



# Azure Synapse Analytics, HDInsight & Azure Databricks

L20-1116 Faiqa Adnan & L20-0921 Aisha M Nawaz



# INTRODUCTION



## Why needed ?

- ❖ Businesses operate in a data-driven environment
- ❖ Data generated at extraordinary rates
- ❖ Challenging to manage & analyze data effectively
- ❖ It provides robust, cloud-based data platform to manage, integrate & analyze data

# Azure Synapse Analytics

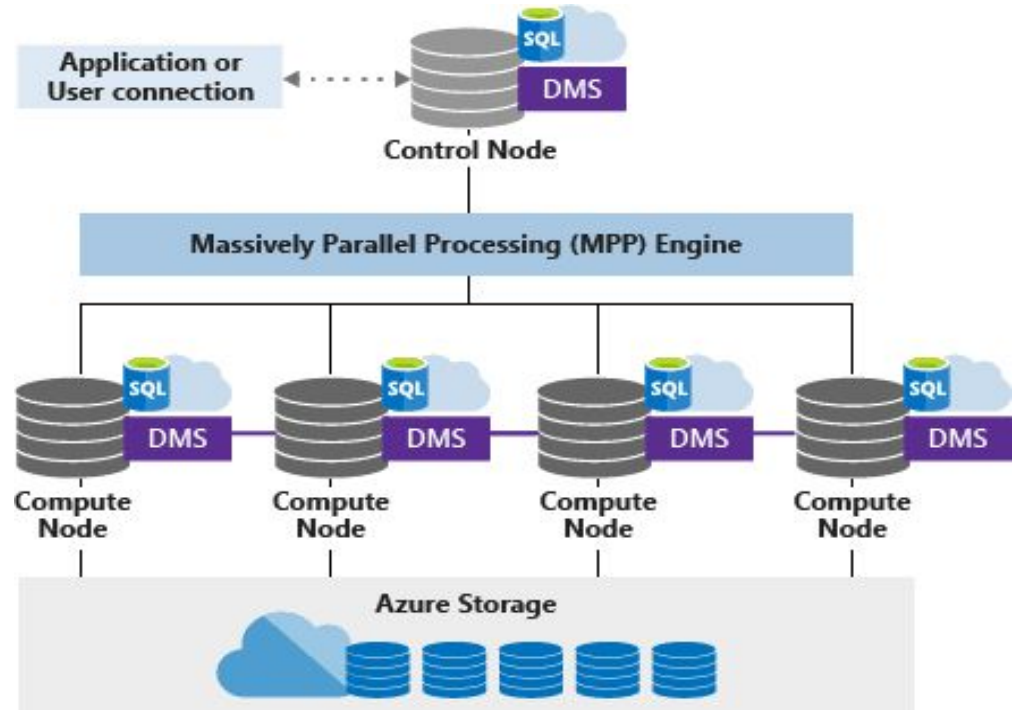
- Big data & Analytics service by Microsoft Azure
- Cloud-based , Quick & Efficient
- Big Data & Data warehousing in one service
- Business can query & analyze large volumes of data using serverless or provisioned resources
- Easily store, analyze & manage data at scale
- Uses massively parallel processing
- Like virtual assistance for data :
  - Intelligent insights & recommendations
  - Help make informed business decisions



# Azure Synapse Analytics

## Features:

- Serverless or Provisioned Resources
- Seamless Integration
- Security Features
- Data Ingestion
- Data Preparation
- Data warehousing
- Big data analytics
- Machine Learning

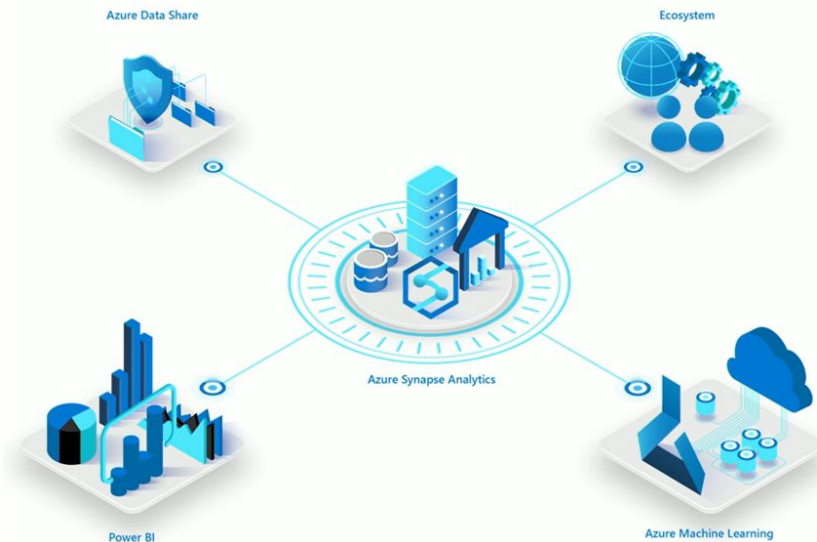


# Azure Synapse Analytics

Supports variety of Data Sources, Structured, Semi-Structured & Unstructured Data

Offers Integration with other Azure Services:

- Azure Machine Learning
- Power BI
- Azure Data Factory
- Azure Stream Analytics





# Azure Synapse Analytics

## Industries using it:

- ❖ Healthcare
- ❖ Retail
- ❖ Financial services

## Business outcomes:

1. Improved decision making
2. Increased Efficiency
3. Enhanced Security
4. Improved Customer Experience
5. Faster Time-to-Market



# Azure Synapse Analytics



## ADVANTAGES:

- Scalability
- Integration with Azure Services
- Simplified data management
- Advance Analytics capability

## DISADVANTAGES:

- Cost
- High Complexity
- Limited compatibility
- Limited customization

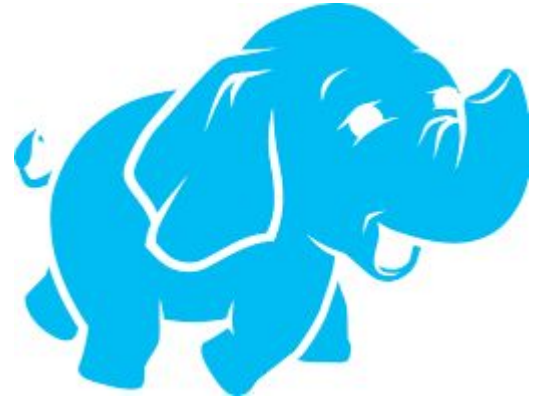


# Azure HDInsight

- ❖ Cloud-based Big Data solution
- ❖ Simplifies process of deploying, managing & scaling Apache Hadoop clusters in the cloud.
- ❖ Allows running of Big Data workloads like:
  - ❖ Batch Processing
  - ❖ Interactive Querying
  - ❖ Real-time stream processing
- ❖ Provides platform for running popular big data processing frameworks like Apache Hadoop, Spark, Hive, HBase & Storm.



# Azure HDInsight



## Features:

- Data Ingestion
  - Data Processing
  - Data Analysis
  - Data Visualization
- Supports Variety of Data Sources : Structured, Unstructured data
- Offers Integration with Azure Services: Azure Blob Storage, Azure Data Lake Storage

# Azure HDInsight



## ADVANTAGES:

- Scalability
- Cost-effective
- Integration
- Security
- Ease of use

## DISADVANTAGES:

- High Complexity
- Dependency on cloud infrastructure
- Limited control
- Data Transfer Cost



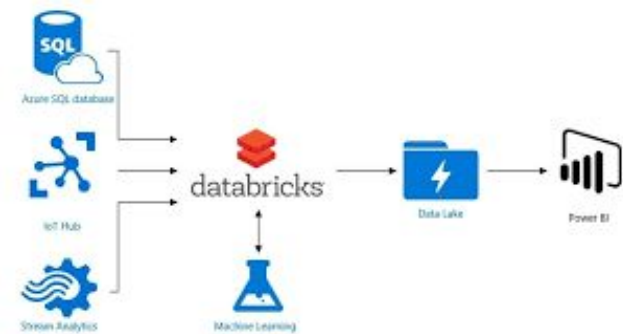
# Azure Databricks

- Fast, easy and collaborative Apache Spark-based analytics
- Allows organizations to quickly build & deploy big data solutions
- Apache Spark + Databricks + Enterprise Cloud = Azure Databricks
- Unified Analytics Platform for :
  - Data Engineering
  - Machine Learning
    - Analytics
- Enables users to collaborate & share insights in real-time

# Azure Databricks

## Features:

- Data Ingestion
  - Data Preparation
  - Data Analysis
  - Machine Learning
- ❖ Supports a variety of data sources: structures, unstructured data
  - ❖ Offers integration with Azure Services: Azure Data Factory, Azure Blob Storage & Azure Data Lake Storage





# Azure Databricks



## ADVANTAGES

- Clusters are easy to manage & configure
- Has an Azure Synapse Analytics Converter & can connect to Azure DB
- Integrated with Active Directory
- Supports Multiple Languages

## DISADVANTAGES

- Does not integrate with Git or any other versioning tool.
- Currently only supports HDInsight
- Does not support Azure Batch or Azure Distributed Data Engineering Toolkit (AZTK)



## Conclusion

- ★ 3 powerful big data & analytics services
- ★ Comprehensive suite of tools for building & managing big data solutions
- ★ For different needs & use cases
- ★ Together provides a complete solution for organizations looking to leverage power of big data to drive their business forward



THANK YOU!