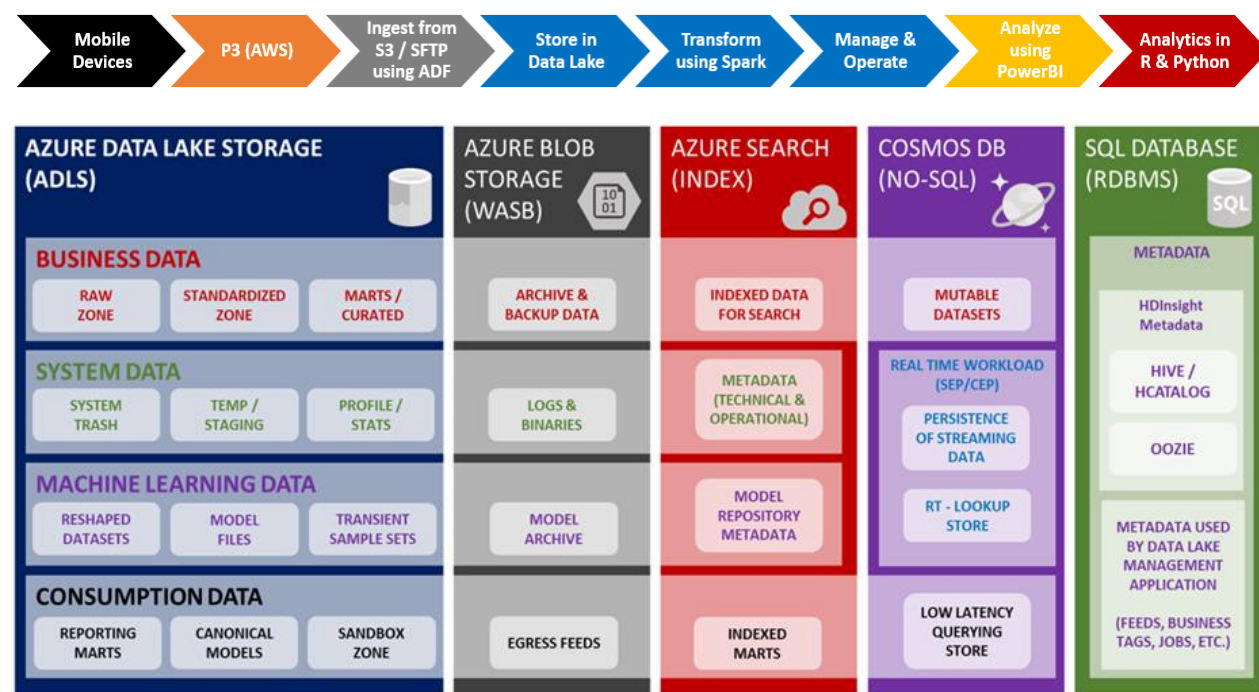


Mobile Broadband Data Analytics on Azure Data Lake

UK government-approved regulatory and competition authority for the communications services

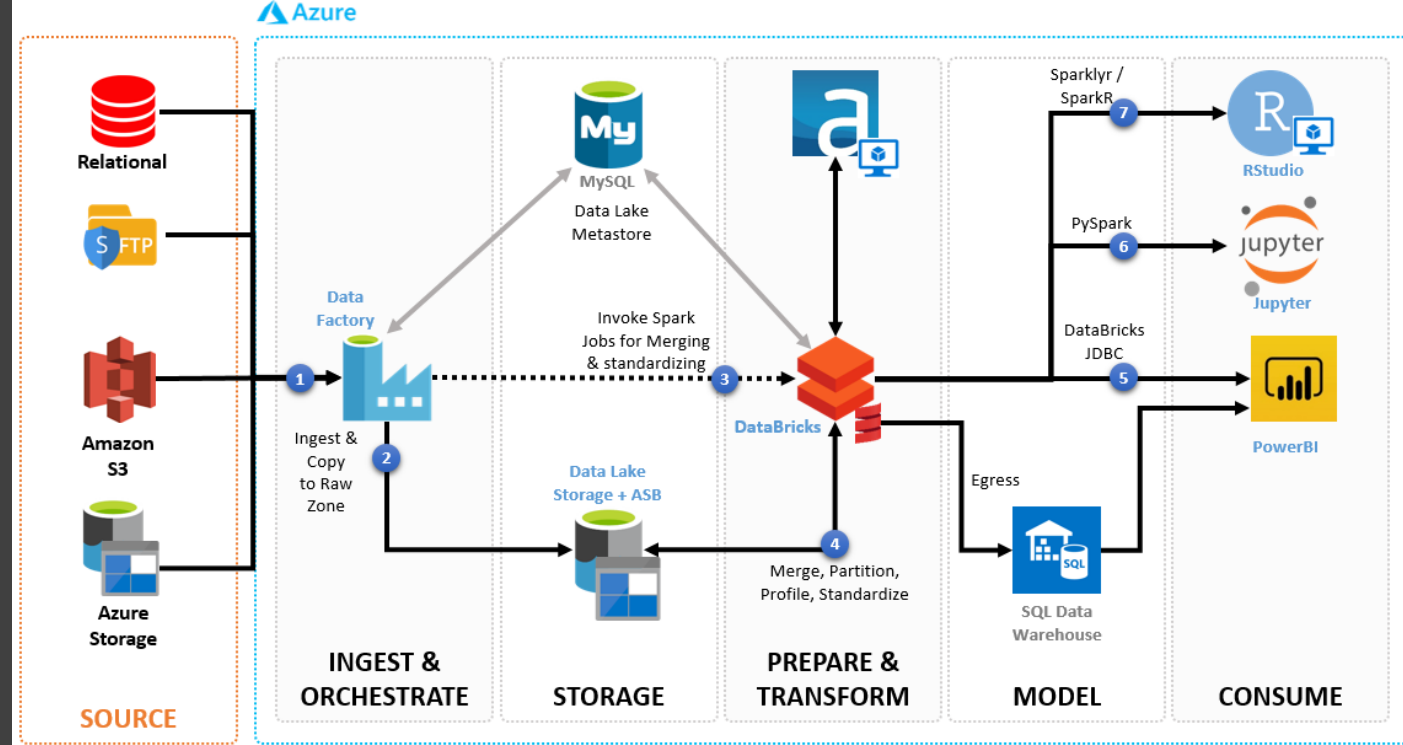
THE CLIENT PROBLEM STATEMENT

- Client wanted to build a Data Lake on Microsoft Azure platform and start with ingesting market research data in a governed manner.
- Client required NIIT implement a Data Lake on Microsoft Azure Cloud with initial use case of ingesting & storing Mobile Broadband Data (from a third-party data provider)
- Primary technologies to be used were Azure Data Factory for Sourcing, Ingestion & Orchestration, ADLS or ASB for Data Lake storage, Azure Databricks for reshaping & processing the data and different analytical tools like PowerBI, RStudio & Jupyter (connected with DataBricks) for doing the analysis.



SOLUTION

- NIIT had setup a Data Lake on Azure Cloud environment leveraging PaaS services on client's Azure EA subscription
- NIIT worked with P3 in designing the landing zone
- Batch data ingestion pipelines were developed using Microsoft Azure Data Factory (ADF)]
- Azure DataBricks Premium Tier were used for Data Analytics Workload (DataBricks)
- Template driven approach for ingestion using ADF were used
- NIIT designed the storage layer with three key zones – Raw, Archive & Standardized, using Azure Storage Blob as the underlying storage technology
- Developed Spark transformations for loading data lake zones using Java/Scala for better performance
- Solution covered governance capabilities like Data Lineage and Captured data profiling statistics in hive tables for every run at column level
- Demonstrated Alteryx integration for Data Preparation on Data Lake using DataBricks as the compute engine
- BI/Dashboard were built using Power BI that connected with DataBricks cluster and leveraged P3 datasets.
- Developed & demonstrated use case using RStudio/Jupyter on P3 data using DataBricks as the compute engine.



BUSINESS OUTCOME



- Readily availability of data to various users for different type of analysis
- Data Lake implementation helped client in storing & analyzing large volumes of P3 data
- Ingestion pipelines helped to bring in Six P3 datasets to Azure data lake and populated various data lake zones.
- Data is now flowing in scheduled weekly batches from P3 to client's data lake and can be queried.

