Wells Fargo – Data platform

Business Case:

Wells Fargo has multiple internal applications that send huge amounts of data (***petabytes***) in CSV format **daily** in the company Azure cloud. We are required to perform **data/schema validation** on this incoming data and then needs to be stored in a **delta table** for downstream systems to make use for analysis.

Task:

1. Validation needs to be applied as follows:
   1. Need to check for **duplicate rows**. If file contains duplicate rows, it’s rejected.
   2. Need to **validate the data format** for all the data fields, the date column names, and the date format is stored in the **Azure SQL server**, if validation fails the file is rejected.
2. Need to move all rejected files to Rejected folder.
3. Need to move all valid files to Staging folder.
4. Write all the validated files as a delta table in **Azure Databricks**