**Azure Databricks**

**Apache Spark Introduction: -**

Apache Spark is a popular framework for large data analysis and is an open-source, rapid cluster computing system. This framework helps to improve performance by processing data in parallel. It's written in Scala, a high-level programming language that also supports Python, SQL, Java, and R APIs.

**Azure Databricks: -**

Databricks is a Microsoft Azure implementation of Apache Spark. Spark clusters, which are completely managed, are used to process big data workloads and aid in data engineering, data exploration, and data visualization utilizing machine learning.

A few very important features of the Azure Databrick are –

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

**Databricks File System (DBFS):** - The Databricks File System (DBFS) is a distributed file system mounted into an Azure Databricks workspace and available on Azure Databricks clusters. DBFS is an abstraction on top of scalable object storage that maps Unix-like filesystem calls to native cloud storage API calls.

**Databricks Notebook: -**

Notebooks in Databricks are used to write spark code to process and transform data. Notebooks support Python, Scala, SQL, and R languages. Whenever we execute a notebook in Databricks, it attaches a cluster (computation resource) to it and creates an execution context.

**Databricks Clusters: -**

An Azure Databricks cluster is a set of computation resources and configurations on which you run data engineering, data science, and data analytics workloads, such as production ETL pipelines, streaming analytics, ad-hoc analytics, and machine learning.