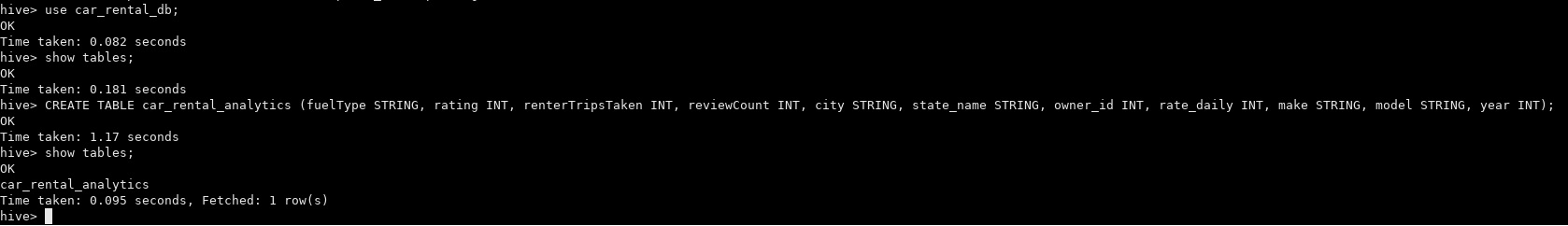
Ejercicio 2: Desarrollo

Alquiler de automóviles Una de las empresas líderes en alquileres de automóviles solicita una serie de dashboards y reportes para poder basar sus decisiones en datos.

Entre los indicadores mencionados se encuentran total de alquileres, segmentación por tipo de combustible, lugar, marca y modelo de automóvil, valoración de cada alquiler, etc.

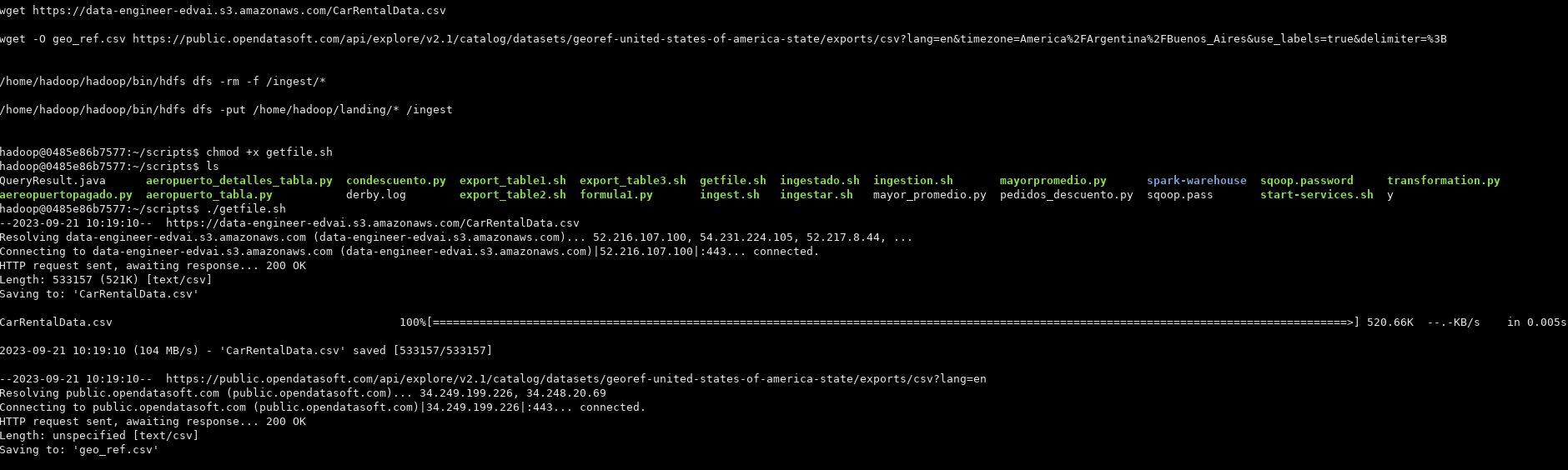
Como Data Engineer debe crear y automatizar el pipeline para tener como resultado los datos listos para ser visualizados y responder las preguntas de negocio.

1. Crear en hive una database car\_rental\_db y dentro una tabla llamada car\_rental\_analytics, con estos campos:



2. Crear script para el ingest de estos dos files

● https://data-engineer-edvai.s3.amazonaws.com/CarRentalData.csv

● https://public.opendatasoft.com/api/explore/v2.1/catalog/datasets/georef-united-st ates-of-america-state/exports/csv?lang=en&timezone=America%2FArgentina%2FBueno s\_Aires&use\_labels=true&delimiter=%3B

3. Crear un script para tomar el archivo desde HDFS y hacer las siguientes transformaciones:

● En donde sea necesario, modificar los nombres de las columnas. Evitar espacios y puntos (reemplazar por \_ ). Evitar nombres de columna largos

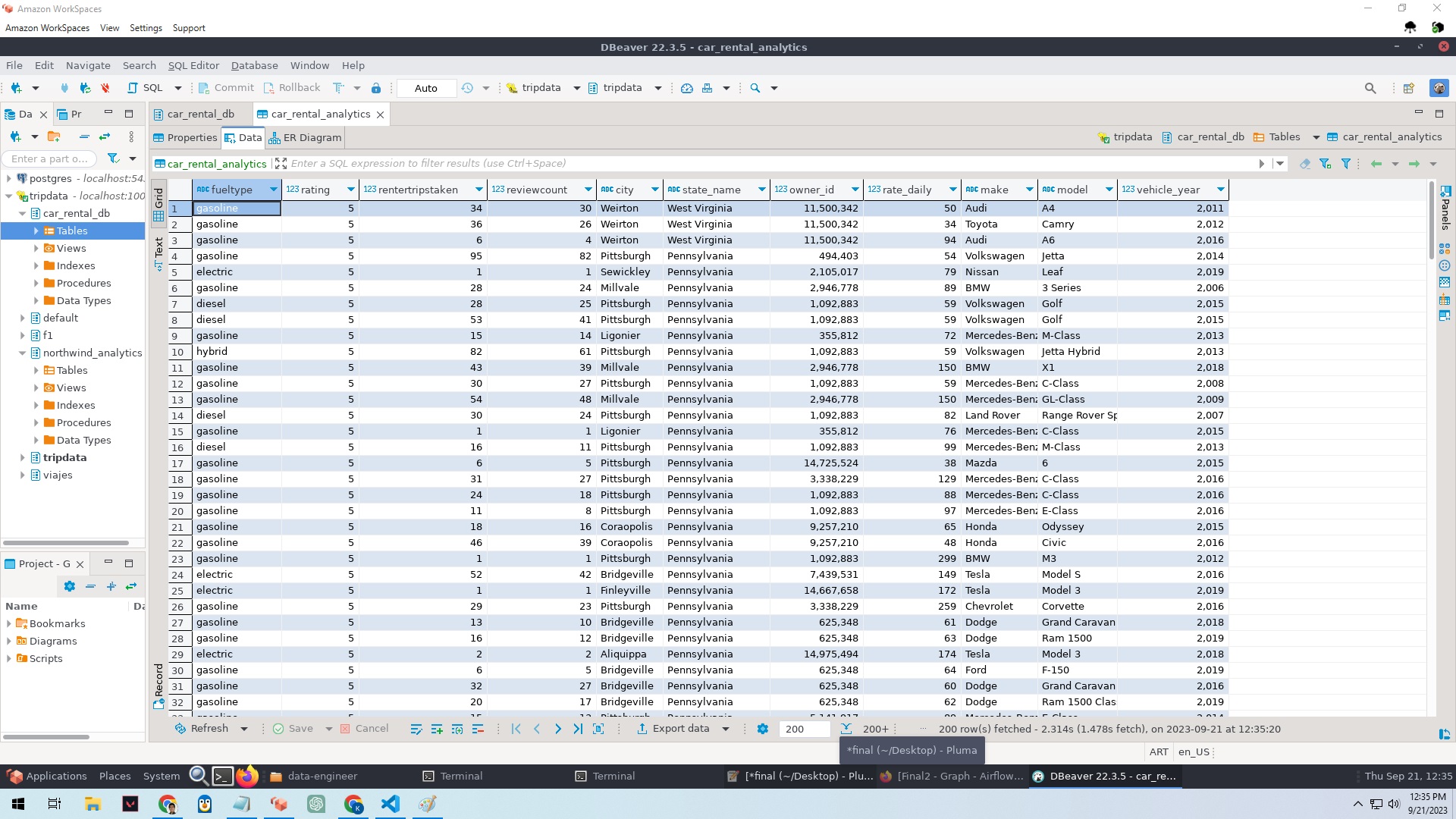
● Redondear los float de ‘rating’ y castear a int.

● Joinear ambos files

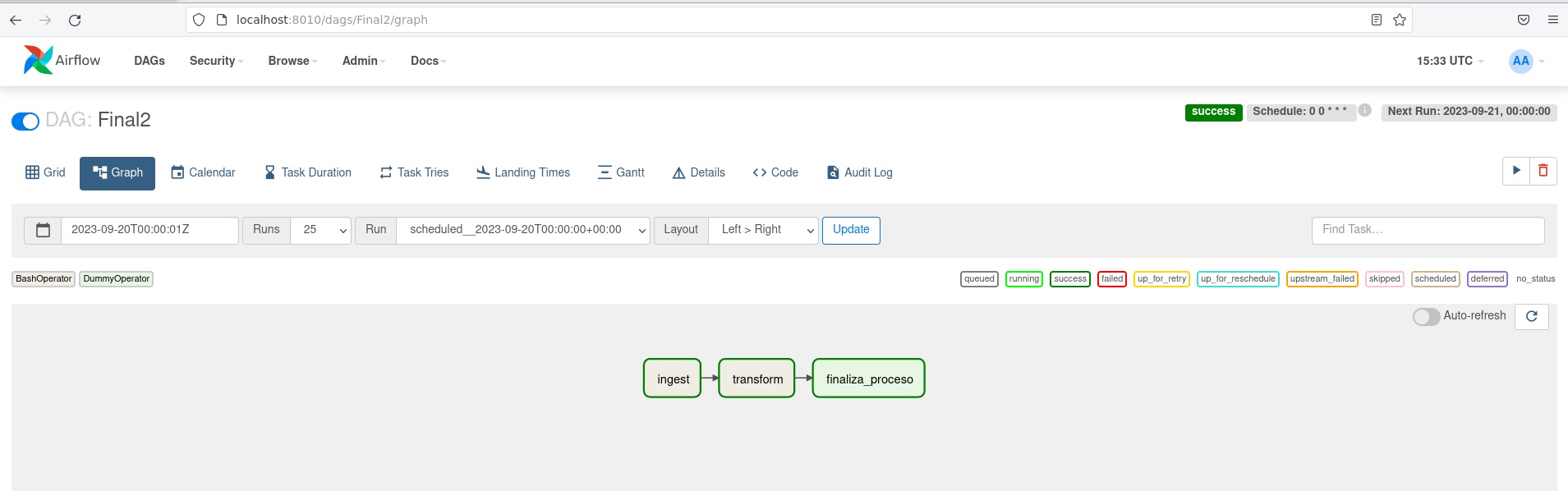
● Eliminar los registros con rating nulo

● Cambiar mayúsculas por minúsculas en ‘fuelType’

● Excluir el estado Texas Finalmente insertar en Hive el resultado



4. Realizar un proceso automático en Airflow que orqueste los pipelines creados en los puntos anteriores. Crear dos tareas: a. Un DAG padre que ingente los archivos y luego llame al DAG hijo b. Un DAG hijo que procese la información y la cargue en Hive



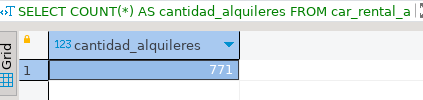
5. Por medio de consultas SQL al data-warehouse, mostrar:

a. Cantidad de alquileres de autos, teniendo en cuenta sólo los vehículos ecológicos (fuelType hibrido o eléctrico) y con un rating de al menos 4.

**SELECT** **COUNT**(\*) **AS** cantidad\_alquileres

**FROM** car\_rental\_analytics

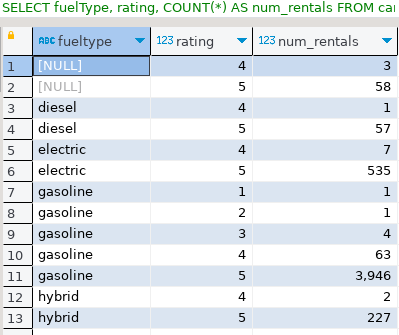
**WHERE** fuelType **IN** ('hybrid', 'electric') **AND** rating >= 4;



**SELECT** fuelType, rating, **COUNT**(\*) **AS** num\_rentals

**FROM** car\_rental\_analytics

**GROUP** **BY** fuelType, rating;



b. los 5 estados con menor cantidad de alquileres (crear visualización)

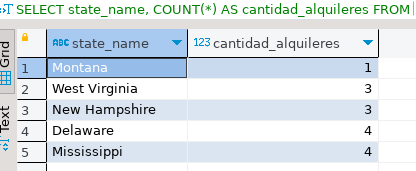
**SELECT** state\_name, **COUNT**(\*) **AS** cantidad\_alquileres

**FROM** car\_rental\_analytics

**GROUP** **BY** state\_name

**ORDER** **BY** cantidad\_alquileres **ASC**

LIMIT 5;



c. los 10 modelos (junto con su marca) de autos más rentados (crear visualización)

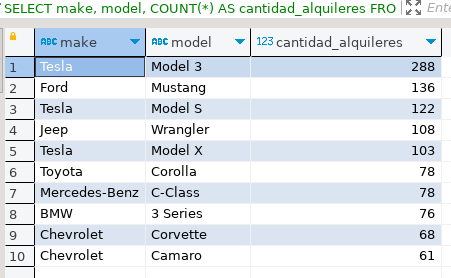
**SELECT** make, model, **COUNT**(\*) **AS** cantidad\_alquileres

**FROM** car\_rental\_analytics

**GROUP** **BY** make, model

**ORDER** **BY** cantidad\_alquileres **DESC**

LIMIT 10;



d. Mostrar por año, cuántos alquileres se hicieron, teniendo en cuenta automóviles fabricados desde 2010 a 2015

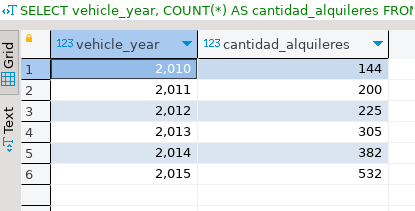
**SELECT** vehicle\_year, **COUNT**(\*) **AS** cantidad\_alquileres

**FROM** car\_rental\_analytics

**WHERE** vehicle\_year **BETWEEN** 2010 **AND** 2015

**GROUP** **BY** vehicle\_year

**ORDER** **BY** vehicle\_year;



e. las 5 ciudades con más alquileres de vehículos ecológicos (fuelType hibrido o electrico) f. el promedio de reviews, segmentando por tipo de combustible

**SELECT** city, **COUNT**(\*) **AS** cantidad\_alquileres

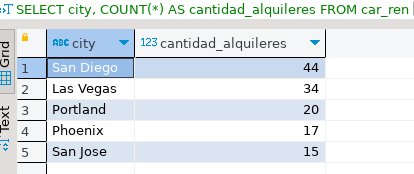
**FROM** car\_rental\_analytics

**WHERE** fuelType **IN** ('hybrid', 'electric')

**GROUP** **BY** city

**ORDER** **BY** cantidad\_alquileres **DESC**

LIMIT 5;



6. Elabore sus conclusiones y recomendaciones sobre este proyecto.

Ciudades como San Diego, Las Vegas y Portland encabezan la lista de alquileres de vehículos ecológicos. Esto puede deberse a factores como políticas medioambientales, incentivos o una mayor conciencia ecológica en estas ciudades sumado al cambio generacional y la tecnología que se encuentra a nuestro alcance, se ve reflejado en la demanda considerable de vehículos ecológicos y que, además, muchos de ellos tienen altas calificaciones, lo que refleja una satisfacción de los mismos hacia estos vehículos.

Los vehículos eléctricos, especialmente de la marca Tesla y el Model 3, son los más rentados, indicando su creciente popularidad en el mercado de alquiler, también los híbridos tienen el mayor promedio de reviews, seguidos por los de gasolina y eléctricos.

por la popularidad y buen rating de los vehículos ecológicos, recomiendo expandir la flota de estos vehículos, especialmente en áreas con baja presencia.

Sugiero investigar las razones de los bajos números de alquileres en ciertos estados y plantear estrategias de marketing específicas para esas zonas.

tener un lugar donde recolectar reviews para mejorar el servicio.

Se propone incentivar a los usuarios a dejar reviews mediante descuentos o promociones.

7. Proponer una arquitectura alternativa para este proceso ya sea con herramientas on premise o cloud (Si aplica)

GCP

Cloud Storage >> DataProc