SETTING UP SPARK DATABRICKS COMMUNITY EDITION

Ramakrishna Addanki
Data Engineering Lead

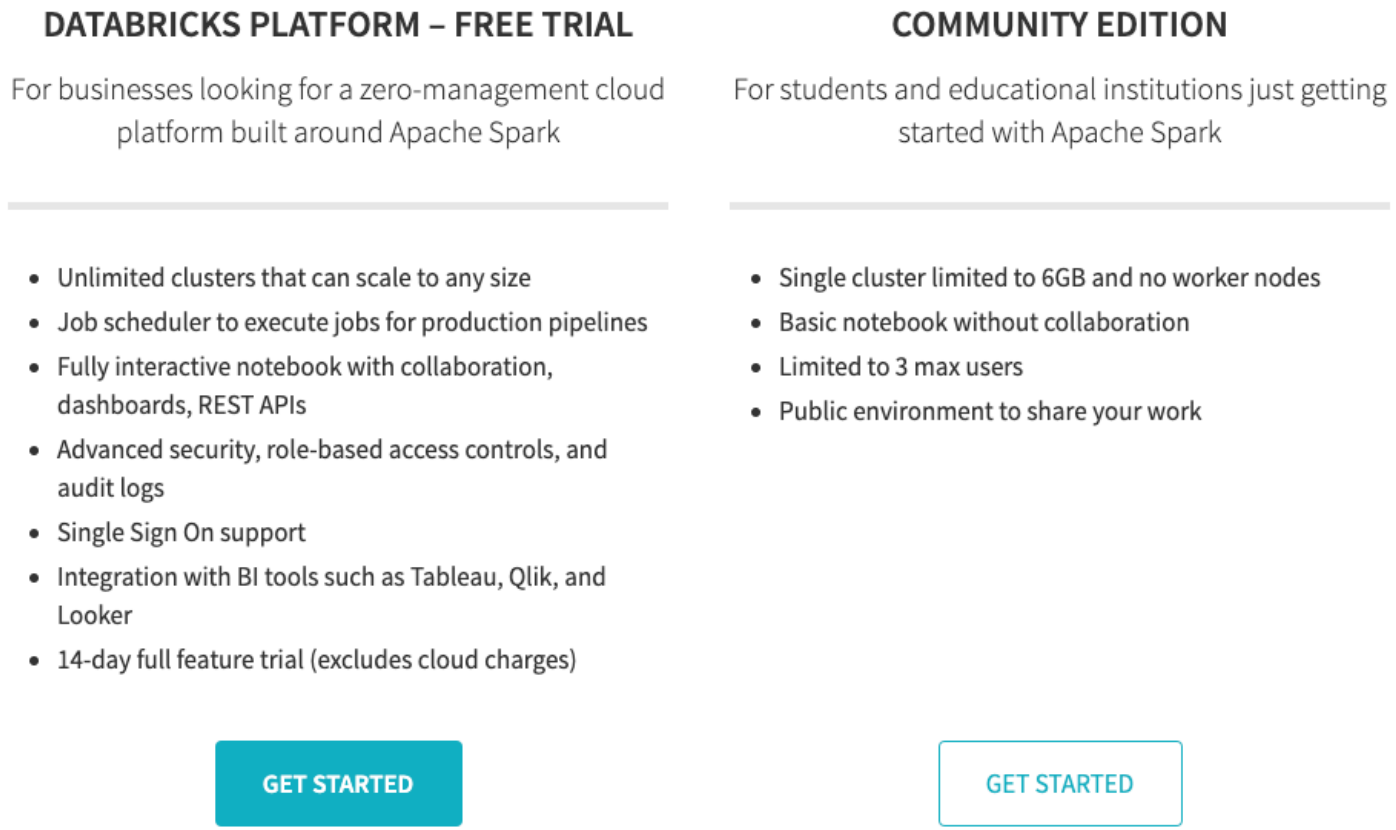
AGENDA

1) Setting up databricks cluster for session.
2) Understanding Databricks Runtime.
3) Creating Cluster
4) Spark UI, LOGS & METRICS
5) Installing Libraries
6) Creating Notebooks
7) Executing code in notebook
8) Git / bitbucket Integration

Step 1: Launch the sign up wizard and select a subscription type

Go to Databricks and click **Try Databricks** at the top right.

Step 2: Click **GET STARTED** to choose Community Edition.



SETTING UP SPARK CLUSTER

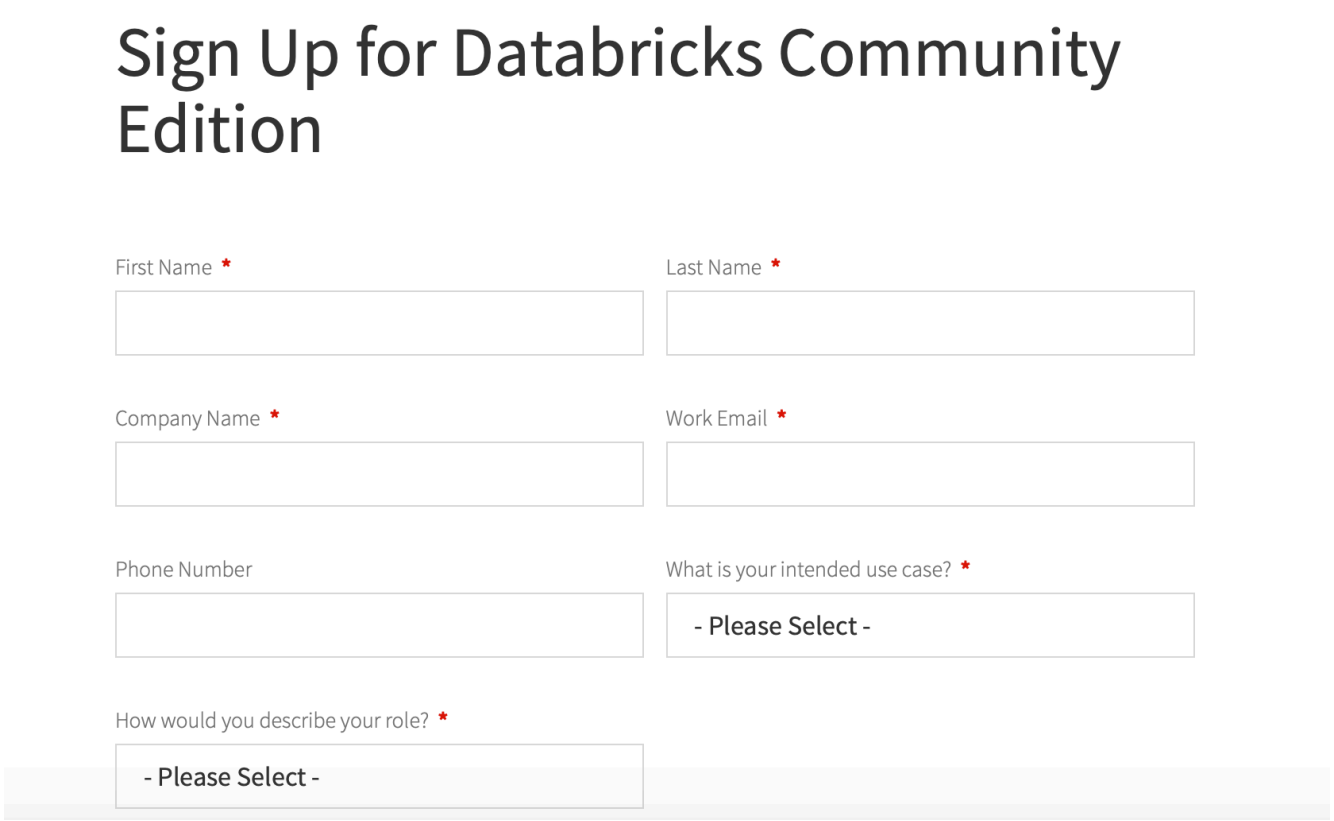
Step 3:When you select Community Edition you'll see a registration form.

Fill in the registration form.

Click **Sign Up**.

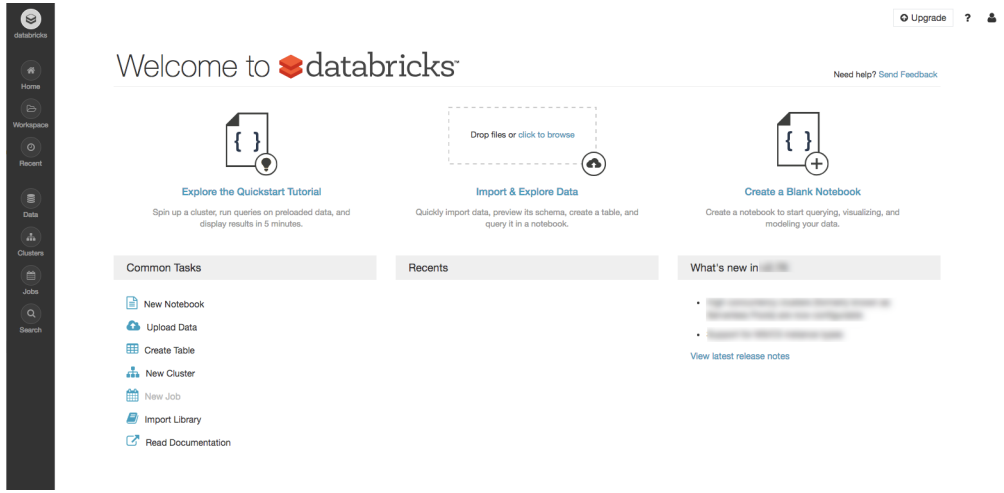
Read the Terms of Service and click **Agree**.

Step 4: When you receive the "Welcome to Databricks" email, click the link to verify your mail address.



SETTING UP SPARK CLUSTER

Step 5: Log into Databricks using the credentials you supplied when you registered. You'll see the Databricks Community Edition home page.



SETTING UP SPARK CLUSTER

SETTING UP SPARK CLUSTER

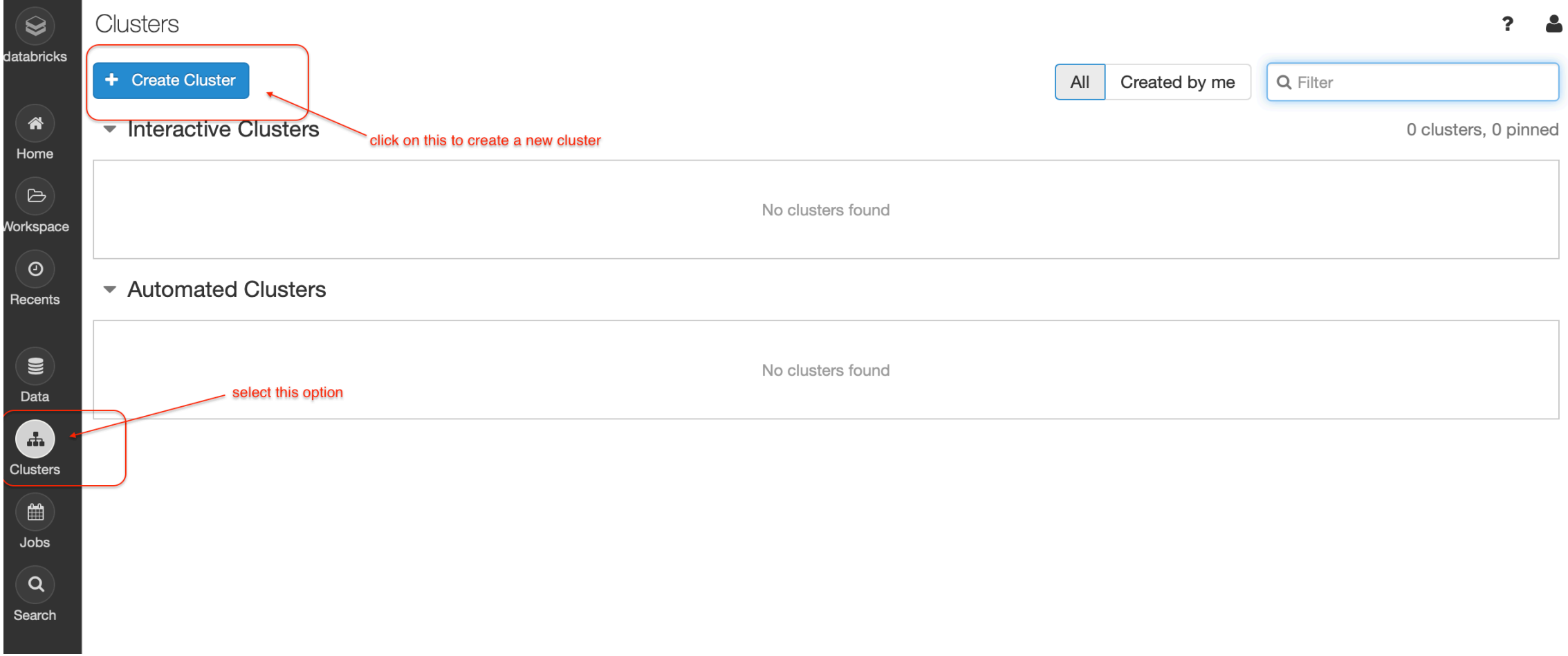
1) Databricks Community Edition is fully resourced;

2) you are not required to supply any compute or storage resources.

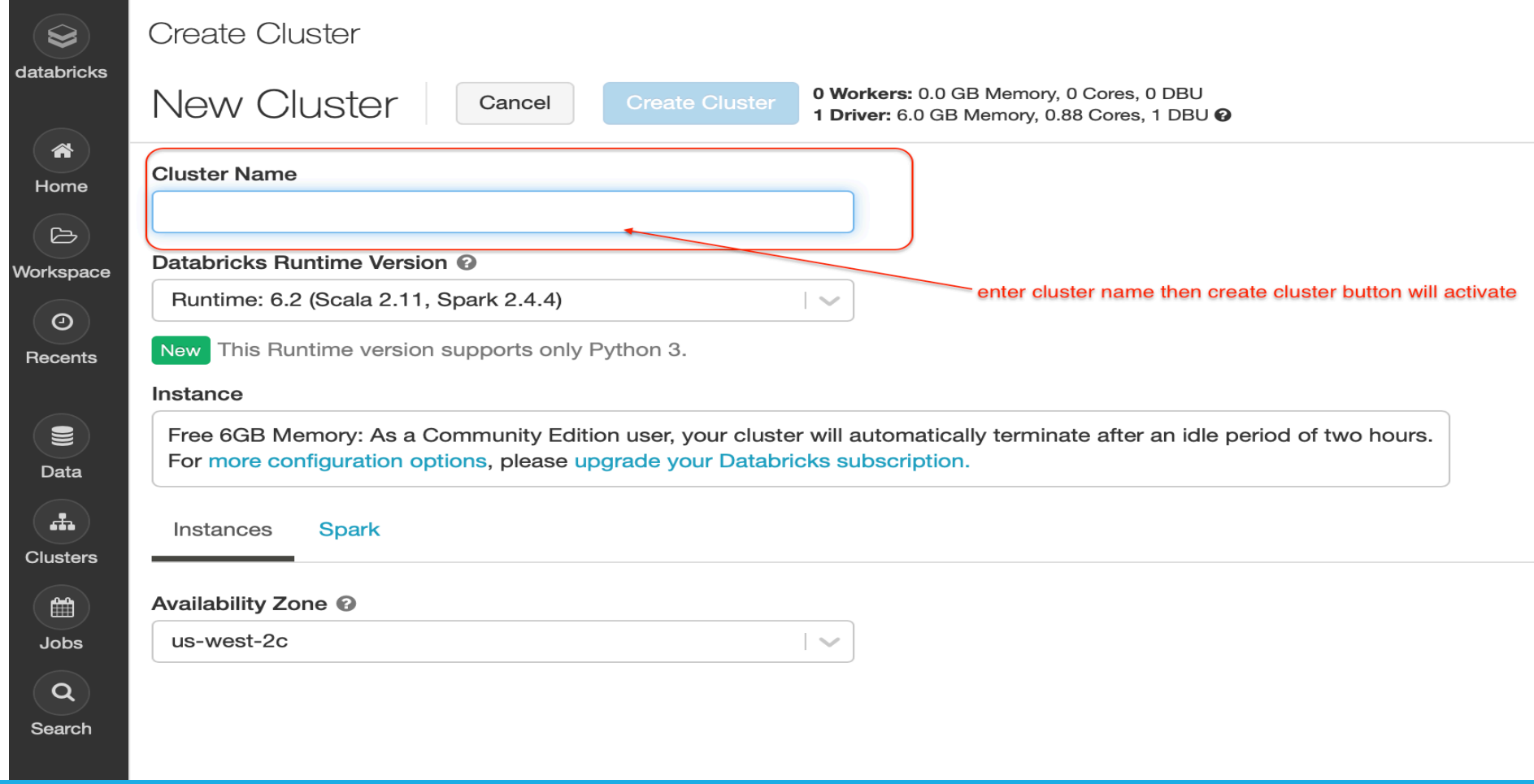
3) However, several features available in the Databricks Platform Free Trial, such as the REST API, are not available in Databricks Community Edition.

For more info refer:

https://databricks.com/product/faq/community-edition



CREATING CLUSTER



CREATING CLUSTER

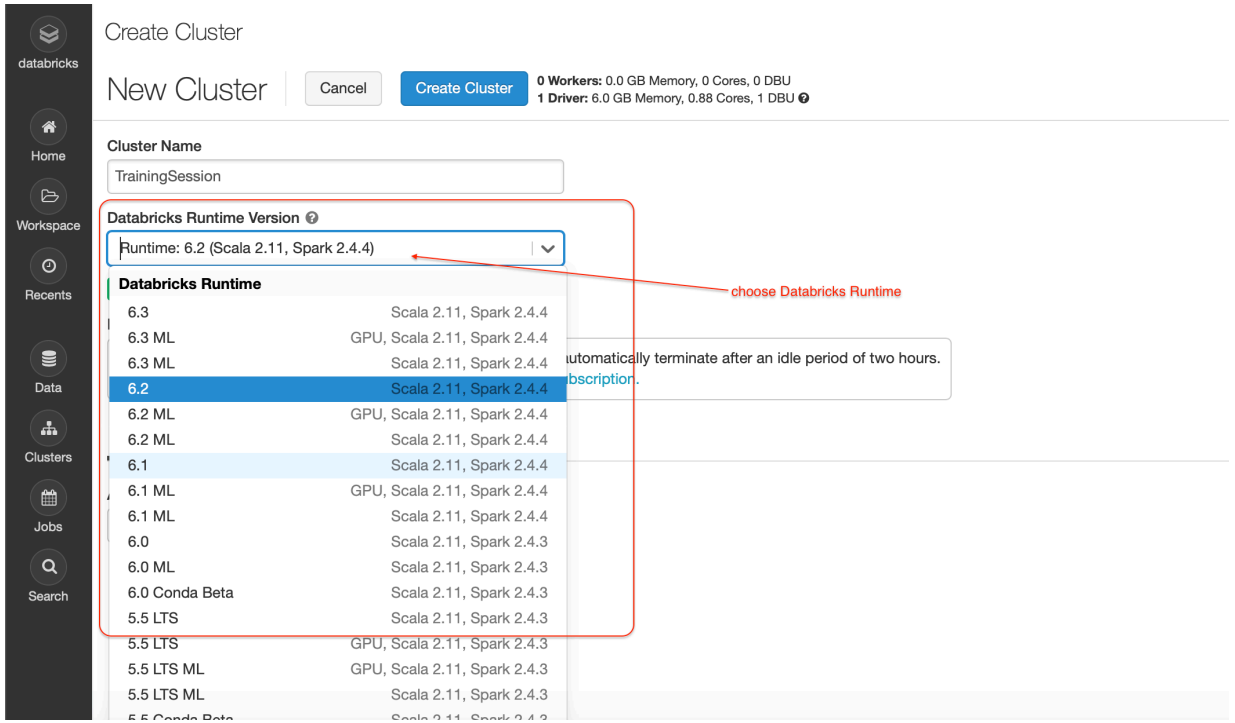
Databricks Runtime includes Apache Spark but also adds a number of components and updates that substantially improve the usability, performance, and security of big data.

Installed Java, Scala, Python, and R libraries.

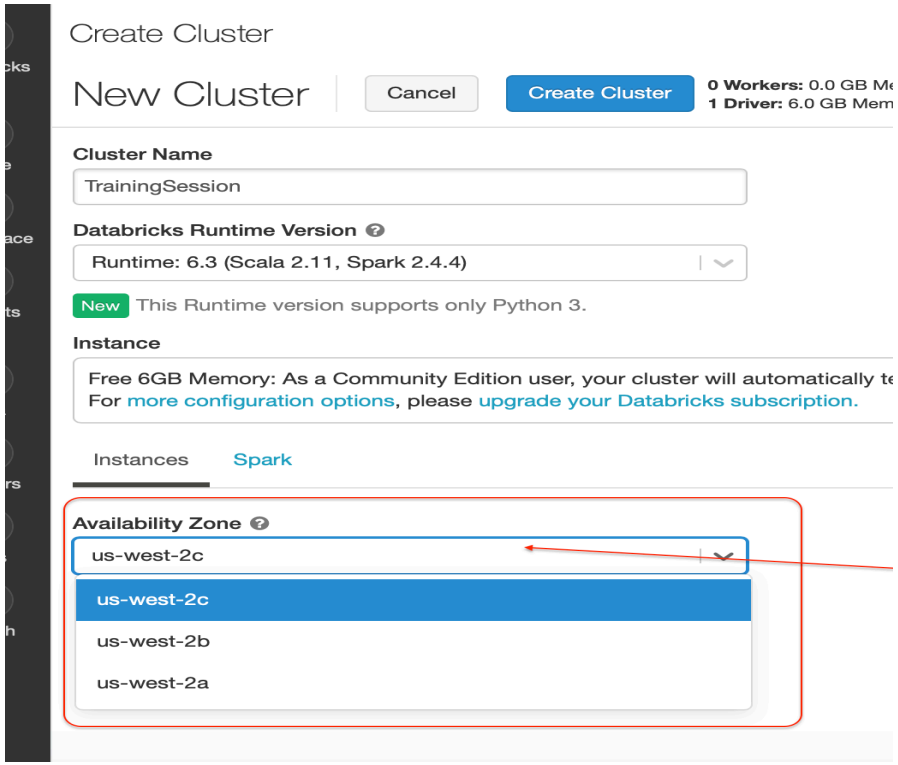
Ubuntu and its accompanying system libraries

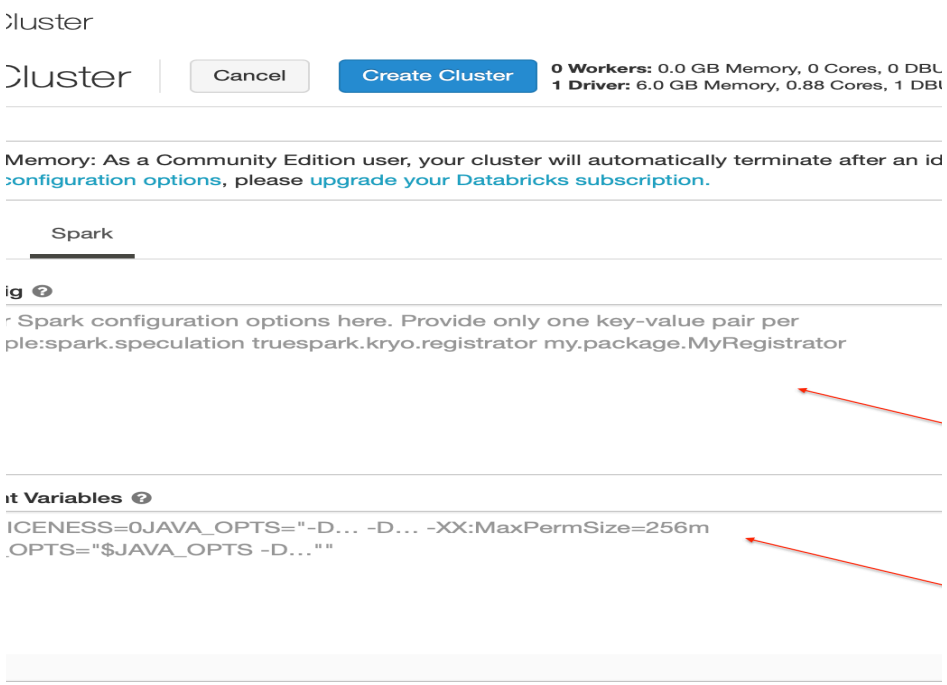
GPU libraries for GPU-enabled clusters.

Databricks services that integrate with other components of the platform, such as notebooks, jobs, and cluster manager

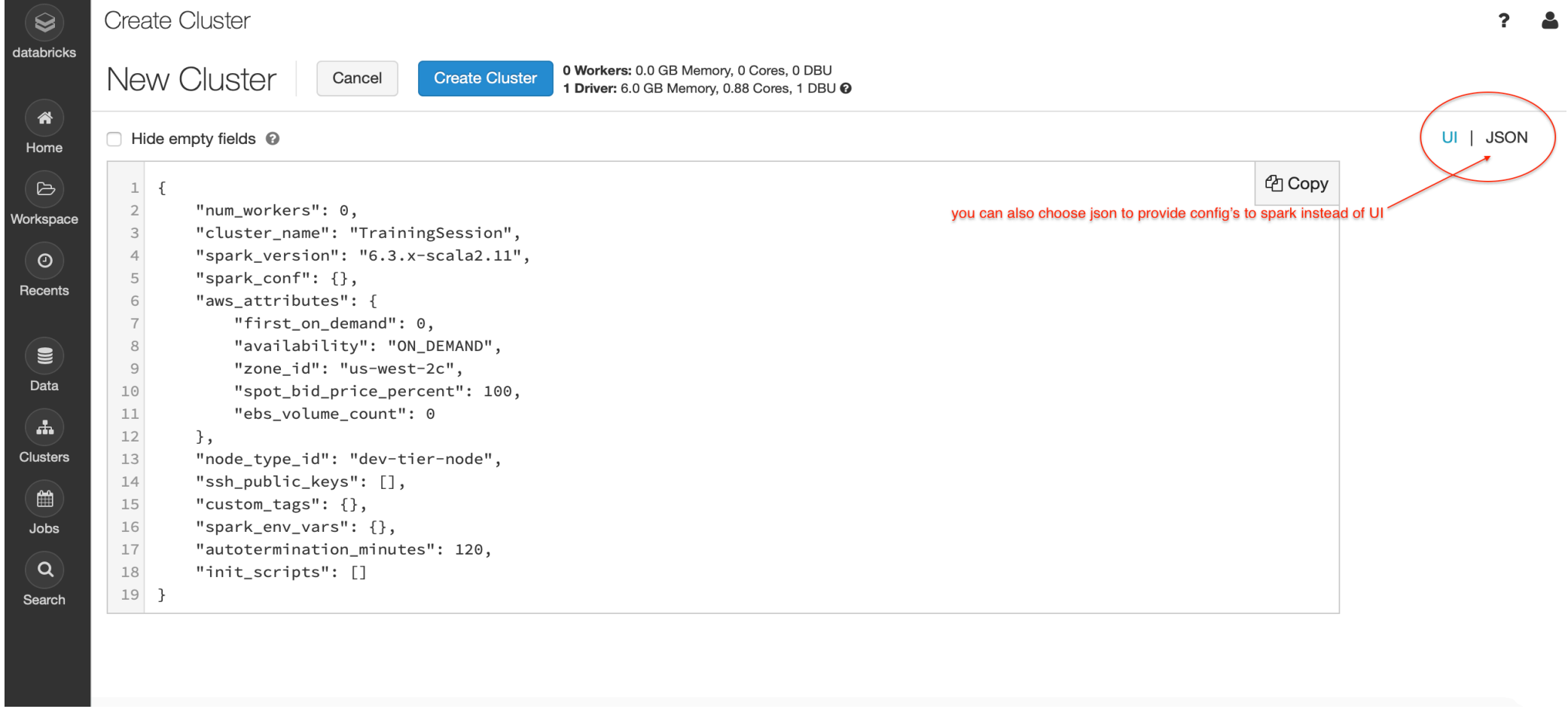


CREATING CLUSTER

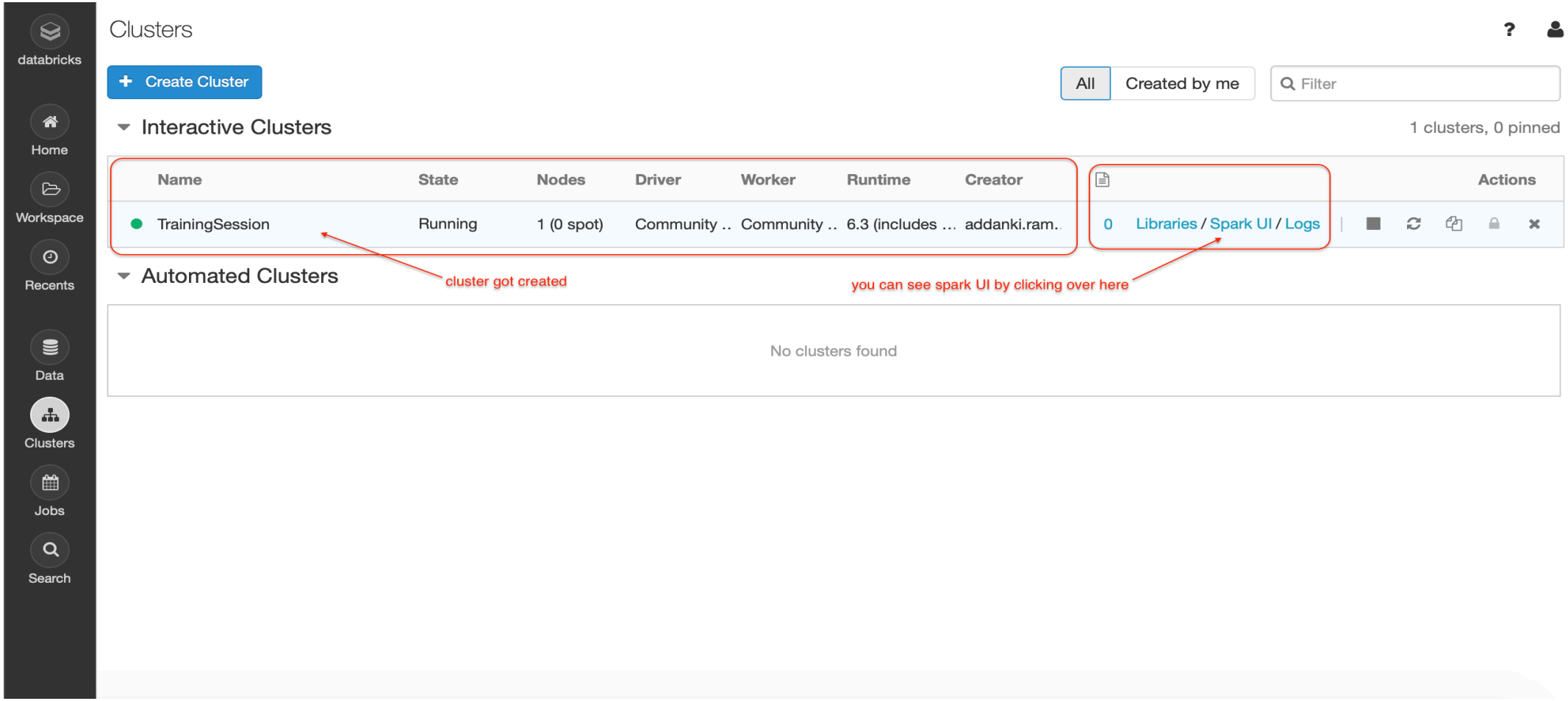




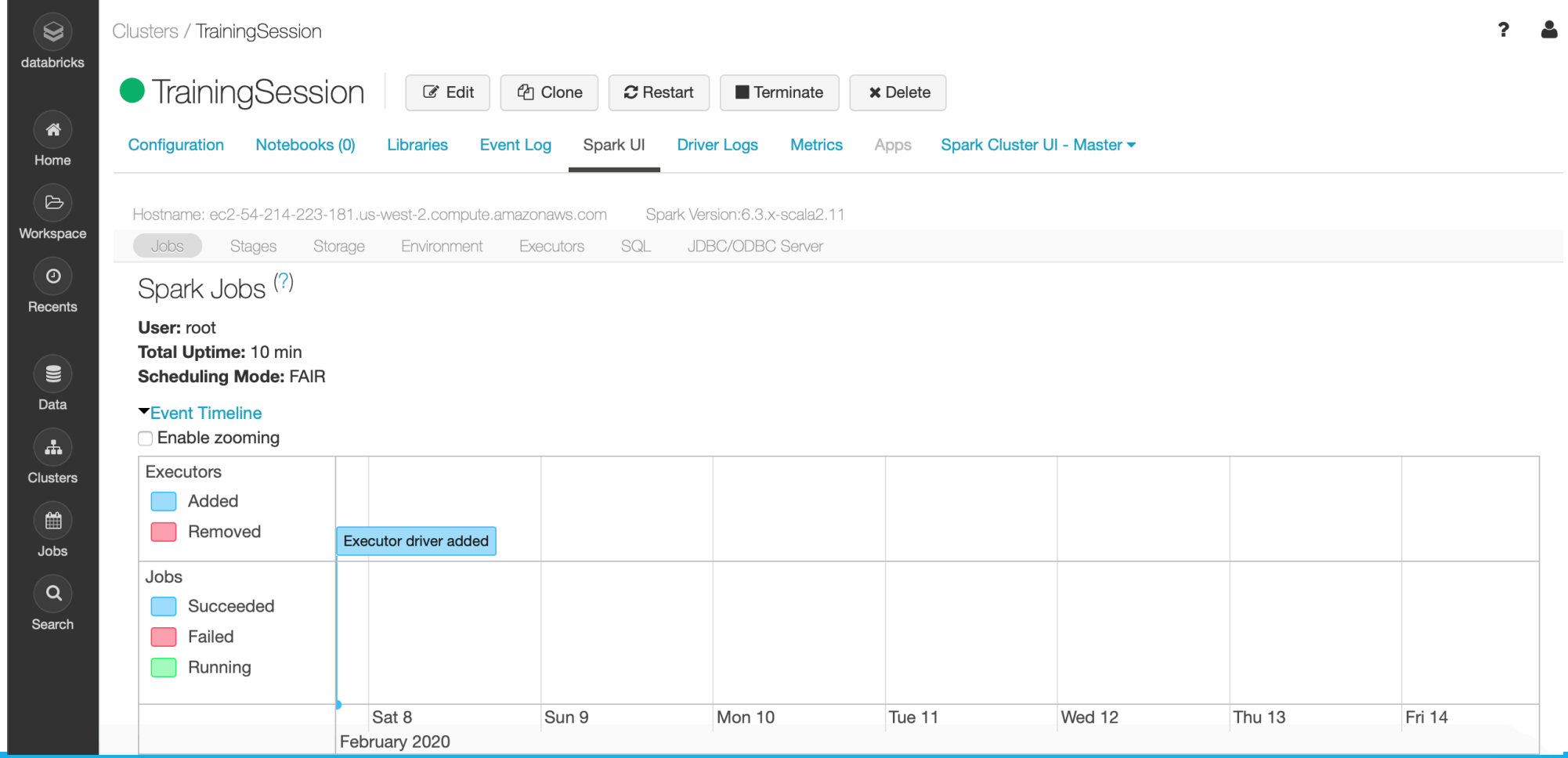
CREATING CLUSTER



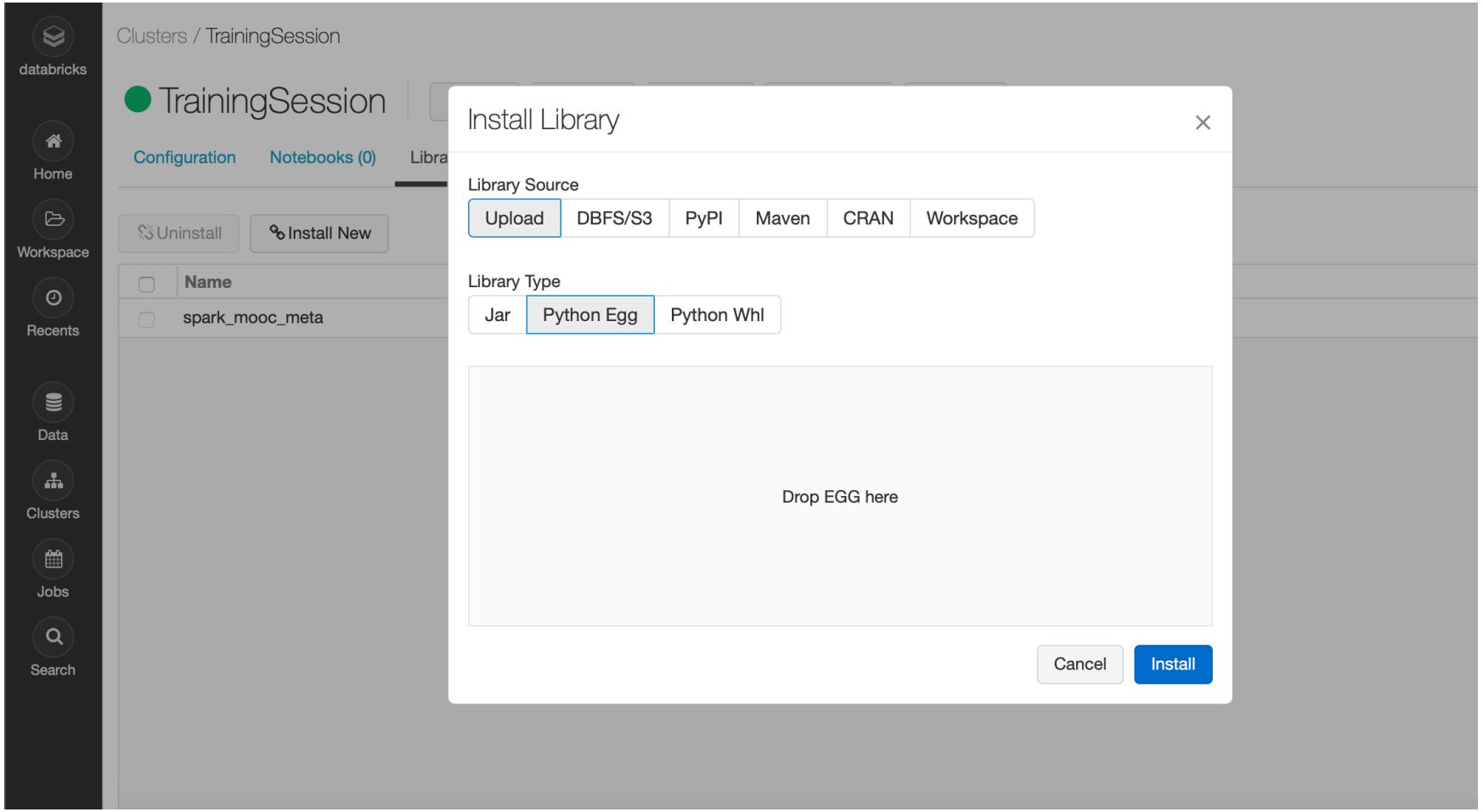
CREATING CLUSTER



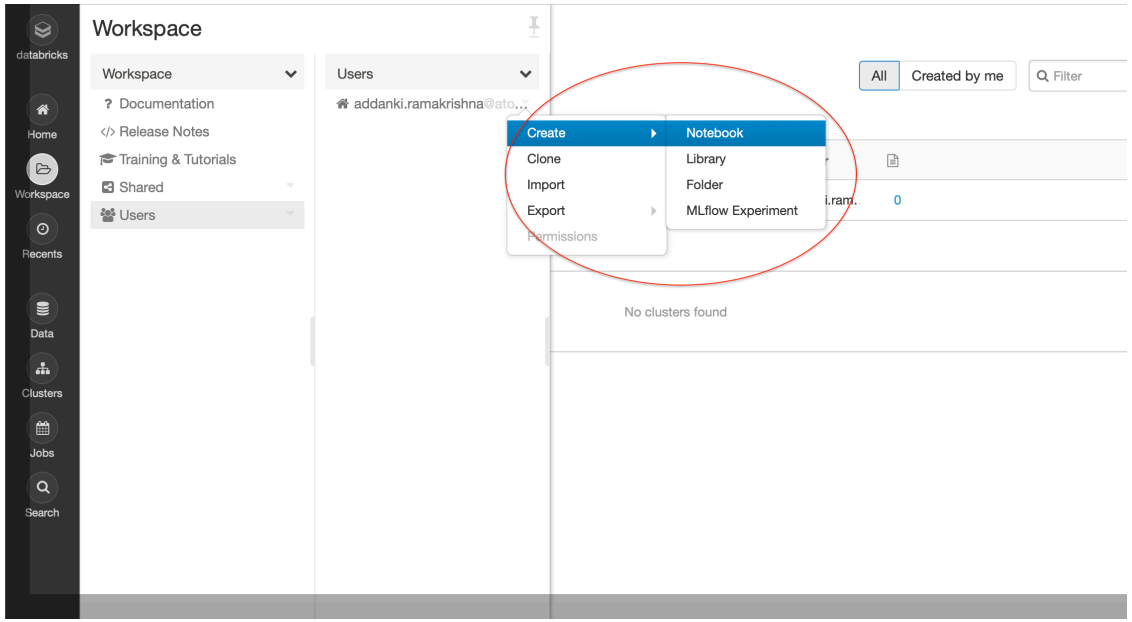
CLUSTER CREATED

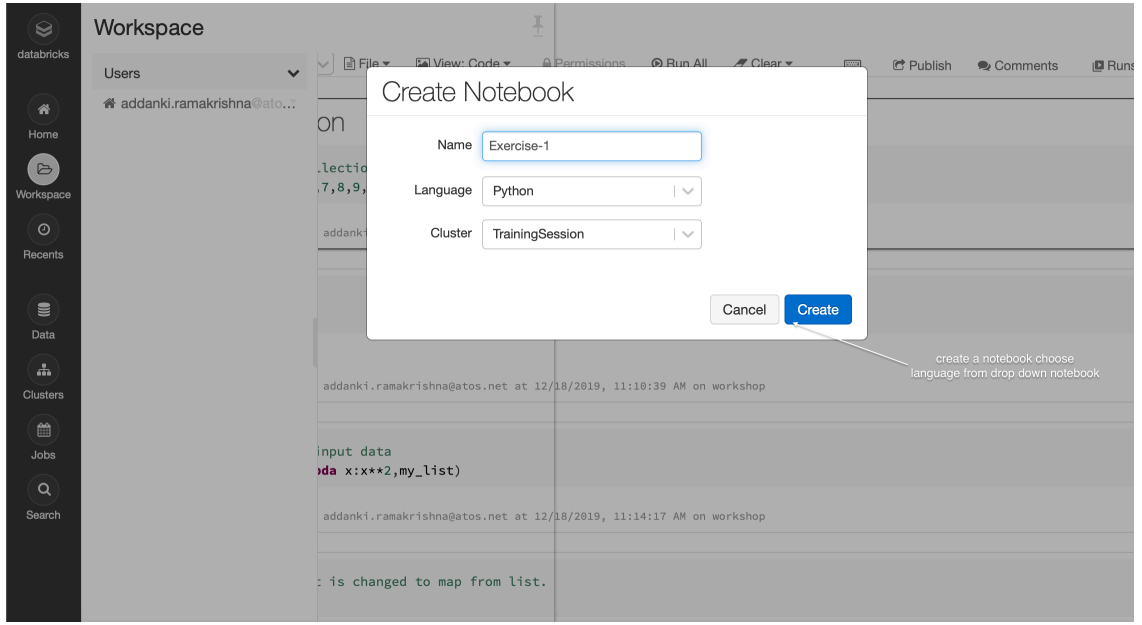


SPARK UI, LOGS & METRICS

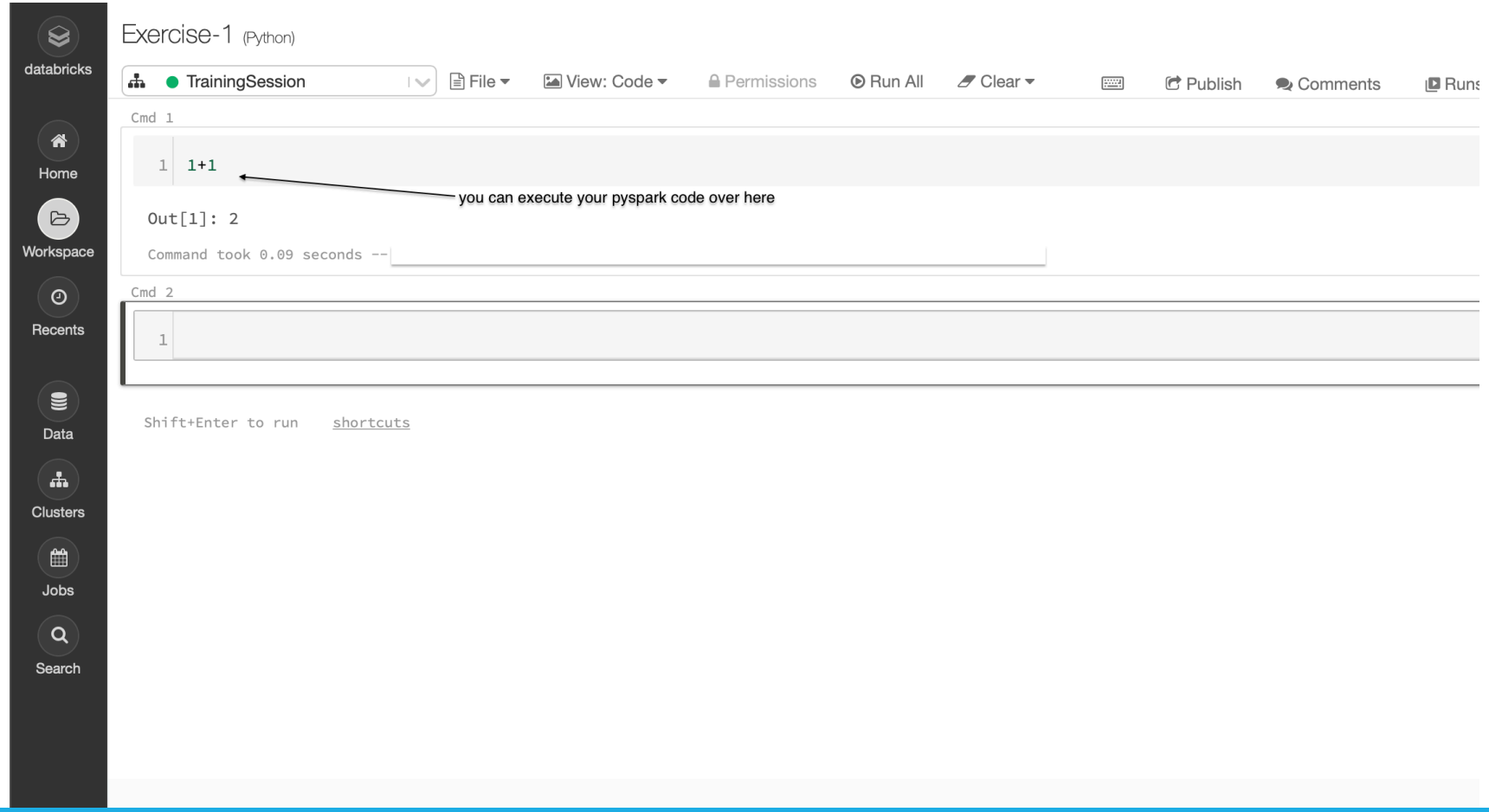


INSTALLING LIBRARIES

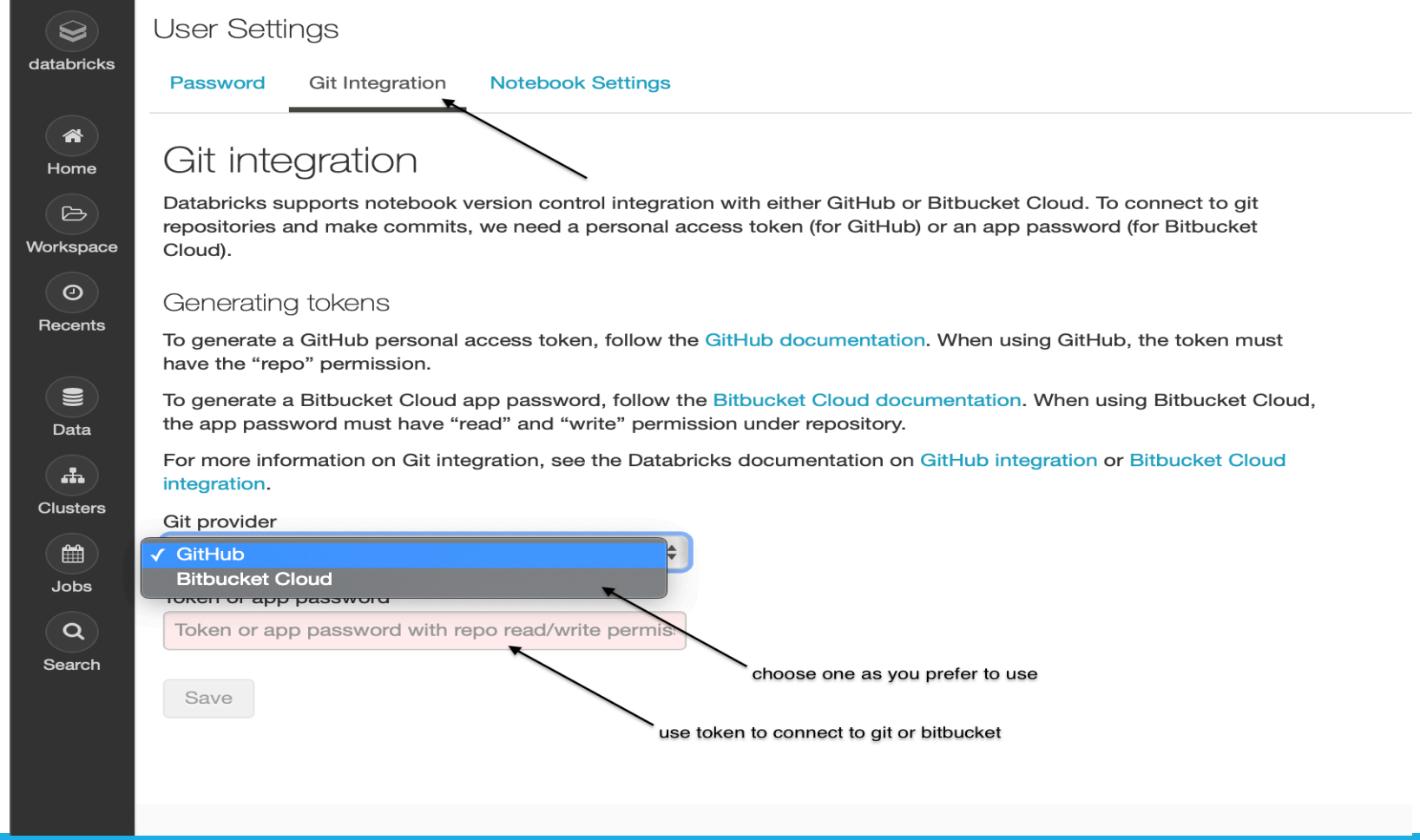




CREATING NOTEBOOKS



EXECUTING CODE IN NOTEBOOK



GIT/ BIT BUCKET INTEGRATION OF NOTEBOOK

QUESTIONS ??


THANK
YOU