Improvements to the data pipeline

1. Scheduling data orchestration tool like Airflow can be used to automate the data pipeline execution
2. The source and target storage is windows folder; this need to changed as database or clouds based storage for parquet file
3. Event driven execution: starting automatic pipeline execution depends on availability of new file can be done using Airflow