

How to Create a Cluster Databricks

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. In a simple way, the cluster executes all databricks code. Workloads that the Databricks cluster can run are ETL pipelines, Machine Learning models, Streaming, Batch analytics, and ad-hoc analytics.

Databricks job scheduler creates these clusters when running on a new job cluster. These are mainly used for running fast and robust automated tasks. They are created to run a job on the new Job Cluster and terminate the Cluster once the job ends. These clusters cannot be restarted.

1. Hover over the **Sidebar** to expand assets.



Figure 1: Hover on the Sidebar in Databricks

2. Select **Create Compute**.

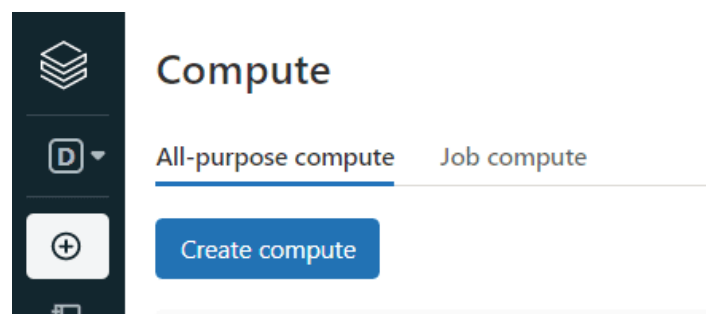


Figure 2: Select Create Compute in Databricks



3. Select **Create a Cluster**.

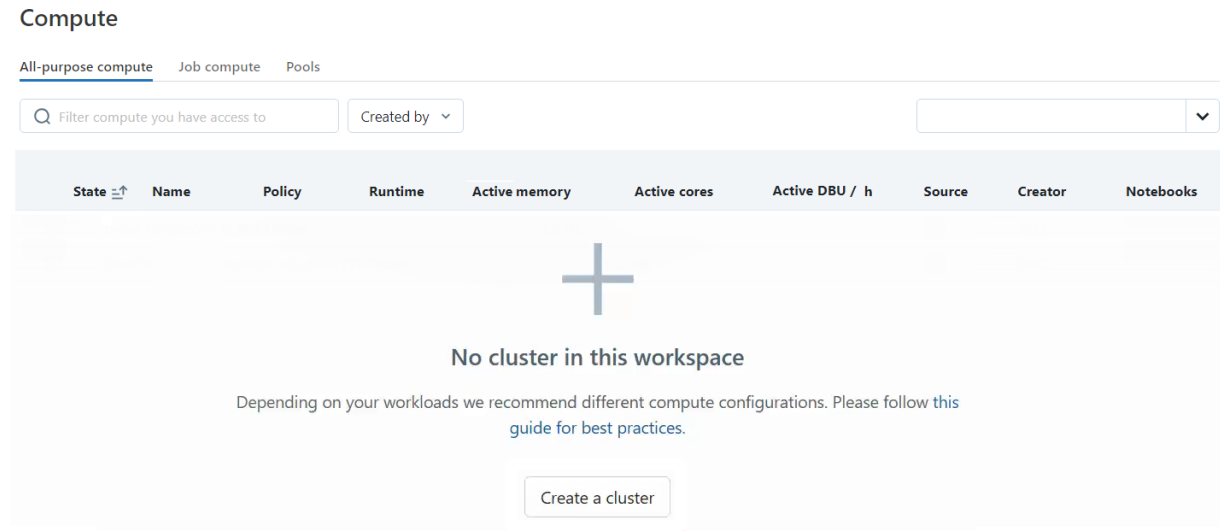


Figure 3: Create a Cluster in Databricks

4. Select **Create Cluster**.

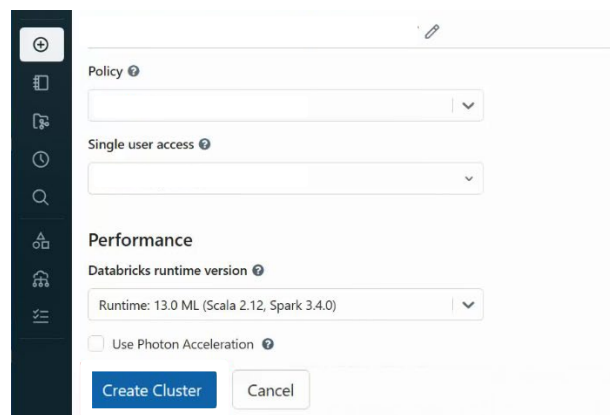


Figure 4: Create Cluster in Databricks

Note: Cluster creation takes 5 minutes.