DATA PIPELINE

Ingestion

Batch Processing

Real-Time Queries

XML, TSV

Data Cleansing
JSON Hive SerDe