

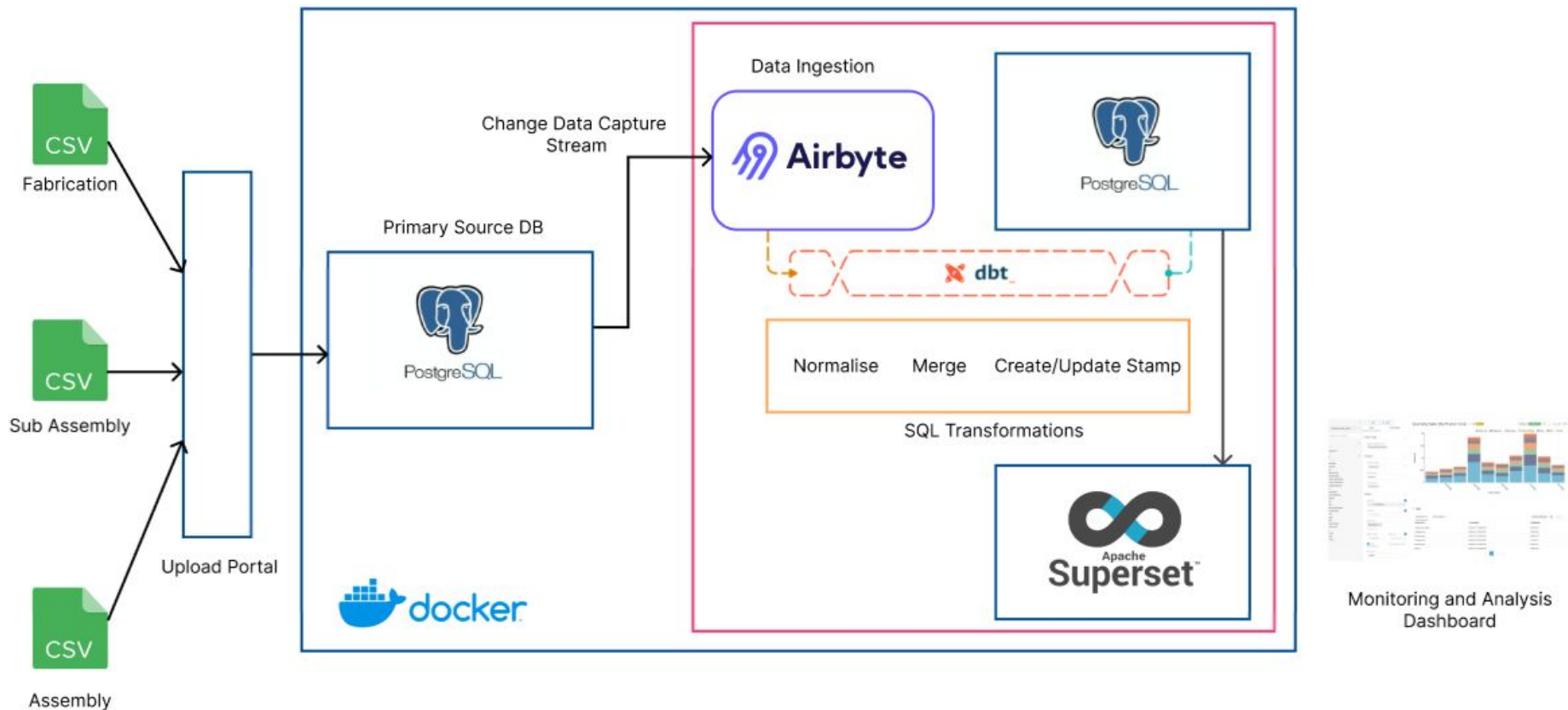
Problem Statement:

To design a Data Lake based architecture for a washing machine manufacturing industry use-case

Team Aviators

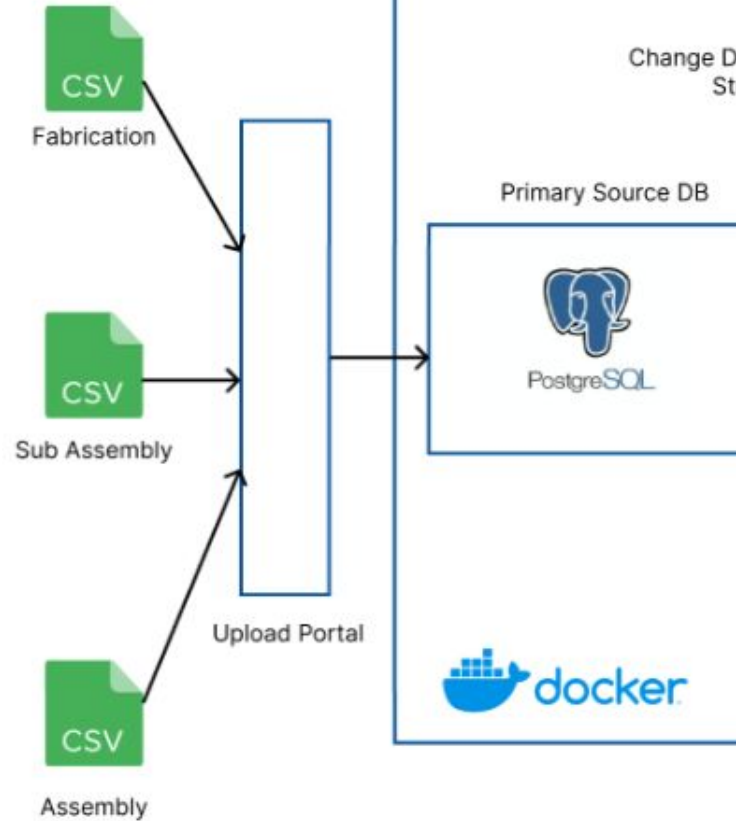


Architecture



Primary Data Source

- ReactJS based platform to upload CSV files.
- CSVs uploaded by each department pushed into primary PostgreSQL DB



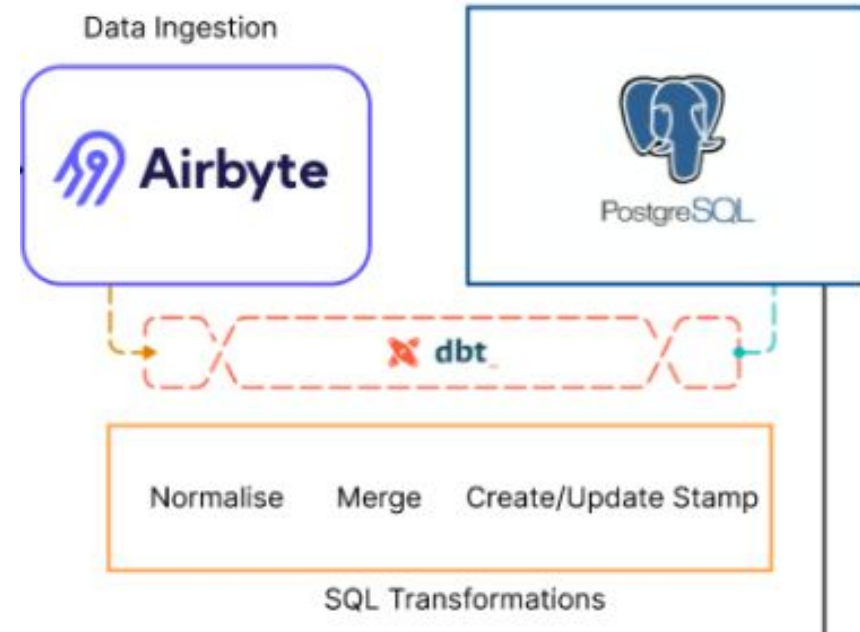


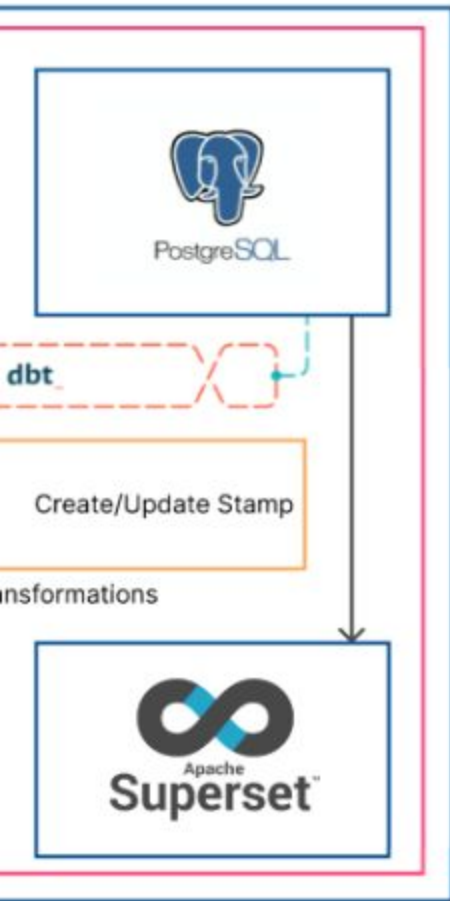
Date Ingestion (Extract and Load)

- Airbyte (an open source tool) used.
- Change Data Capture based Incremental Update method is used.
- Cron jobs run sync every 15 mins ensuring Near Real Time sync.

Transformation

- SQL Transformation applied using dbt (Data Build Tool) (an open source project)
- Data is normalised. Department wise data is merged to create new views.
- Historical data is not deleted, rather time stamping is done.





Monitoring and Analysis
Dashboard

Visualisation

- Apache Superset (another open source BI and analytics tool) used to provide visualisations via dashboard.
- Departments get insight into average inventory days of intermediate products given the demand and supply plots.

Thank You

