Explore azure data bricks and its studio. List down the components created during the Databricks

**Azure Databricks**

A databricks workspace is managed cloud-based analytics platform that provides an environment for big data and AI/ML workloads.

The workspace facilitates collaboration by enabling to run notebooks, jobs and SQL Queries efficiently.

When a databricks is created, the following backend resources are automatically created.

Compute resources: Clusters for executing workloads.

Storage resources: Databricks File system backed by Azure Blob storage.

Networking Components: Virtual networks (VNets) to securely connect resources.

**UI Features**

**Clusters:** Provides managed spark clusters to run analytics

**SQL Warehouses** – Allows execution of SQL queries on structured data.

**Serverless SQL**: Optimized for running SQL queries without managing infrastructure

**Resource deletion**: Deleting the databricks workspace will also delete all associated resources, including clusters and storage.

**Compute options in Databricks**

**Types:**

**All-purpose compute**

Designed for interactive workloads such as developing and running notebooks.

All purpose compute used for notebook execution and job execution.

Supports multiple concurrent users working on the same cluster.

**Job Compute**

Optimized for running scheduled jobs.

Runs in an isolated environment, preventing interference with interactive users.

Types of Computes:

1. Create with personal compute

2. Create with Power User compute

3. Create with Legacy Shared compute

**Cluster Autoscaling**

Databricks allows clusters to dynamically scale based on workload.

It automatically increases or decreases the number of worker nodes as needed, optimizing performance and reducing costs.

**Cluster Size**

Databricks supports up to 1,00,000 nodes, but most practical workloads requires significant fewer machines.

**Inactivity timeout:**

Clusters that are not actively used can be automatically terminated to avoid unnecessary costs.

Single node clusters

Provides a cost-effective option for small workloads or testing.

Reduces expenses by using a single machine instead of a distributed cluster.

**Databricks units (DBU) and access modes.**

DBU is a consumption-based pricing metric.

DBU bills are based on:

1.Type of compute used: Standard vs Premium clusters

Duration of Usage: Compute time is measured in DBUs per hour.

Access Modes:

1.Shared mode

Allows execution of python and SQL.

Multiple users can share the same cluster.

Single-User Mode

Allows execution of multiple programming languages (Python, Scala, SQL, R)

Provides better performance isolation for individual users.