Azure Databricks Cluster Management

Template Version: 2.0

Introduction

During this lab, you will learn how to manage Azure Databricks cluster.

Estimated Time

20 minutes

Objectives

At the end of this lab, you will be able to:

- Manage Azure Databricks clusters.

Table of Contents

Lab: Cluster Management

During this lab, you will learn cluster management.

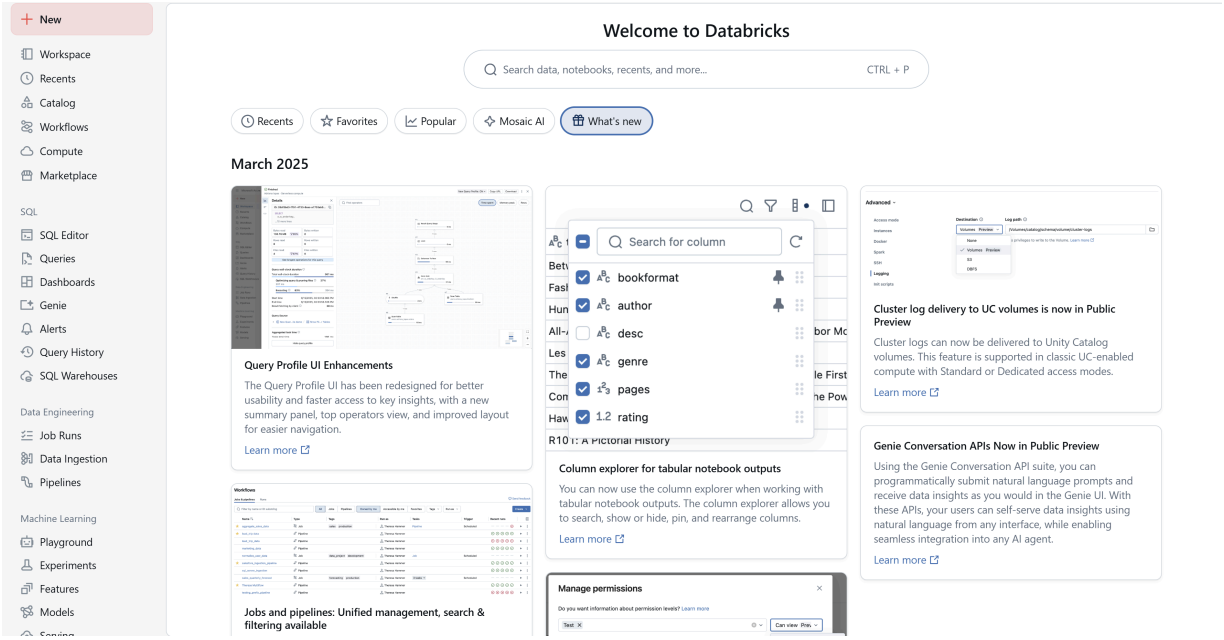
Exercise 1: Azure Databricks Cluster Management

Tasks: Cluster Management Options

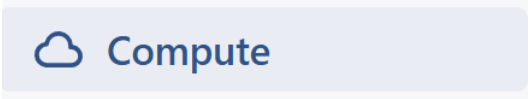
1. Orient yourself to the Databricks UI

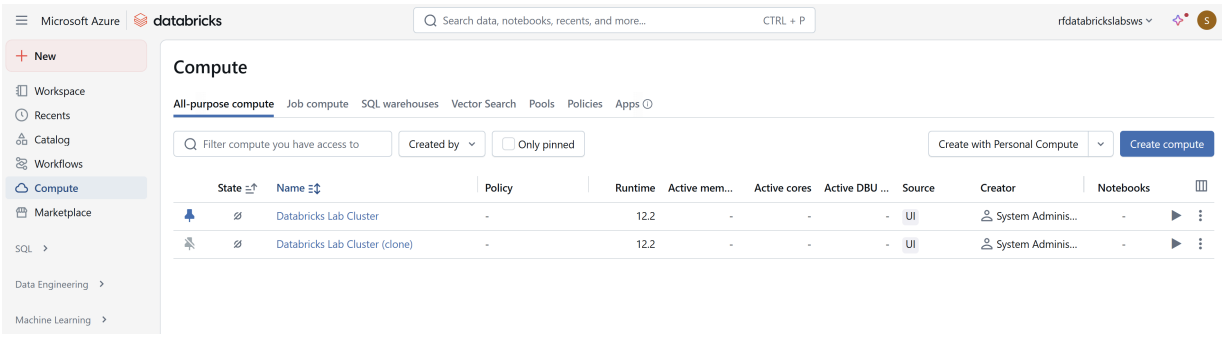
   From the sidebar at the left and the **Get Started** list on the home page, you access fundamental Azure Databricks entities:Workspace, Repos, Data , Compute and Workflows. The Workspace is the special root folder that stores your Azure Databricks assets, such as notebooks and libraries, and the data that you import.

   You can change between different Azure Databricks environments, Data Science, Databricks SQL and Databricks Machine Learning, which will show you new options related to these environments.



2. In this section, it is assumed that you've already created an Azure Databricks cluster as part of previous demo/lab.

a. To display the clusters in your workspace, click the Compute icon in the sidebar. 



b. The Clusters page displays two lists: All-Purpose Clusters and Job Clusters.

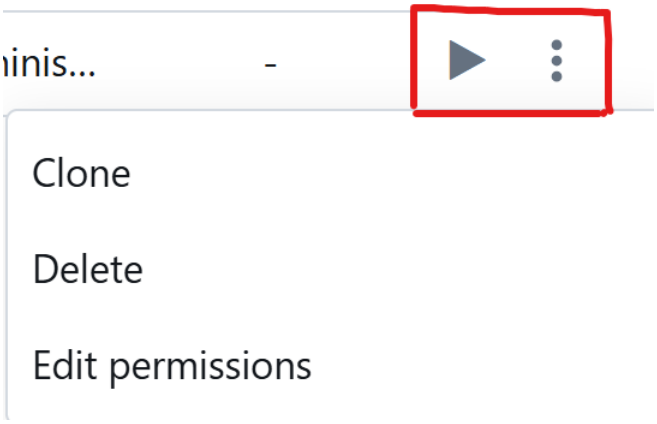
c. In addition to the common cluster information, the All-Purpose Clusters list shows the numbers of notebooks attached to the cluster.

d. In this demo Job cluster is currently not configured.

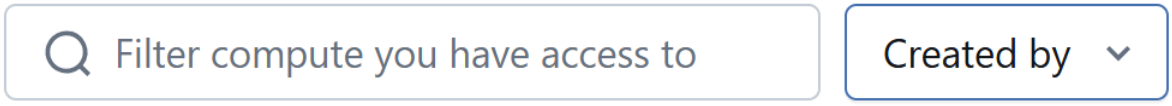
e. An icon to the left of the cluster name indicates whether the cluster is pinned, whether the cluster offers a high concurrency cluster, and whether table access control is enabled:

- Pinned
- Starting
- Terminating
- Standard cluster
  - Running
  - Terminated
- High concurrency cluster
  - Running
  - Terminated
- Access Denied
  - Running
  - Terminated
- Table ACLs enabled
  - Running
  - Terminated

f. Links and buttons at the far right provide access to the start/terminate, restart, clone, permissions, and delete actions.



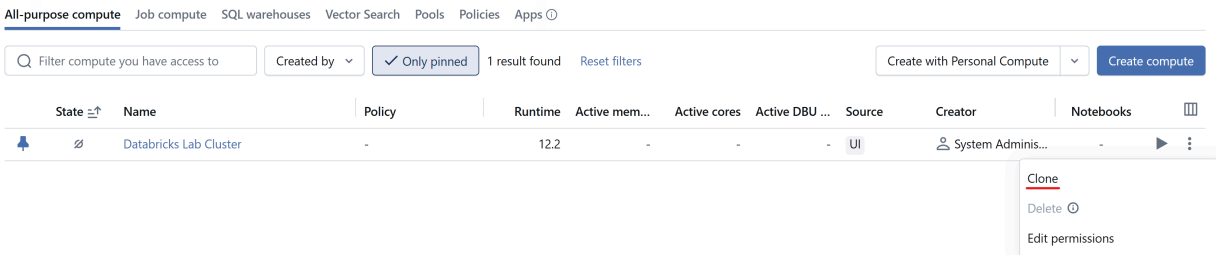
g. You can filter the cluster lists using the buttons and Filter field at the top right:



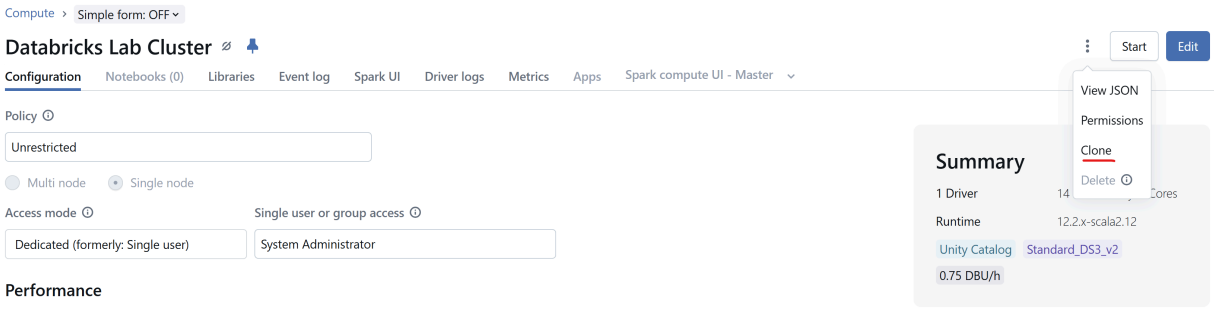
- To display only clusters that you created, click Created by me.
- To display only clusters that are accessible to you created by another user click Created by and browse to user who created.
- To filter by a string that appears in any field, type the string in the Filter text box.

3. Clone a Cluster

You can create a new cluster by cloning an existing cluster. Go to the **cluster list** and click on **Clone** button on the far-right corner.



Another option is to click on the cluster go to the **cluster details page**. On Configuration tab, click on **Clone** button:

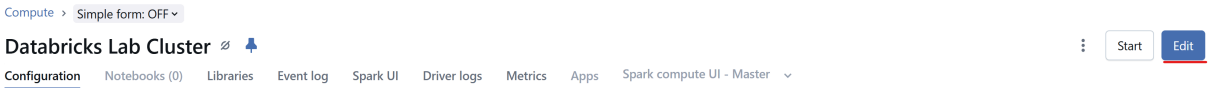


The cluster creation form is opened pre-filled with the cluster configuration. The following attributes from the existing cluster are not included in the clone:

- Cluster permissions
- Installed libraries
- Attached notebooks

4. Editing a Cluster

You can edit a cluster configuration from the cluster detail page.



- Libraries, notebooks, and jobs that were attached to the cluster remain attached after editing.
- If you edit any attribute of a running cluster (except for the cluster size and permissions), you must restart it. This can disrupt users who are currently using the cluster.
- You can edit only running or terminated clusters. You can, however, update permissions for clusters that are not in those states on the cluster details page.

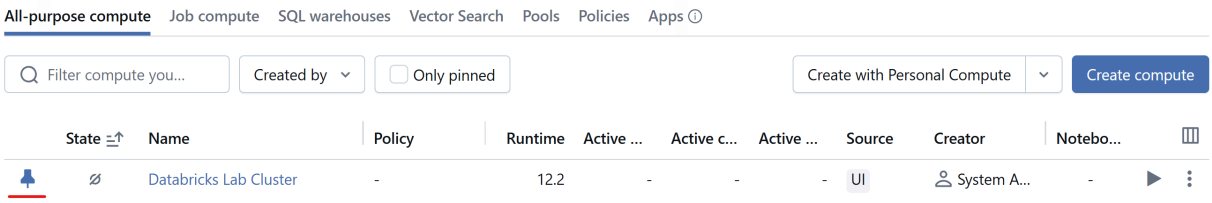
5. Pinning a Cluster

TTo keep an All-Purpose Cluster interactive cluster configuration even after a cluster has been terminated for more than 30 days, an administrator can pin the cluster. Up to 20 clusters can be pinned.

You can pin a cluster from the:

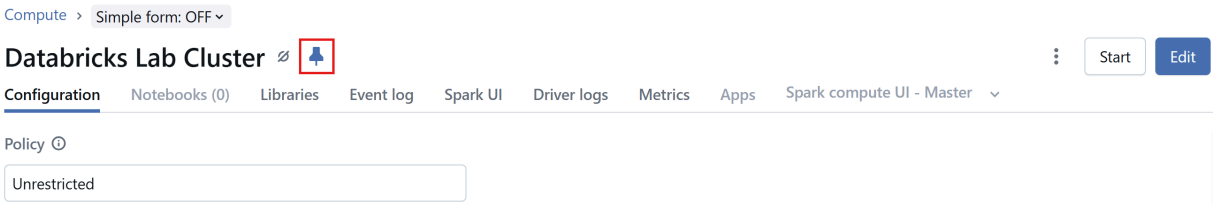
a. Cluster list

To pin or unpin a cluster, click the pin icon to the left of the cluster name.



b. Cluster detail page

To pin or unpin a cluster, click the pin icon to the right of the cluster name.



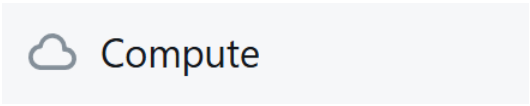
Exercise 1 has been completed

Exercise 2: Azure Databricks Cluster Event Logs

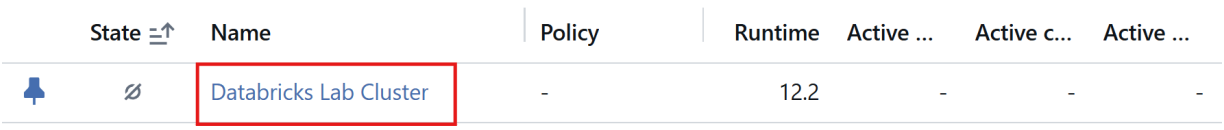
Tasks: Managing Cluster Event Logs

The cluster event log displays important cluster life cycle events that you trigger manually or are automatically triggered by Azure Databricks. Such events affect the operation of a cluster as a whole and the jobs running in the cluster.

1. Click the **Compute** icon in the sidebar.

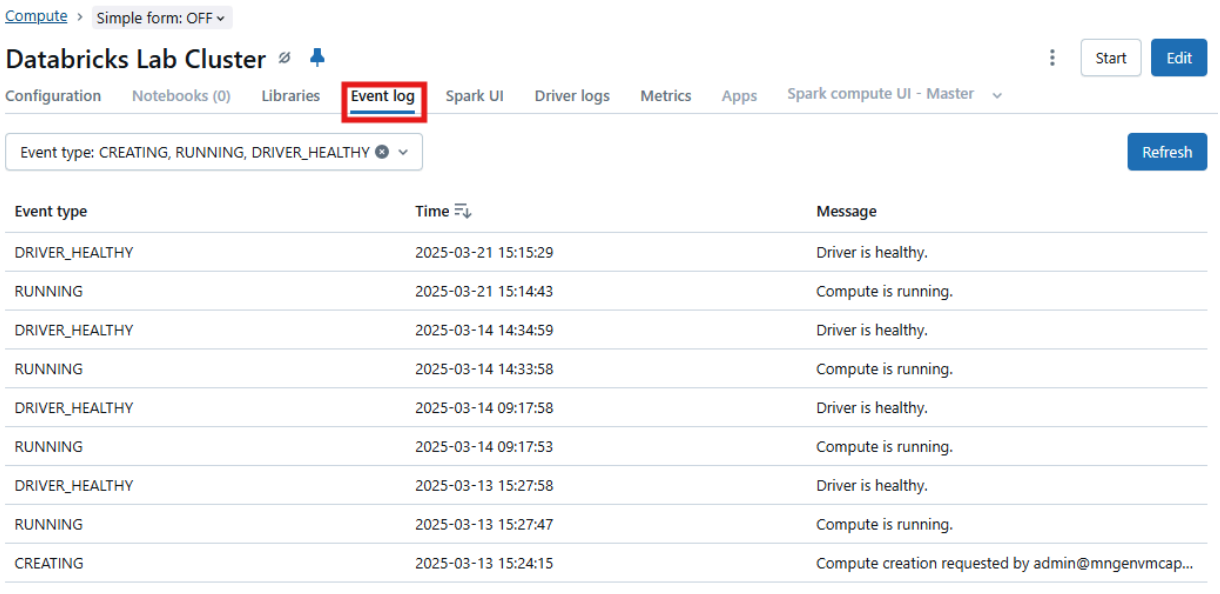


2. Click **Cluster Name** to go to cluster details page



3. Click Event Log tab

For more information about an event, click its row in the log and then click the JSON tab for details.



NOTE: Once you complete this lab, you must terminate your cluster to save cost. Do not delete the cluster as we'll use the same cluster in subsequent labs.

4. To filter the events, click the drop down button in the **Filter by Event Type** field and select one or more event type checkboxes.

5. For more information about an event, click its row in the log and then click the JSON tab for details.

NOTE: Once you complete this lab, you must terminate your cluster to save cost. Do not delete the cluster as we'll use the same cluster in subsequent labs.

Exercise 2 has been completed