# DRAW YOUR OWN ARCHITECTURE

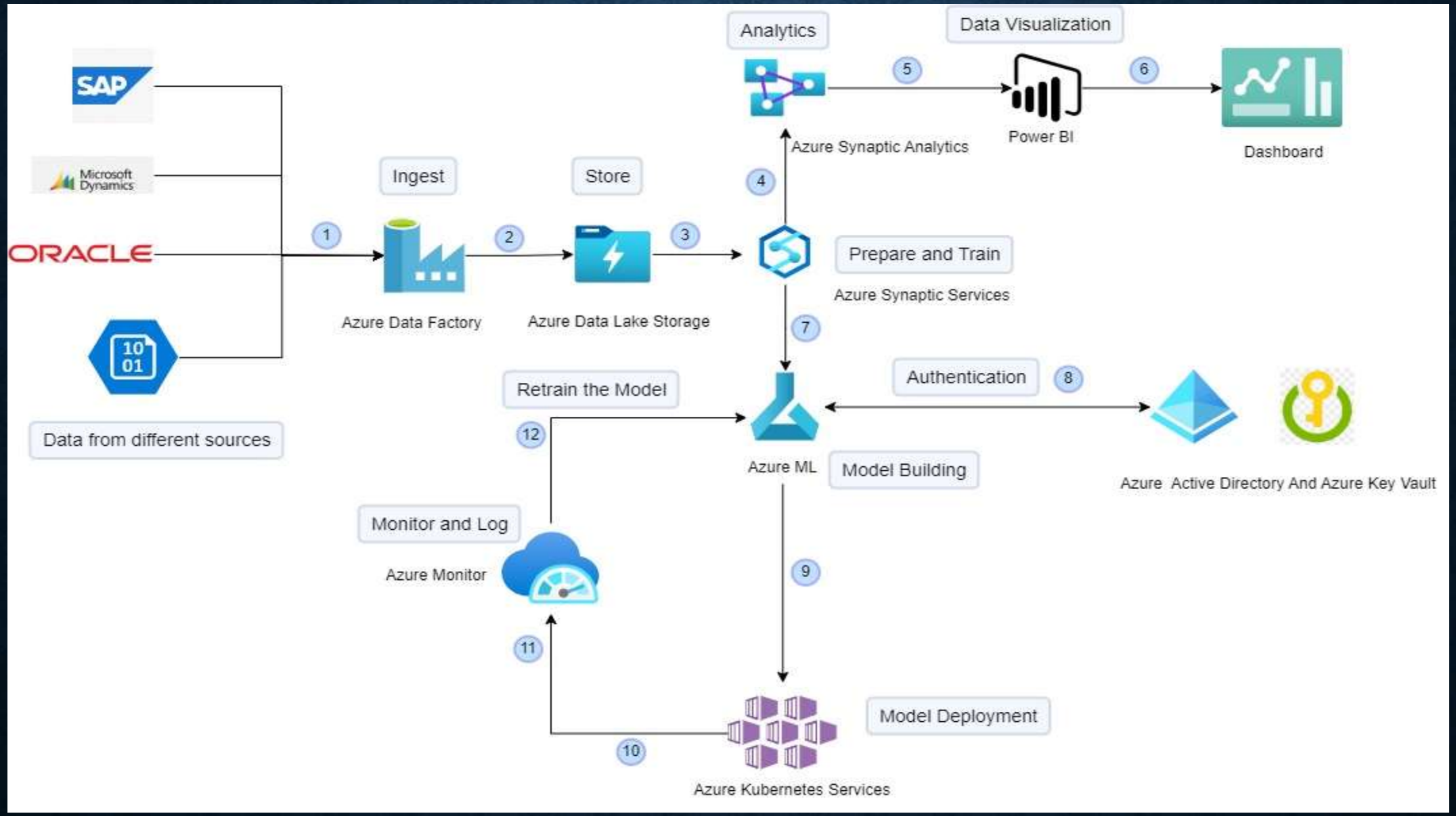
Assignment: 3

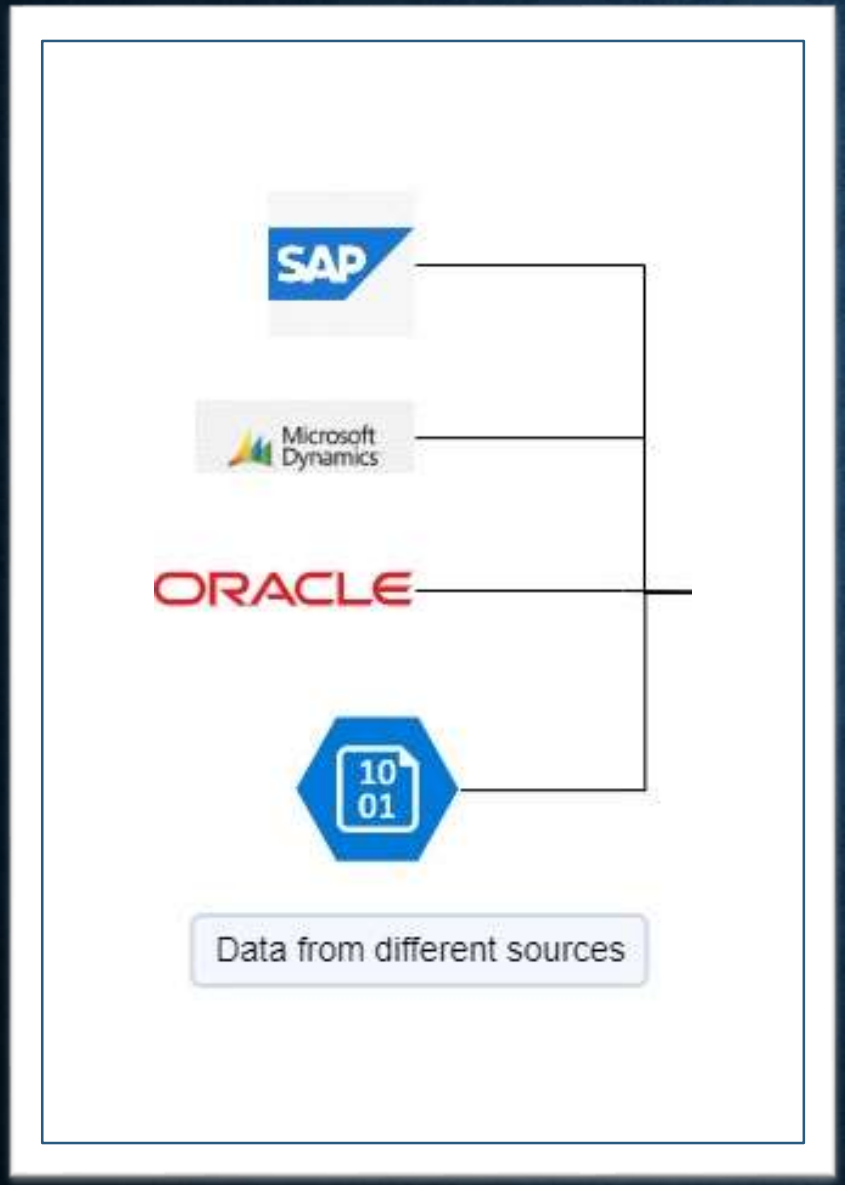
Samrat Thapa – 200481763

# PROBLEM STATEMENT

The STK company is a brand-new start-up responsible to deliver 95% of all the products sold by Amazon in Canada. This company is using SAP, ORACLE, and Microsoft Dynamics 365 CRM, as the main data sources. The company also has some data stored in a blob storage service on Azure (CSV files and unstructured data). The main idea is to move ALL the data to a cloud instance (Azure). They need a unique place to store all the data and to help them to explore the data, generating data analysis, and prep the structure for future AI projects.

# Architecture

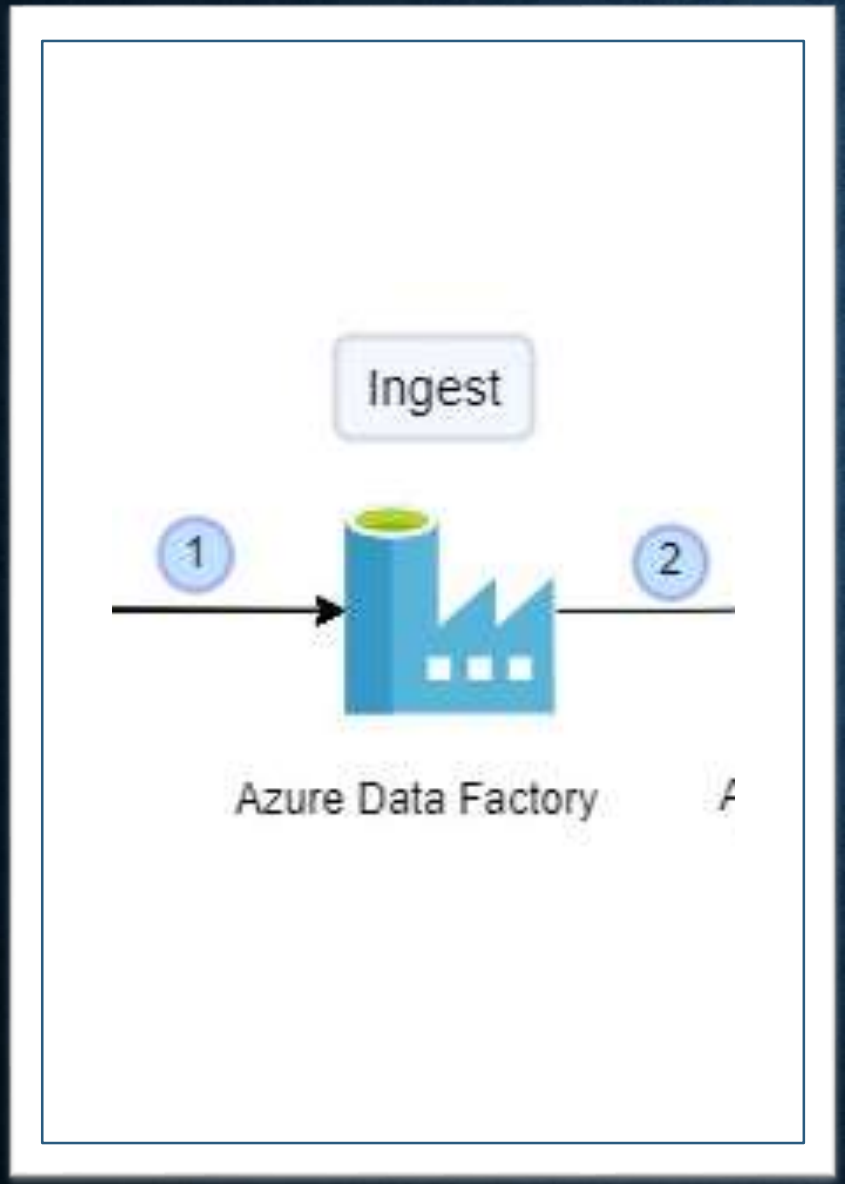





# DATA SOURCE

In our scope, we are given 4 data source form which we have to extract our data. The data source are:

1. SAP
2. Microsoft Dynamic 365 CRM
3. Oracle
4. Azure Blob Storage

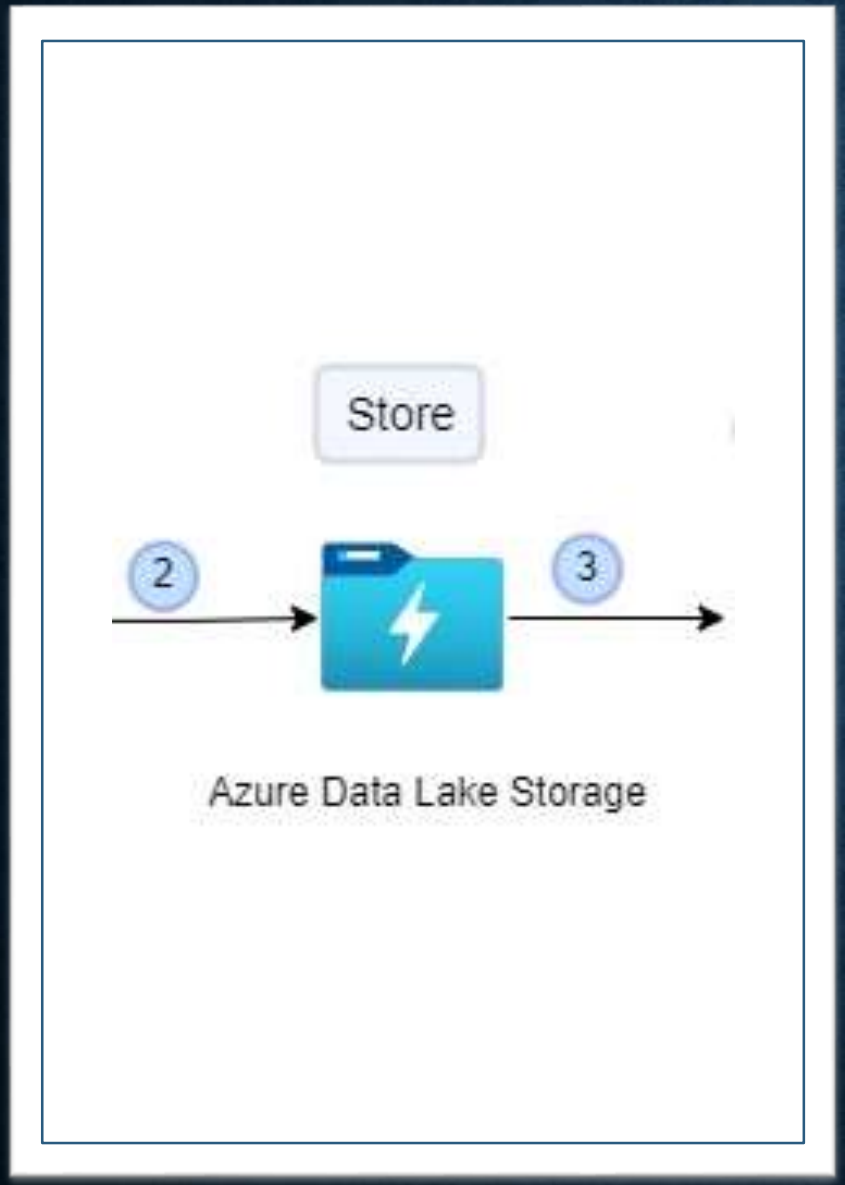



# DATA INGEST

All the data which are from our data sources either structured or unstructured are then send into data factory.

Data Factory is a cloud-based ETL and data integration service  that allows to create data driven workflows for data movement.

It helps us to move data from different sources and store it in azure service like Azure Data Lake.

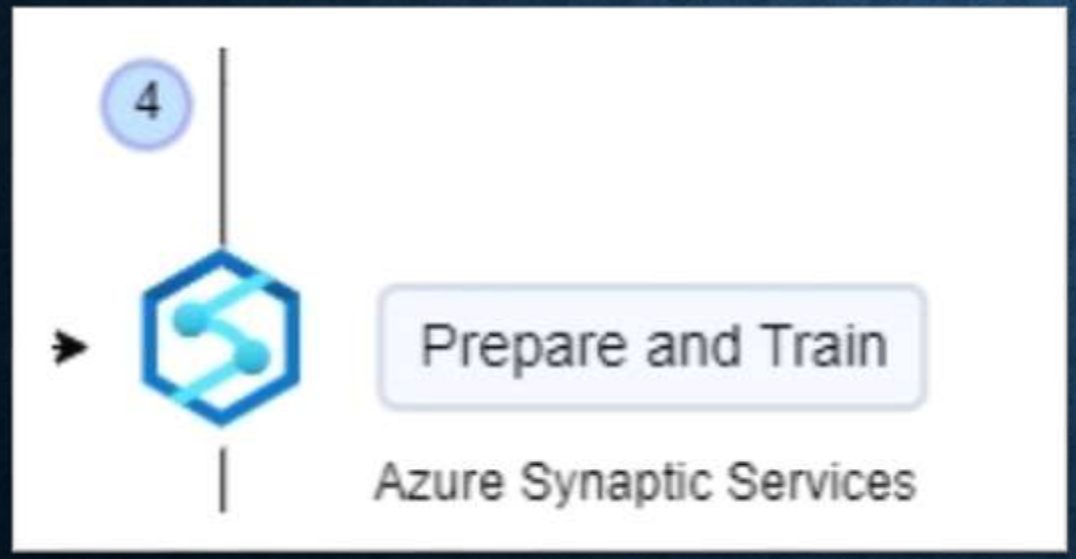



# DATA STORE

All the data which Data Factory gets are then stored in Azure Data Lake Storage.

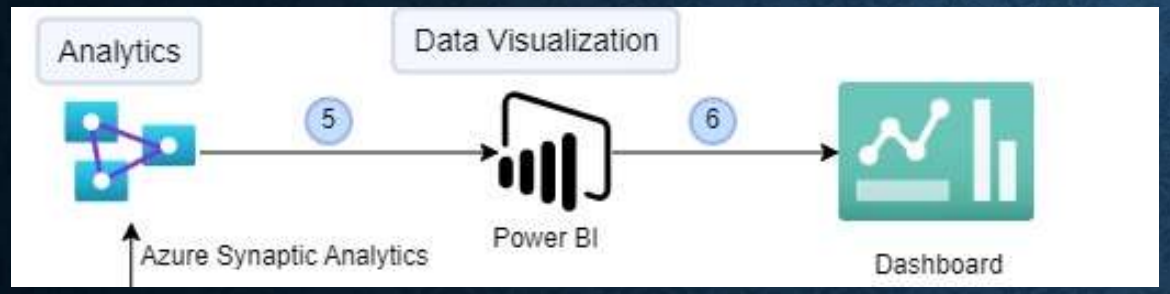
It is a centralized repository designed to store, process and secure large amount of structured, semistructured, and unstructured data. It can store data in its native format and process any variety of it, ignoring size limits.

# PROCESS




The data has been stored in the Data Lake. Azure Synapse Services is used. It helps to bring together data integration, enterprise data warehousing and big data analytics. We are able to query data and play around for prepare, transform, manage, and serve data for BI and machine learning needs.

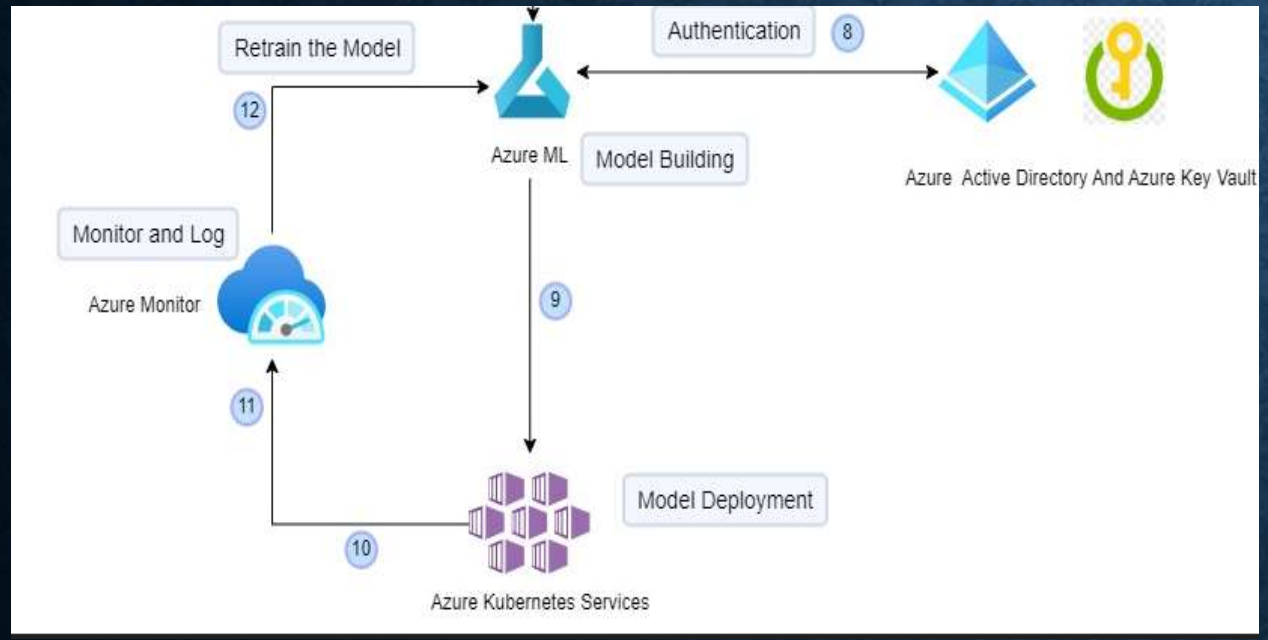
# DATA ANALYTICS




For the data analytics, the data which has been modified in Azure Synapse Services were taken and using Azure Synapse Analytics, it was further modified to get the data we want for the visualization and dashboard building.

Visualization is done on Power BI. Power BI is a free to use data visualization tool which produces dynamic dashboards for insight identification and extraction.

# BUILD AND TRAIN A MODEL




After performing necessary transformation and clean-up of the data in Azure Synapse Analytics, the data is used for training the AI/ML model.

For this, we will be using Azure Machine Learning Service. This is a low code platform. It requires just the working knowledge of ML and AI but no code. However, we need to have authorization and key for using the model. This is handled by Azure Active Directory and Azure Key Vault.

After building of model, the deployment was done on Azure Kubernetes Services. It provides us a quick way of deploying, configure and manage our model.

With the use of Azure Monitor Service, the model performance is logged. It helps to maximize the availability and performance of the application or service.

If we think it needs more training after analysis of logs and result, we can retrain the model using Azure Machine Learning Services again.

# SUMMARY

To conclude, we have made an Azure based architecture to tackle our problem. All the input data source are taken in by Azure Data Factor and is stored on Azure Data Lake. After that, for modification and data pre-processing, the Azure Synapse Services is used. Followed by that, we used Azure Synapse Analytics to perform data analytics. The result after the data analytics, were taken in by PowerBI, where we were able to make a dashboard for insight identification and extraction.

For the AI part, all the data necessary for training the model is taken from Azure Synapse Service. The trained model is deployed to Azure Kubernetes Service and is monitored through Azure Monitor. After checking the log file, we can decide if we need to retrain the model or not.

The data for future AI projects can be stored in one Azure Synapse Service and can be reused again and again as the data has been pre-processed.

# THANK YOU