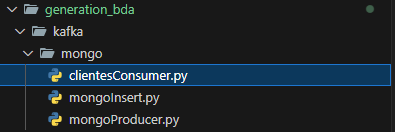
 

mongoInsert.py

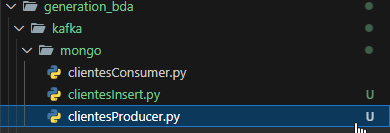
# -> inserta los registros en mongo provenientes de clientes.json

clientesConsumer.py

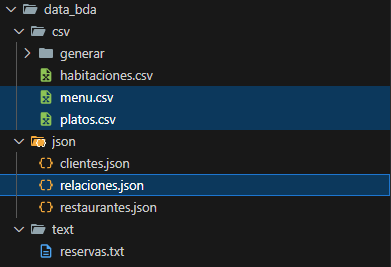
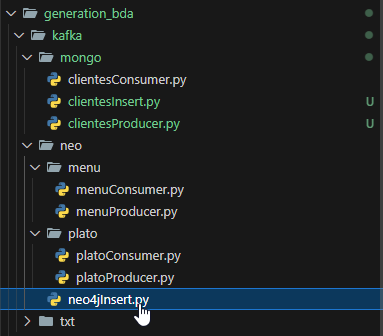
# -> Se pone a la escucha para cuando se le mande la información del Producer.



clientesProducer.py

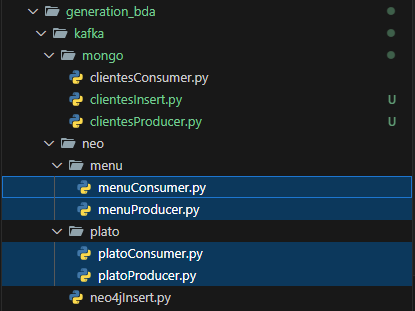
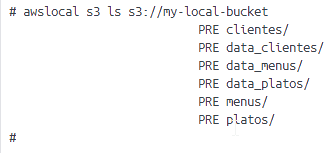
# -> Se encarga de leer desde mongo\_db y mandarla (mensaje estructurado) al Consumer. 

3 Neo4j



neo4jinsert.py

# -> inserta los registros en neo4j provenientes de menus.csv, platos.csv y relaciones.json

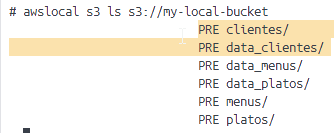
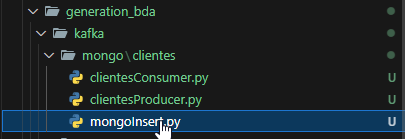
Los Producers se encargan de leer los archivos y mandarlos a los consumers.

Los Consumers se encargan de recibir los mensajes y crear en el bucket los archivos csv menus y platos.

Ya tenemos los archivos.

3.1 Postgres

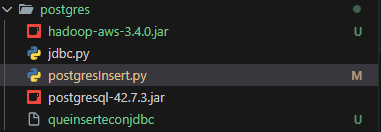
mongoInsert.py



# -> Leer los archivos y los inserta en mongo\_db

clientesConsumer.py

clientesProducer.py

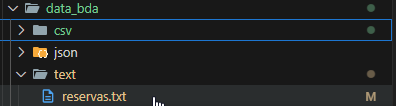
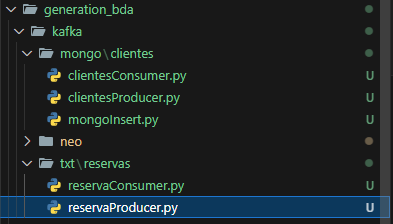
3.1 Postgres empleados. csv y hoteles.csv

postgresinsert.py

# -> inserta en postgres empleados y hoteles

# -> Existe otro archivo que inserta con spark mediante jdbc a postgres desde un csv

3.2 Kafka



Producer

# -> Lee el .txt y manda el mensaje al Consumer

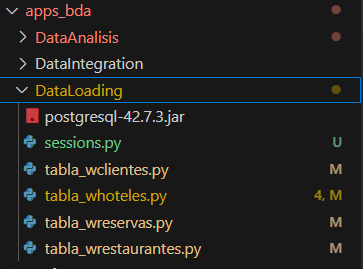
Consumer

# \_> Lee el mensaje e inserta los registros en postgres.

Preparación del WareHouse

Tabla clientes:

* Se crea la tabla con los atributos que nos interesan y se juntan dataframes para insertar los registros.



Consultas con Spark:

Consulta1

1 ¿Cuáles son las preferencias alimenticias más comunes entre los clientes?

Consulta2

2 ¿Qué restaurante tiene el precio medio de menú más alto?

¿Existen tendencias en la disponibilidad de platos en los distintos restaurantes? Hay mismos platos en id\_menu diferentes OK

Consulta3

3 ¿Cuál es la duración media de la estancia de los clientes de un hotel?

¿Existen periodos de máxima ocupación en función de las fechas de reserva?")

Consulta4

4 ¿Cuántos empleados tiene de media cada hotel?

Consulta5 “5.2.5 Ocupación e ingresos del hotel

5 ¿Cuál es el índice de ocupación de cada hotel y varía según la categoría de habitación?" ¿Podemos estimar los ingresos generados por cada hotel basándonos en los precios de las habitaciones y los índices de ocupación?

Consulta6 5.2.6 Análisis de menús

¿Qué platos son los más y los menos populares entre los restaurantes?

¿Hay ingredientes o alérgenos comunes que aparezcan con frecuencia en los platos?

Consulta7

¿Existen pautas en las preferencias de los clientes en función de la época del año?

#- Será hoteles preferidos en épocas por mes/o año

¿Los clientes con preferencias dietéticas específicas tienden a reservar en restaurantes concretos?

Consulta8

¿Existen discrepancias entre la disponibilidad de platos comunicada y las reservas  
reales realizadas? Esto qué es?

Consulta9

¿Cómo se comparan los precios de las habitaciones de los distintos hoteles y existen valores atípicos? Es ver los precios de todas y ver si hay valores que se disparan?

Consulta11