Let’s build a complete End to End Azure Data Engineering Project. In this project we are going to create an end-to-end data platform right from Data Ingestion, Data Transformation, Data Loading and Reporting. The tools that are covered in this project are,

1. Azure Data Factory
2. Azure Data Lake Storage Gen2
3. Azure Databricks
4. Azure Synapse Analytics
5. Azure Key vault
6. Azure Active Directory (AAD) and
7. Microsoft Power BI

The use case for this project is building an end-to-end solution by ingesting the tables from on-premises SQL Server database using Azure Data Factory and then store the data in Azure Data Lake. Then Azure databricks is used to transform the RAW data to the most cleanest form of data and then we are using Azure Synapse Analytics to load the clean data and finally using Microsoft Power BI to integrate with Azure synapse analytics to build an interactive dashboard.

Also, we are using Azure Active Directory (AAD) and Azure Key Vault for the monitoring and governance purpose.

Architecture

A diagram of a blockchain

Description automatically generated

A screenshot of a computer

Description automatically generated

Plan:

1. **Azure Data Factory**

**Setup integration runtime to connect the onpremises to the cloud:**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A pipeline that:**

* **Copy all data from on premises to Azure DataLake storage (parquet Format)**
* **Using Databricks to mount storage, do 2 steps cleanup and transformations (ETL)**
  + **Keep the raw data in the bronze layer of Azure Data Factory:**
  + **Mount a silver storage to do 1st hand transformation and cleanup:**
  + **Mount a gold storage for a final transformation.**
  + **Bronze to Silver transformations**
  + **Silver to Gold transformations**

*A screenshot of a computer

Description automatically generated*

*A screenshot of a computer

Description automatically generated*

*A screenshot of a computer

Description automatically generated*

1. **Azure Synapse Analytics**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Microsoft Power BI**

**A screenshot of a computer

Description automatically generated**