MANDAPALLI HEMA

ASSESSMENT-21

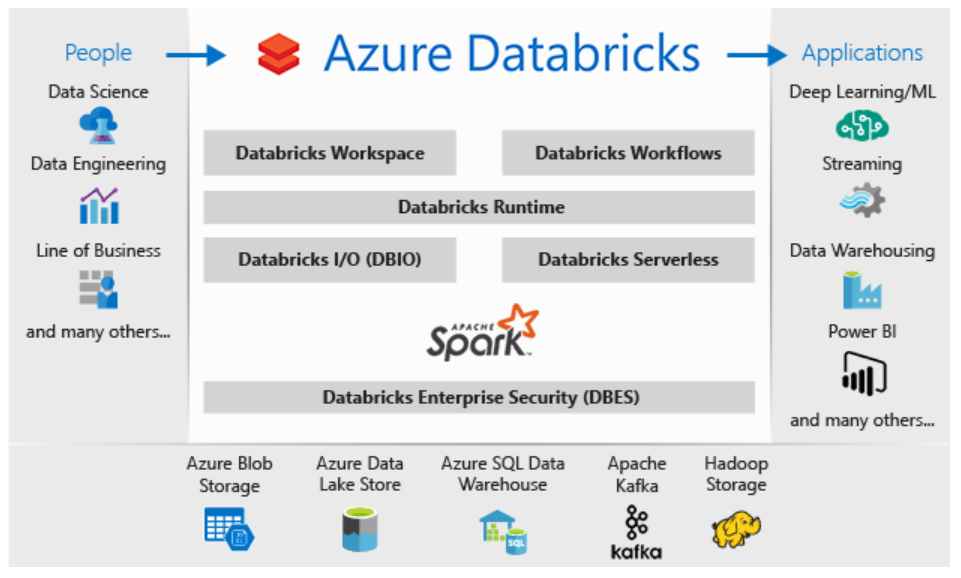
Azure Databricks operates out of a control plane and a compute plane. The control plane includes the backend services that Azure Databricks manages in your Azure Databricks account. Notebook commands and many other workspace configurations are stored in the control plane and encrypted at rest.

Databricks is deployed within the customer's account in their cloud provider, in this case Azure. Databricks is primarily composed of two layers; a Control Plane (internal) and a Data Plane (external/client).

Databricks is an industry-leading, cloud-based data engineering tool used for processing and transforming massive quantities of data and exploring the data through machine learning models. Recently added to Azure, it's the latest big data tool for the Microsoft cloud.

Why use Apache Spark on Databricks: The Databricks platform provides a secure, collaborative environment for developing and deploying enterprise solutions that scale with your business. Databricks employees include many of the world's most knowledgeable Apache Spark maintainers and users.

This platform integrates with cloud storage and security. After that Databricks manages and deploys your cloud infrastructure without your manual interference. It allows you to use it for multiple purposes such as processing, storing, cleaning, sharing, analyzing, modeling, monetize, and more with your datasets.



A workspace is an environment for accessing all of your Databricks assets. A workspace organizes objects (notebooks, libraries, dashboards, and experiments) into folders and provides access to data objects and computational resources.

Databricks Workflows is a managed orchestration service, fully integrated with the Databricks Lakehouse Platform. Workflows lets you easily define, manage and monitor multi-task workflows for ETL, analytics and machine learning pipelines.

DATA BRICKS CLUSTERS:

A cluster in Databricks is a group of virtual machines that are configured with Spark/PySpark and has a combination of computation resources and configuration on which your application can run.

Types of Clusters in Databricks:

There are mainly two types of clusters in Databricks

All-Purpose compute: Used to analyze data collaboratively using interactive notebook. You can create, terminate, and restart this compute using the UI, CLI, or REST API.

Job compute: Used to run fast and robust automated jobs. The Azure Databricks job scheduler creates a job compute when you run a job on a new compute.

Azure Databricks is a fast, easy, and collaborative Apache Spark-based analytics platform that is built on top of the Microsoft Azure cloud. A collaborative and interactive workspace allows users to perform big data processing and machine learning tasks easily.