

04 Databricks Architecture

What is Databricks?

Databricks is a cloud-based platform built on Apache Spark that provides a collaborative environment for big data processing and analytics. It offers an integrated workspace where data engineers, data scientists, and analysts can work together to leverage the power of Spark for various use cases.

Databricks is important because it makes it easier to use a Apache Spark. Instead of having to worry about all the technical stuff behind the scenes, Databricks gives you a simple and friendly way to use Spark. It takes care of all the complicated setup and management stuff so that you can focus on working with your data and doing cool analytics tasks. It's like having a magic helper that takes care of the boring stuff, so you can have more fun exploring and analyzing your data.

Databricks Architecture

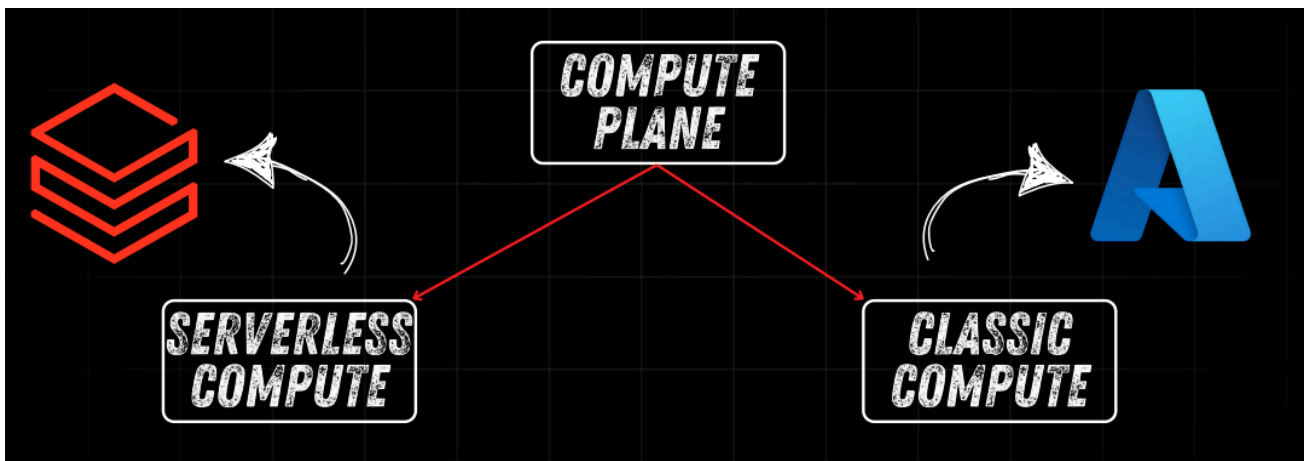
1. Control Plane
2. Compute Plane

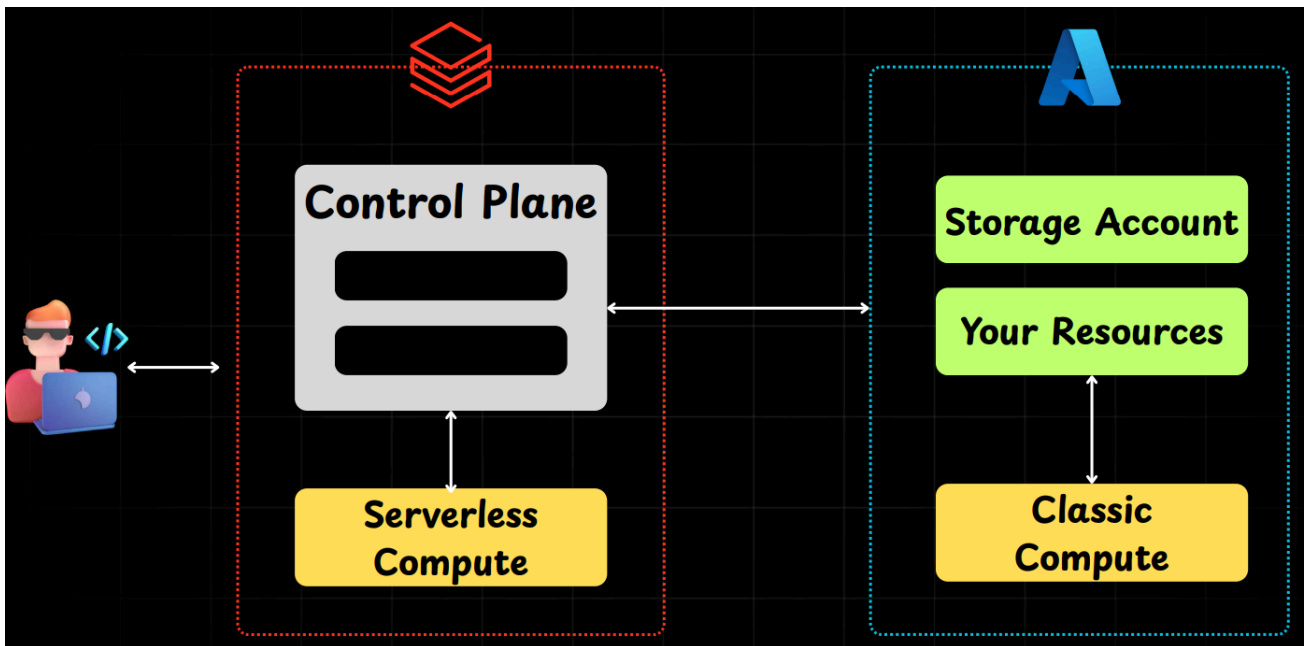
Control Plane

The control plane includes the backend services that Azure Databricks manages in your Azure Databricks account. The web application is in the control plane.

Compute Plane

The compute plane is where your data is processed.





<https://learn.microsoft.com/en-us/azure/databricks/getting-started/overview>

High-level architecture

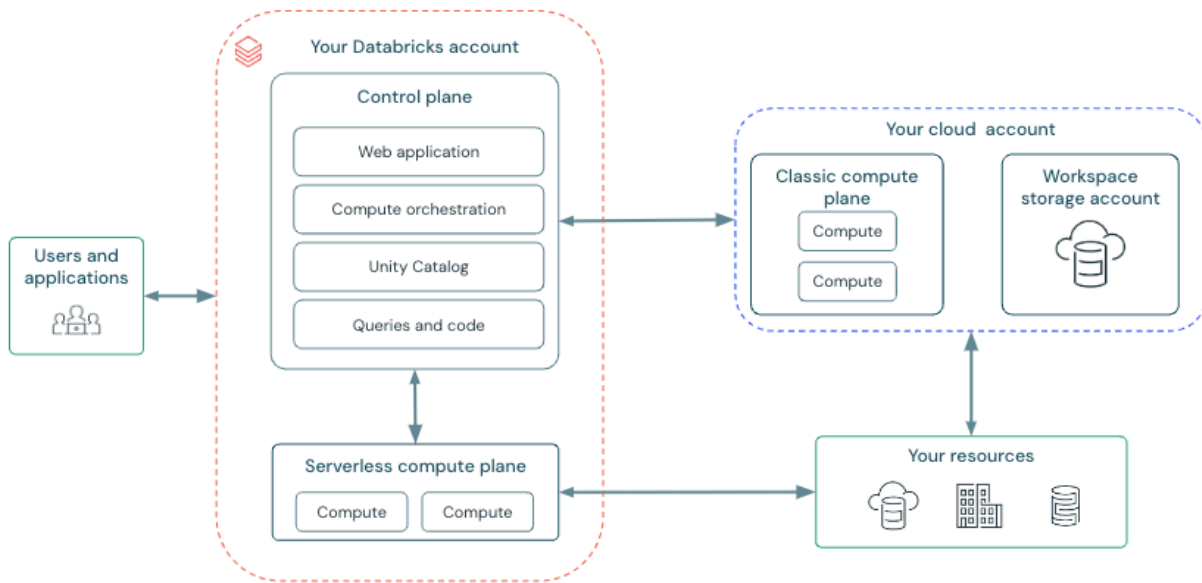
Azure Databricks operates out of a **control plane** and a **compute plane**.

- The **control plane** includes the backend services that Azure Databricks manages in your Azure Databricks account. The web application is in the control plane.
- The **compute plane** is where your data is processed. There are two types of compute planes depending on the compute that you are using.
 - For serverless compute, the serverless compute resources run in a *serverless compute plane* in your Azure Databricks account.
 - For classic Azure Databricks compute, the compute resources are in your Azure subscription in what is called the *classic compute plane*. This refers to the network in your Azure subscription and its resources.

To learn more about classic compute and serverless compute, see [Compute](#).

Each Azure Databricks workspace has an associated storage account known as the **workspace storage account**. The workspace storage account is in your Azure subscription.

The following diagram describes the overall Azure Databricks architecture.



Serverless compute plane

In the serverless compute plane, Azure Databricks compute resources run in a compute layer within your Azure Databricks account. Azure Databricks creates a serverless compute plane in the same Azure region as your workspace's classic compute plane. You select this region when creating a workspace.

To protect customer data within the serverless compute plane, serverless compute runs within a network boundary for the workspace, with various layers of security to isolate different Azure Databricks customer workspaces and additional network controls between clusters of the same customer.

To learn more about networking in the serverless compute plane, [Serverless compute plane networking](#).

Classic compute plane

In the classic compute plane, Azure Databricks compute resources run in your Azure subscription. New compute resources are created within each workspace's virtual network in the customer's Azure subscription.

A classic compute plane has natural isolation because it runs in each customer's own Azure subscription. To learn more about networking in the classic compute plane, see [Classic compute plane networking](#).

Workspace storage account

When you create a workspace, Azure Databricks creates a account in your Azure subscription to use as the workspace storage account.

The workspace storage account contains:

- **Workspace system data:** Workspace system data is generated as you use various Azure Databricks features such as creating notebooks. This bucket includes notebook revisions, job run details, command results, and Spark logs

- **DBFS:** DBFS (Databricks File System) is a distributed file system in Azure Databricks environments accessible under the `dbfs:/` namespace. DBFS root and DBFS mounts are both in the `dbfs:/` namespace. Storing and accessing data using DBFS root or DBFS mounts is a deprecated pattern and not recommended by Databricks. For more information, see [What is DBFS?](#).
 - **Unity Catalog workspace catalog:** If your workspace was enabled for Unity Catalog automatically, the workspace storage account contains the default workspace catalog. All users in your workspace can create assets in the default schema in this catalog. See [Get started with Unity Catalog](#).
-

