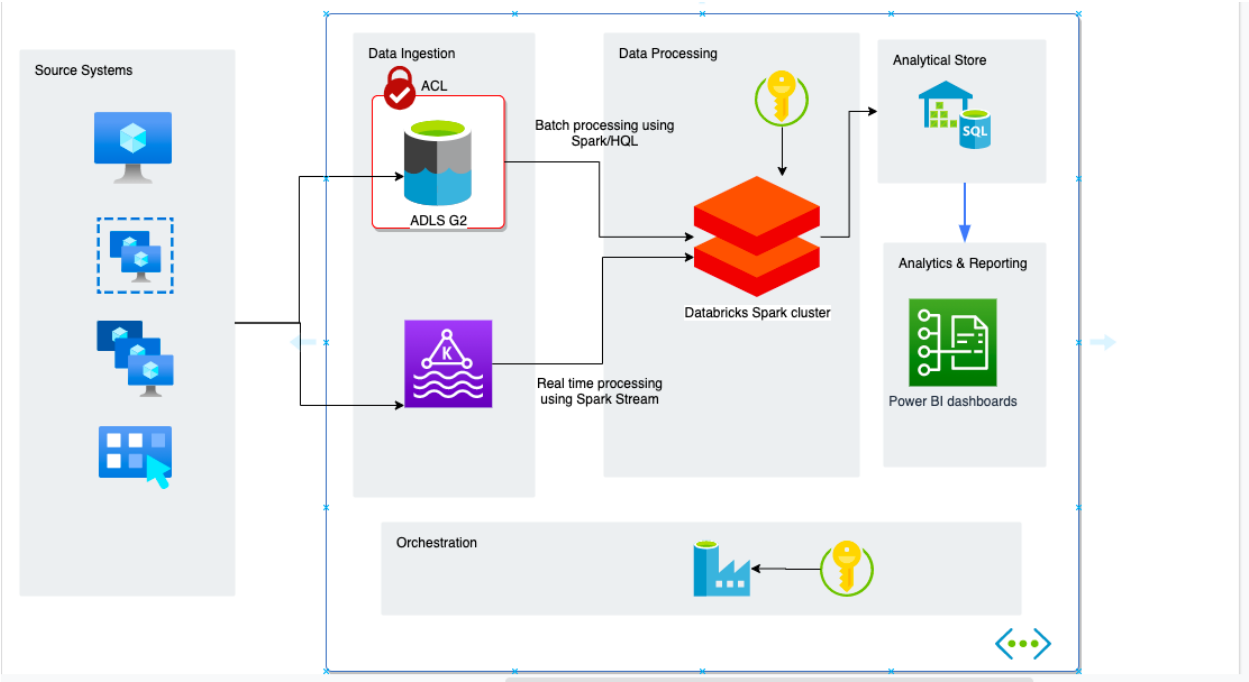

Terms used

- AKV : Azure KeyVault
- ADB : Azure Databricks
- ADF : Azure DataFactory
- ADLS : Azure Data Lake Gen
- AAD : Azure Active Directory
- DW : DataWareHouse
- ACL : Access Control List

Source Systems

- We can have multiple source system from where we will consume metrics data.
- Metrics includes usage/cost of
    - CPU
    - Disk
    - Networks

Data Ingestion

- We will storing data consumed from sources into Azure Data Lake Gen2
- Data ingestion will be done using ADF pipelines which will have copy activities to copy metrics data to ADLs
- We will also configure Kafka streaming/HDinsight Kafka Cluster to consume real time metrics

Data Processing

- Batch Processing
    - Daily ADF pipeline will copy metrics to ADLS.
    - We will create external hive tables on top of ADLS storage.
    - We can use hive queries to process our data using databricks spark cluster
- Real Time Processing
    - We will use Spark Streaming APIs to process real time data.

Analytics Store

- Processed data will be stored into SQL DW.

Report/Dashboard Generation

- Power BI dashboard will be generated and presented to end user .
- Power BI will consume the data from SQL DW

Security

- We will store our credentials - keys & secrets in AKV.
- We will register our application in AAD and will be using service principal authentication in ADF as well as in ADB.
- Data stored in ADLS/ADB/ADF Will be encrypted using customer managed keys stored in AKV. This covers encryption at rest. Azure by default supports encryption in transit
- ACL will be applied to ADLS containers.

Networking

- Virtual network will be created with appropriate address space and peering.
- Private endpoints for blobs and containers will be created.