Task:

Scenario: Please write code to process the OWNERSHIP OG DATA

Steps:

1. We need our code to be written using a distributed processing framework to process file from a location to a table.
2. Image we are processing a big file, our goal should be to process the files in multiple nodes could be at least 2
3. Add an orchestration framework like airflow to execute the process at a certain time.
4. Optional: show the result of the processing (successful or failure ) in any dashboard.

The result will be to show all the data in the table.

Scenario: Please write code to process the OWNERSHIP UPDATED DATA

Steps:

1. We process this new updated data file.
2. It should only update the records which are only updated and we should not touch the records which are not matching.

The final result will be the data which should show the updated data as well as the old data which were inserted the first time.