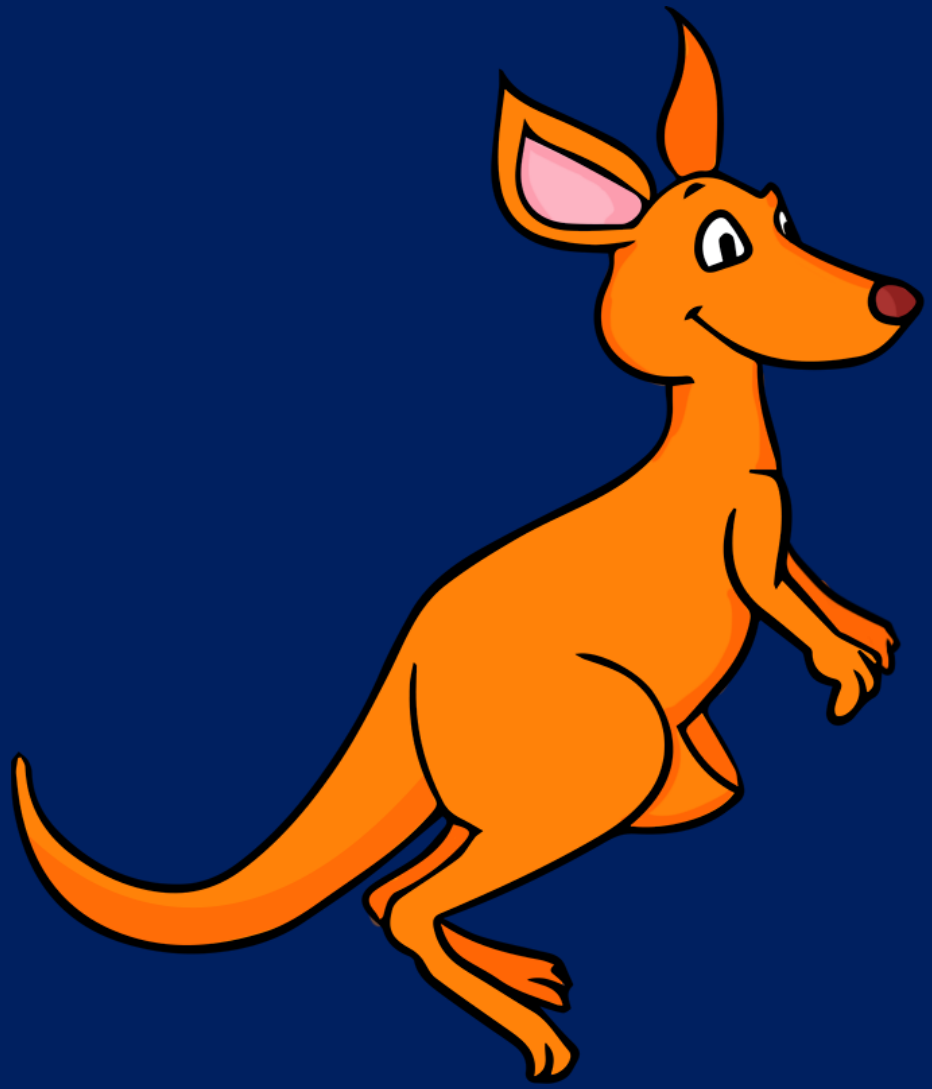
Databricks

Security & Topics

Bryan Cafferky

https://github.com/bcafferky/shared

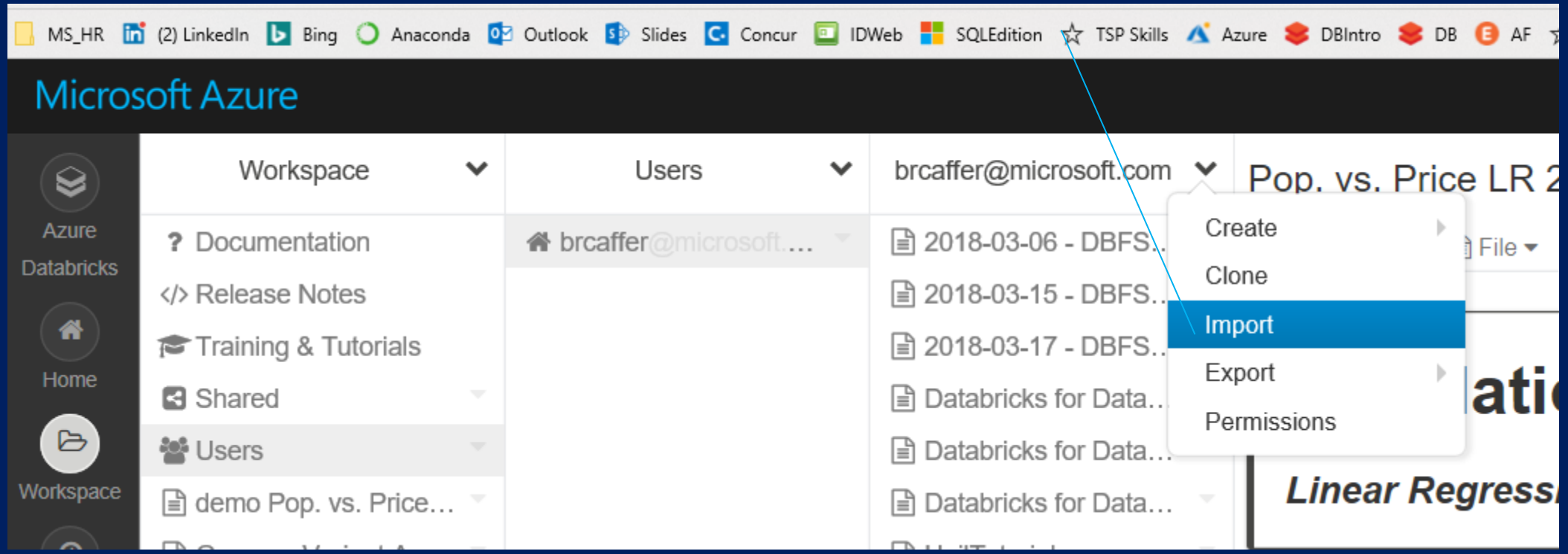
Jumpstart with Sample Notebooks



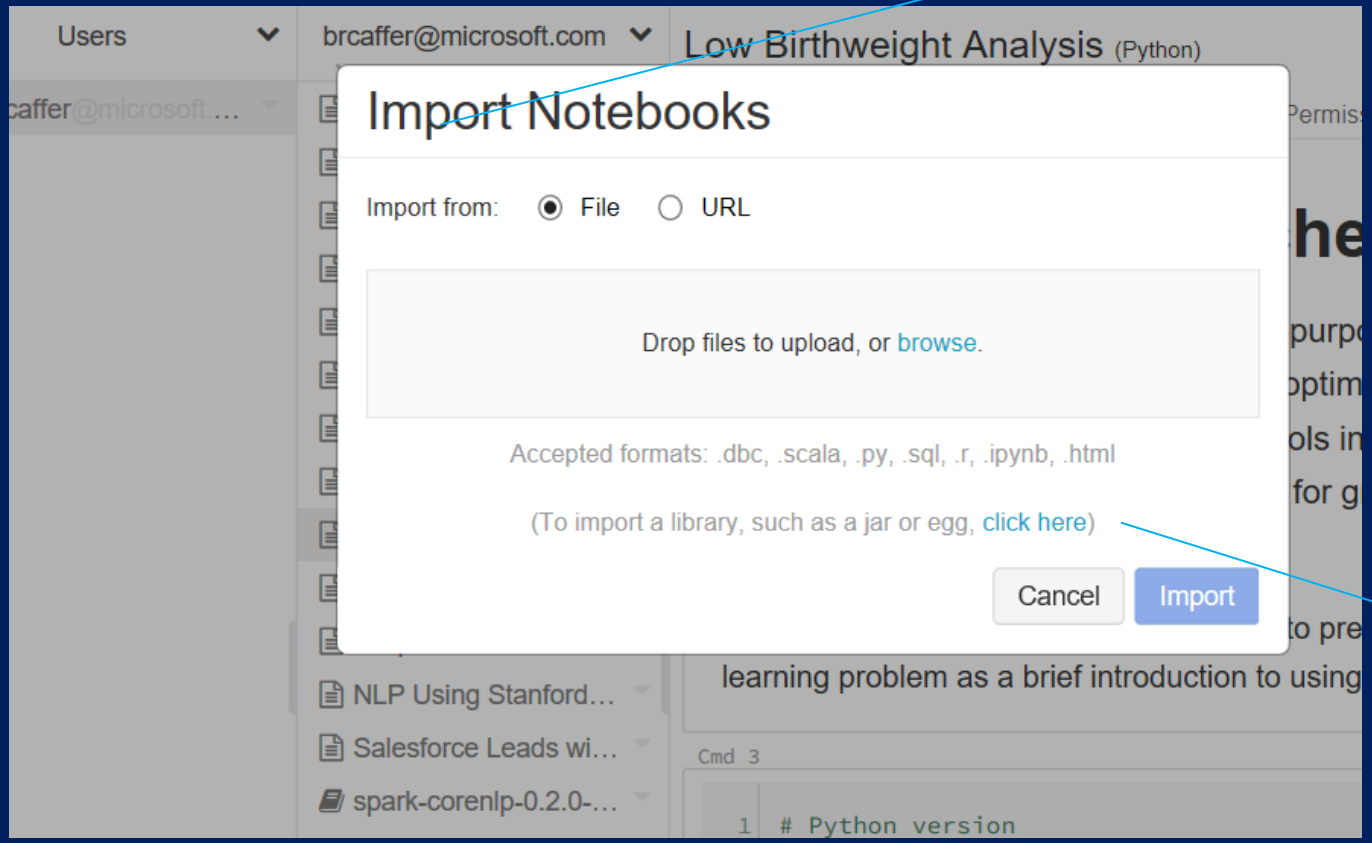
https://clipartion.com/free-clipart-12075/

Importing Notebooks

Under Workspace/Users/User – Select Import from the Dropdown



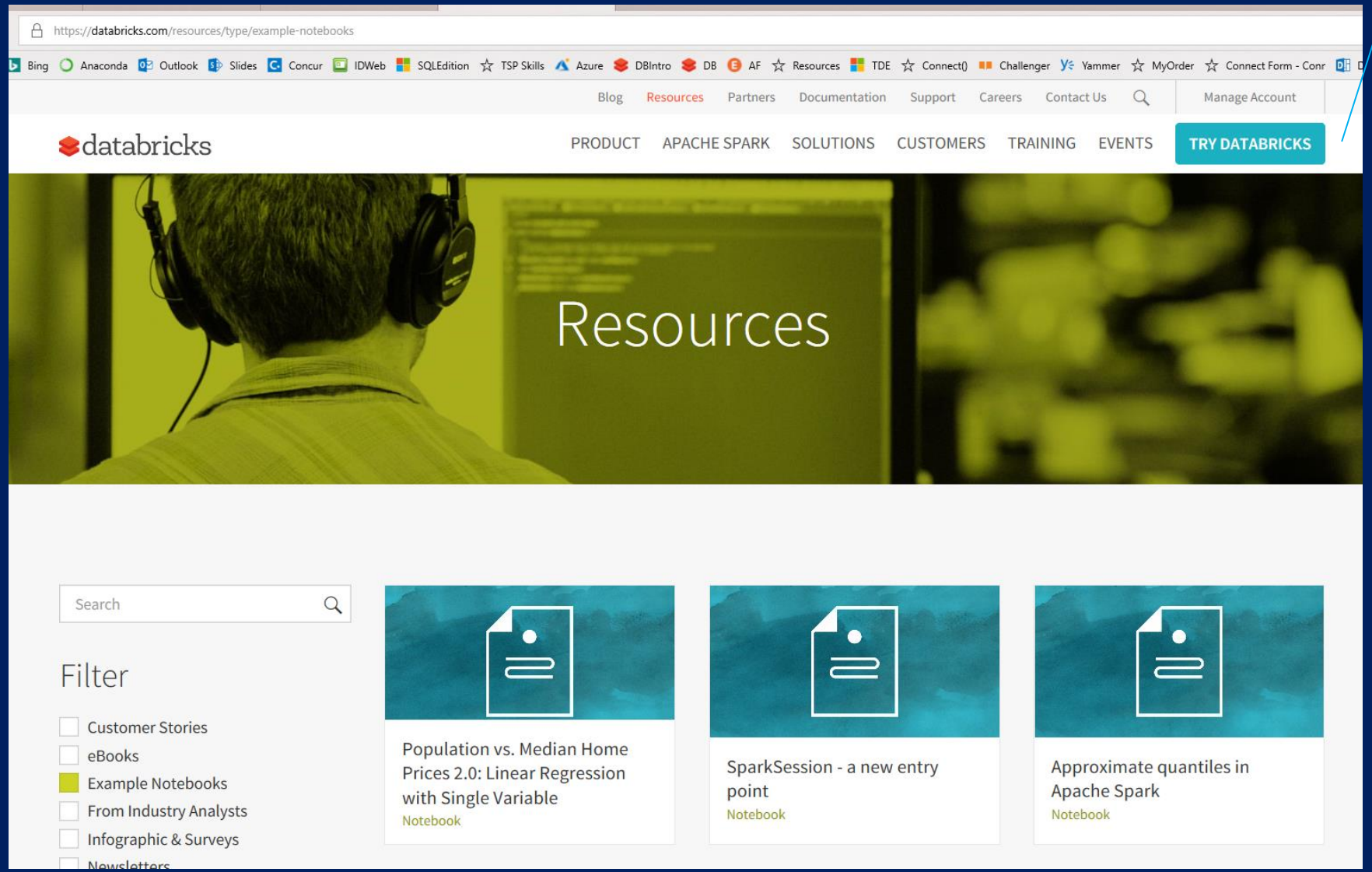
Importing Notebooks



From file or URL

Import library

Importing Notebooks

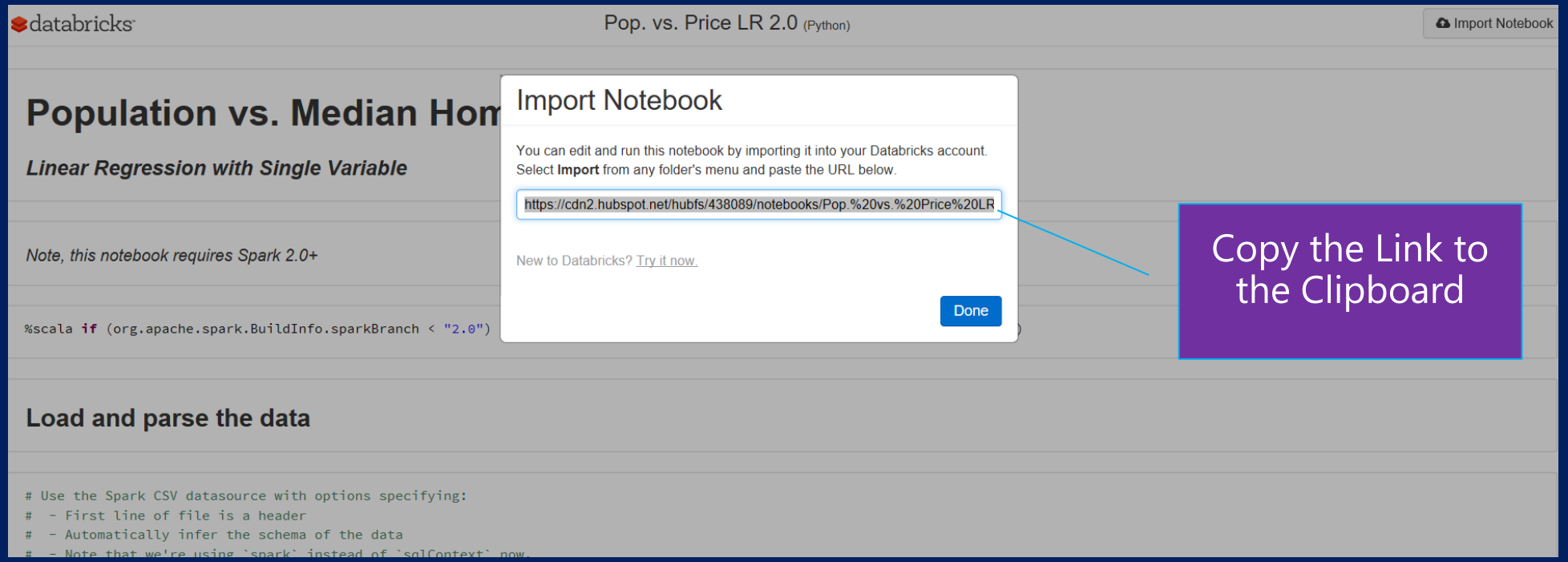


Lot of sample notebooks to try

https://databricks.com/resources/type/example-notebooks

Importing Notebooks

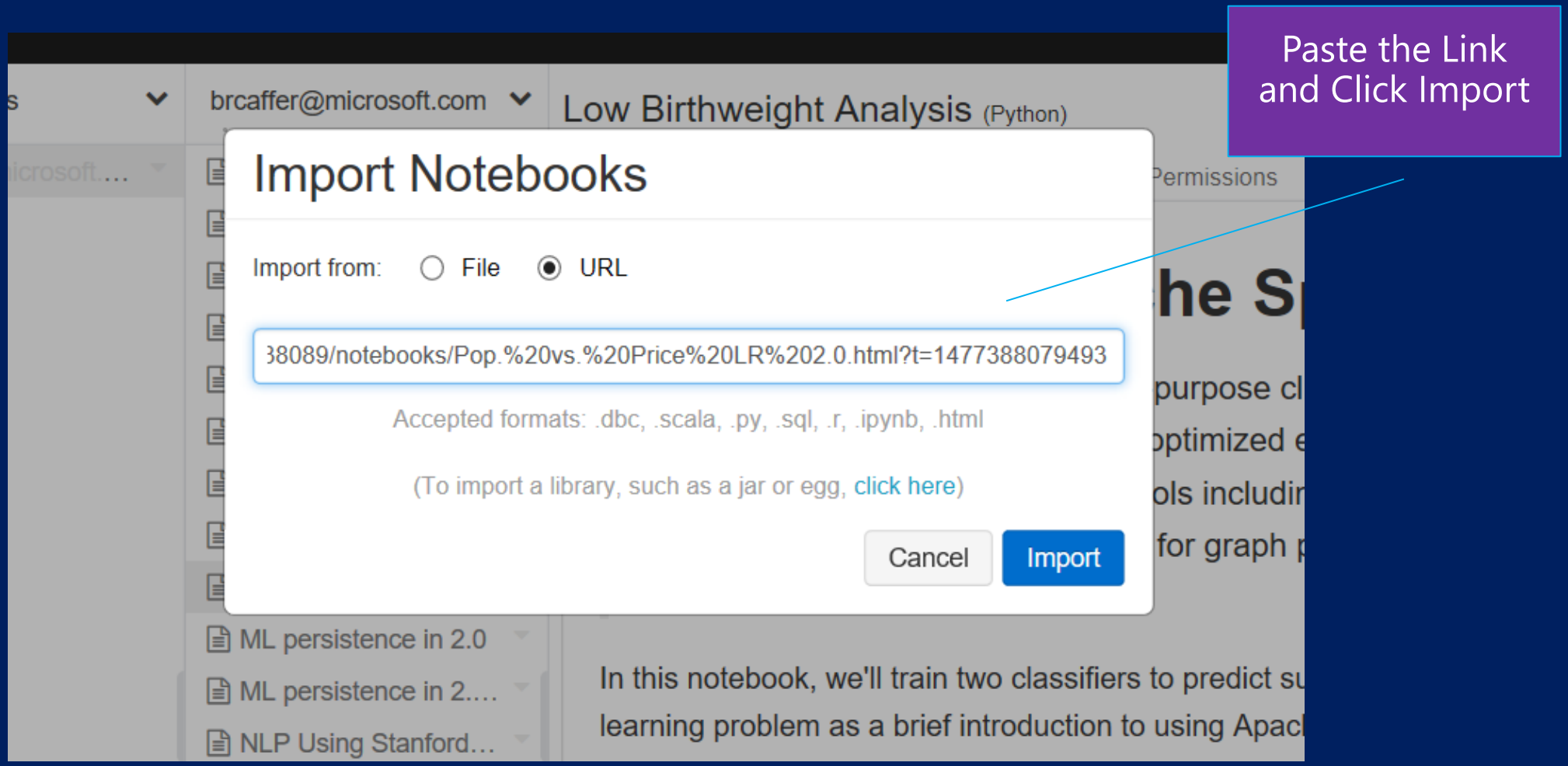
Open and Click on the Import Notebook Button



Copy the Link to the Clipboard

https://databricks.com/resources/type/example-notebooks

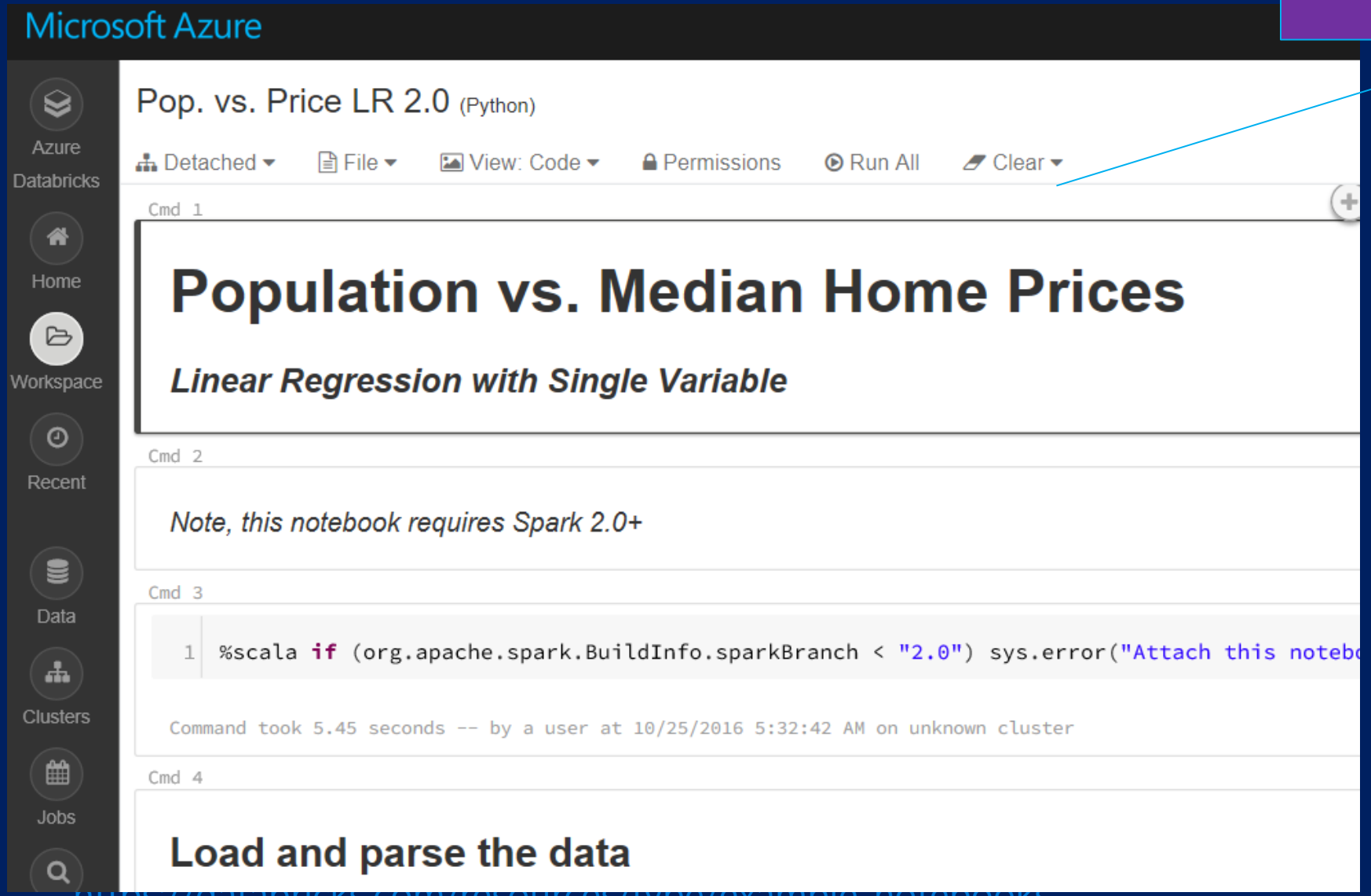
Importing Notebooks



https://databricks.com/resources/type/example-notebooks

Importing Notebooks

Imported Notebook



https://databricks.com/resources/type/example-notebooks

Securing Azure Databricks

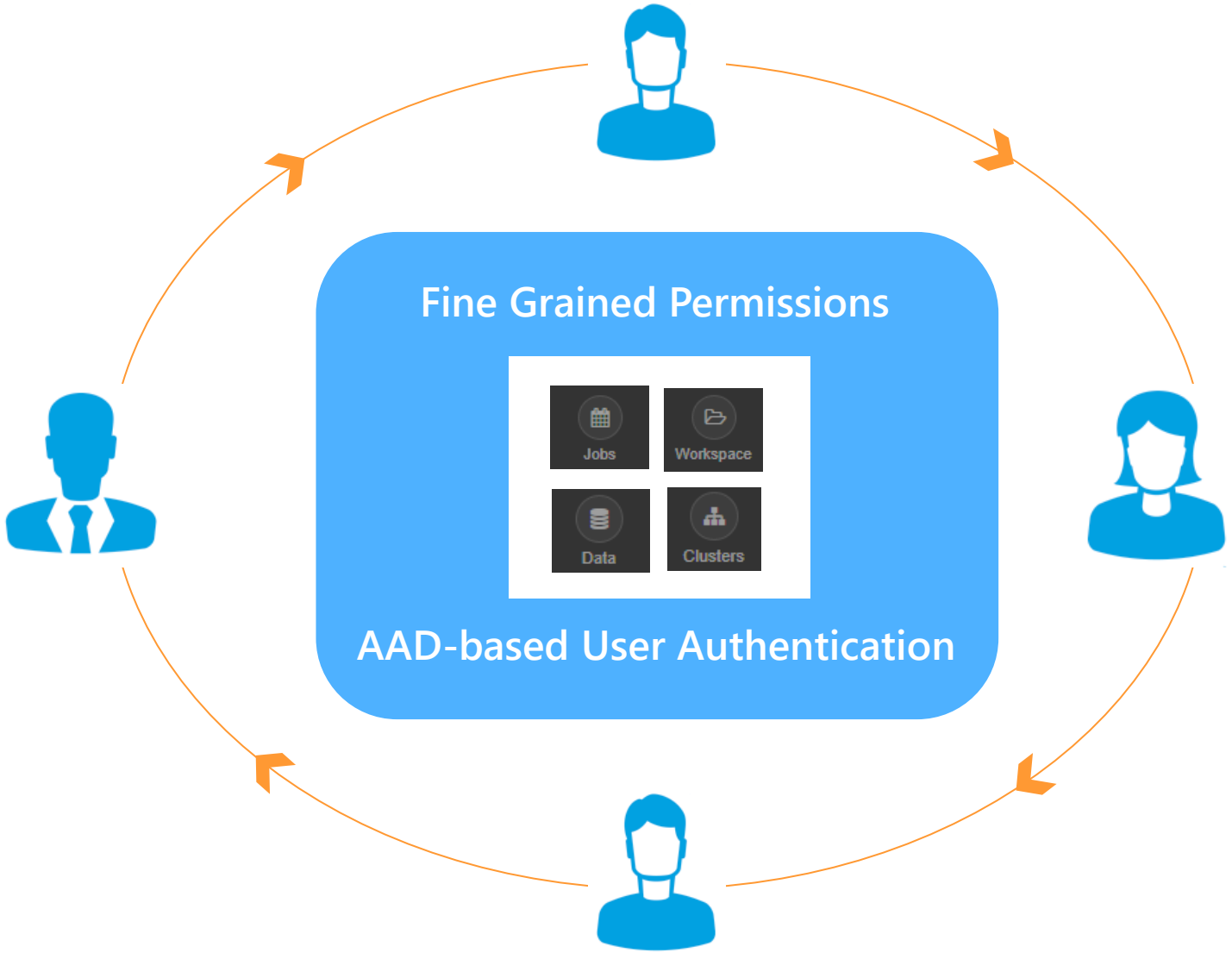
SECURE COLLABORATION

Azure Databricks enables *secure* collaboration between colleagues

- With Azure Databricks colleagues can *securely share* key artifacts such as Clusters, Notebooks, Jobs and Workspaces
- Secure collaboration is enabled through a combination of:

Fine grained permissions: Defines who can do what on which artifacts (access control)

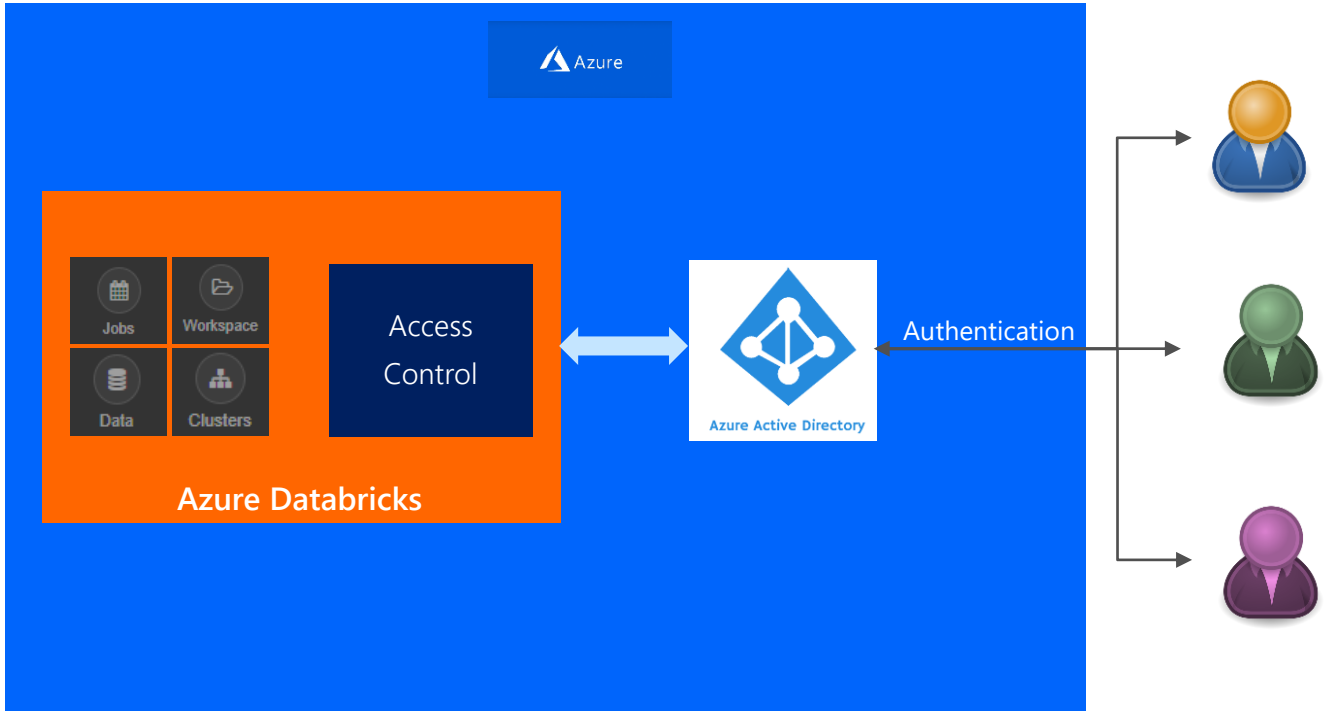
AAD-based authentication: Ensures that users are actually who they claim to be



AZURE DATABRICKS INTEGRATION WITH AAD

Azure Databricks is integrated with AAD—so Azure Databricks users are just regular AAD users

- There is no need to define users—and their access control—separately in Databricks.
- AAD users can be used directly in Azure Databricks for all user-based access control (Clusters, Jobs, Notebooks etc.).
- Databricks has delegated user authentication to AAD enabling single-sign on (SSO) and unified authentication.
- *Notebooks, and their outputs, are stored in the Databricks account*. However, AAD-based access-control ensures that only authorized users can access them.



DATABRICKS ACCESS CONTROL

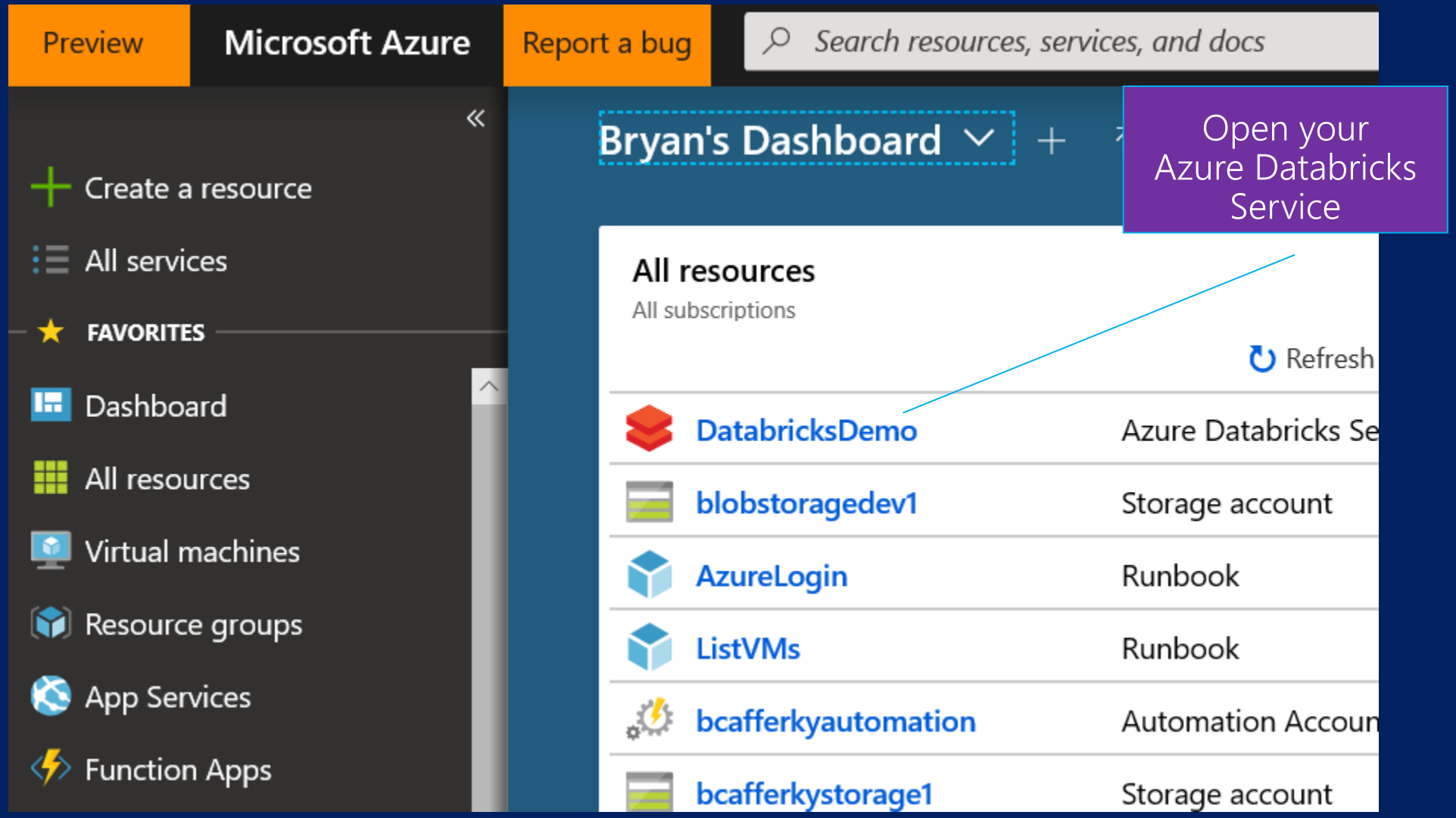
Access control can be defined at the user level via the Admin Console

Access Control can be defined for Workspaces, Clusters, Jobs and REST APIs

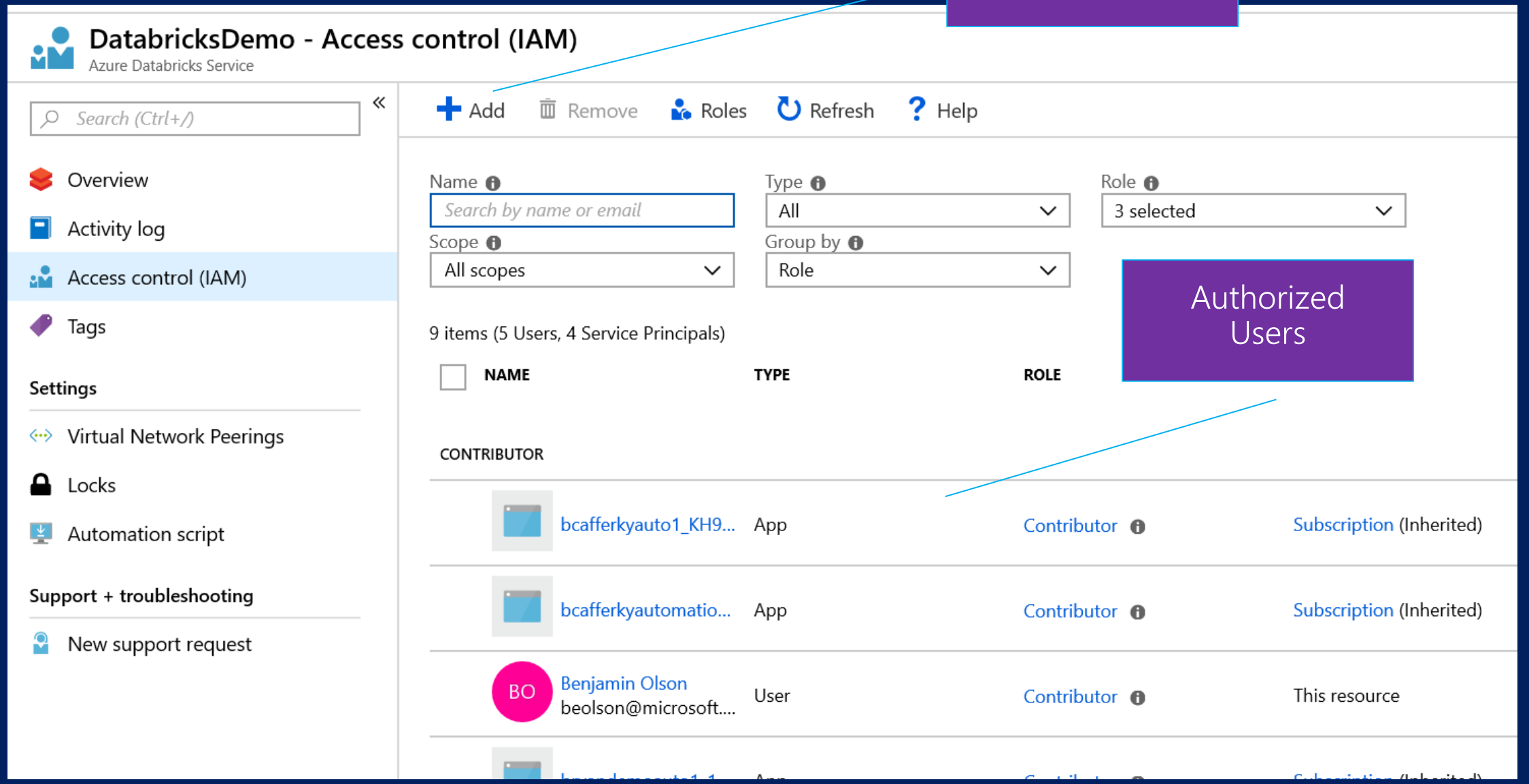
Databricks Access Control

Workspace Access Control	Defines who can who can view, edit, and run notebooks in their workspace
Cluster Access Control	Allows users to who can attach to, restart, and manage (resize/delete) clusters. Allows Admins to specify which users have permissions to create clusters
Jobs Access Control	Allows owners of a job to control who can view job results or manage runs of a job (run now/cancel)
REST API Tokens	Allows users to use personal access tokens instead of passwords to access the Databricks REST API

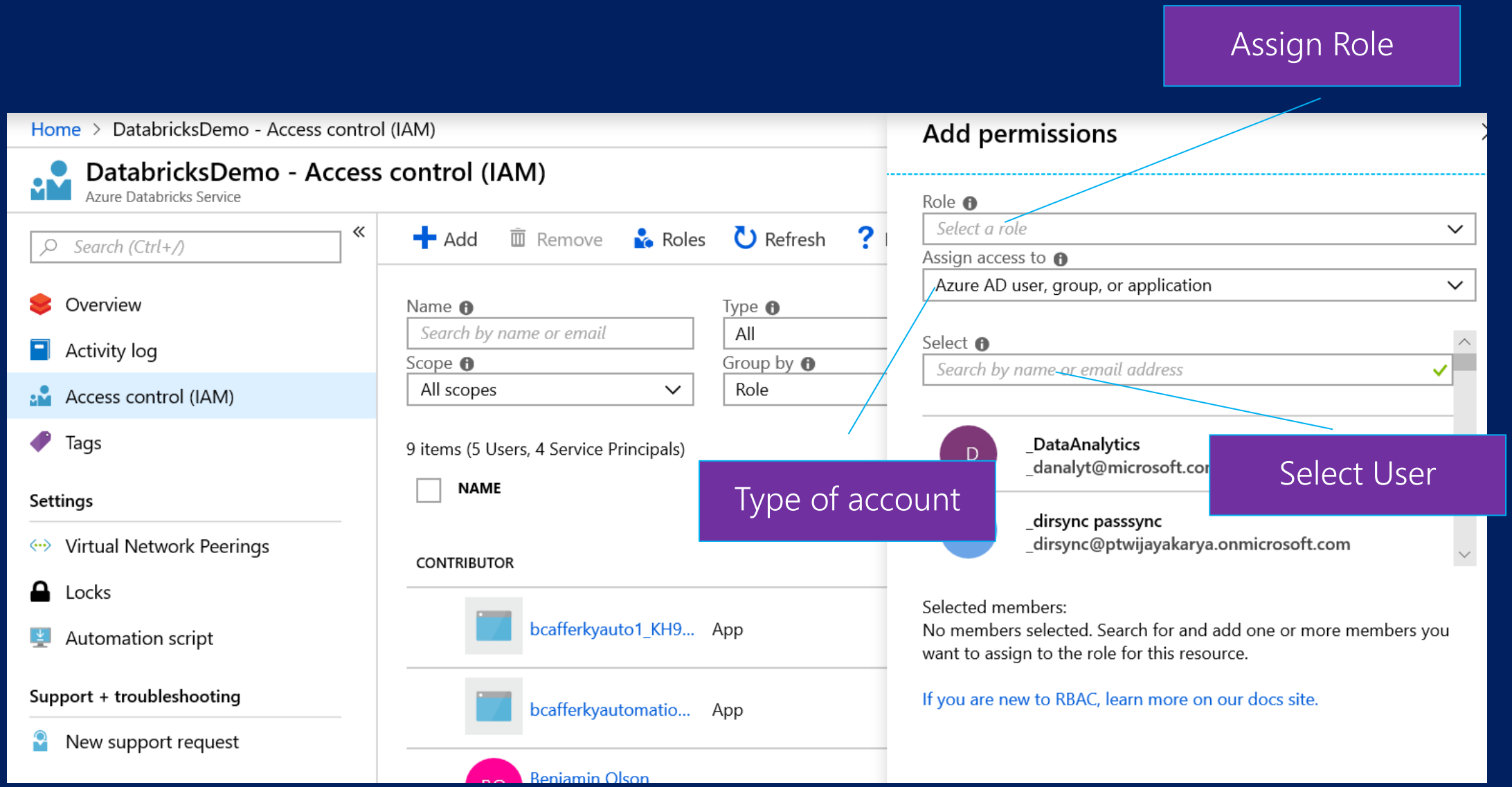
Securing Azure Databricks



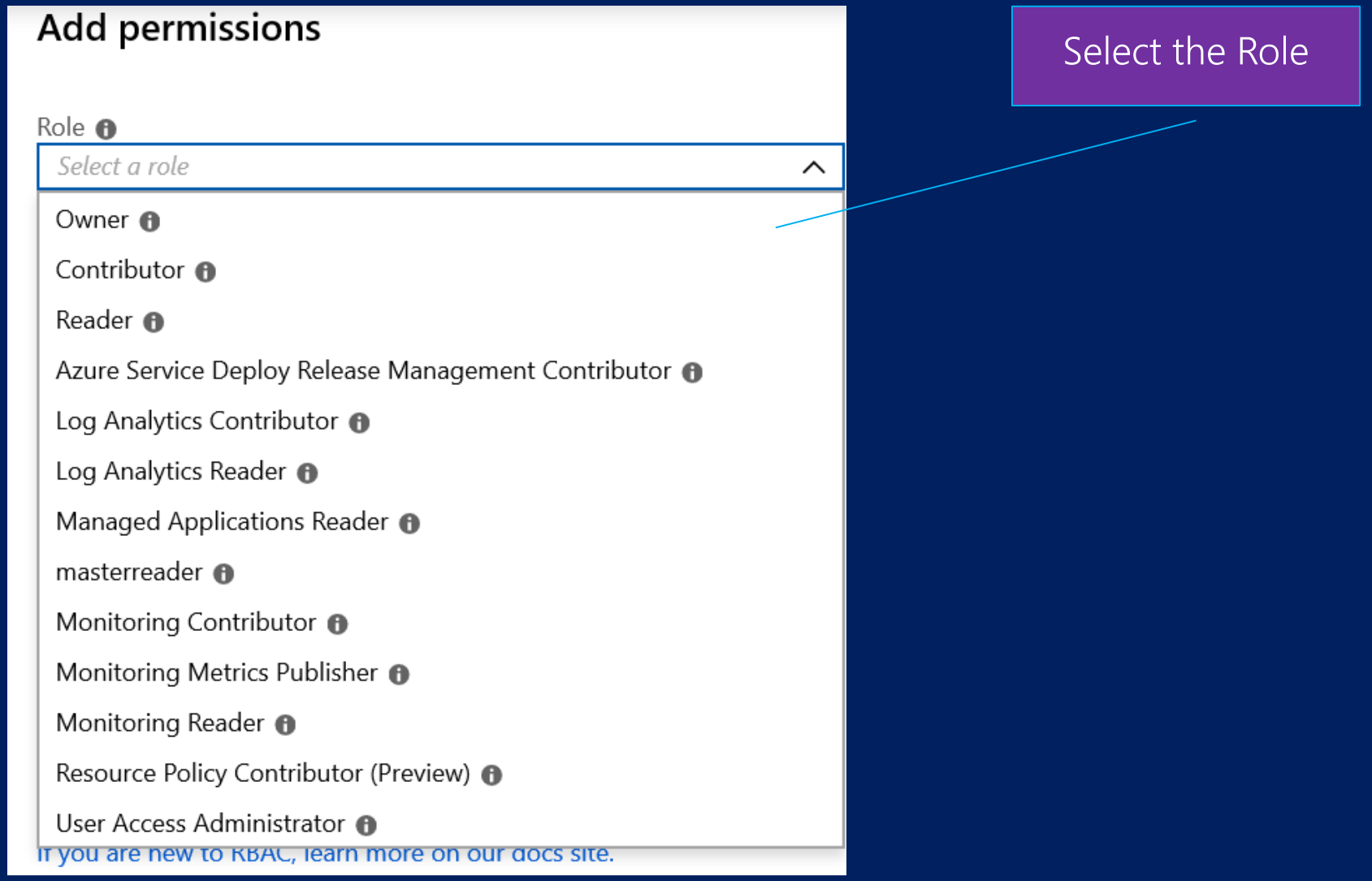
Securing Azure Databricks



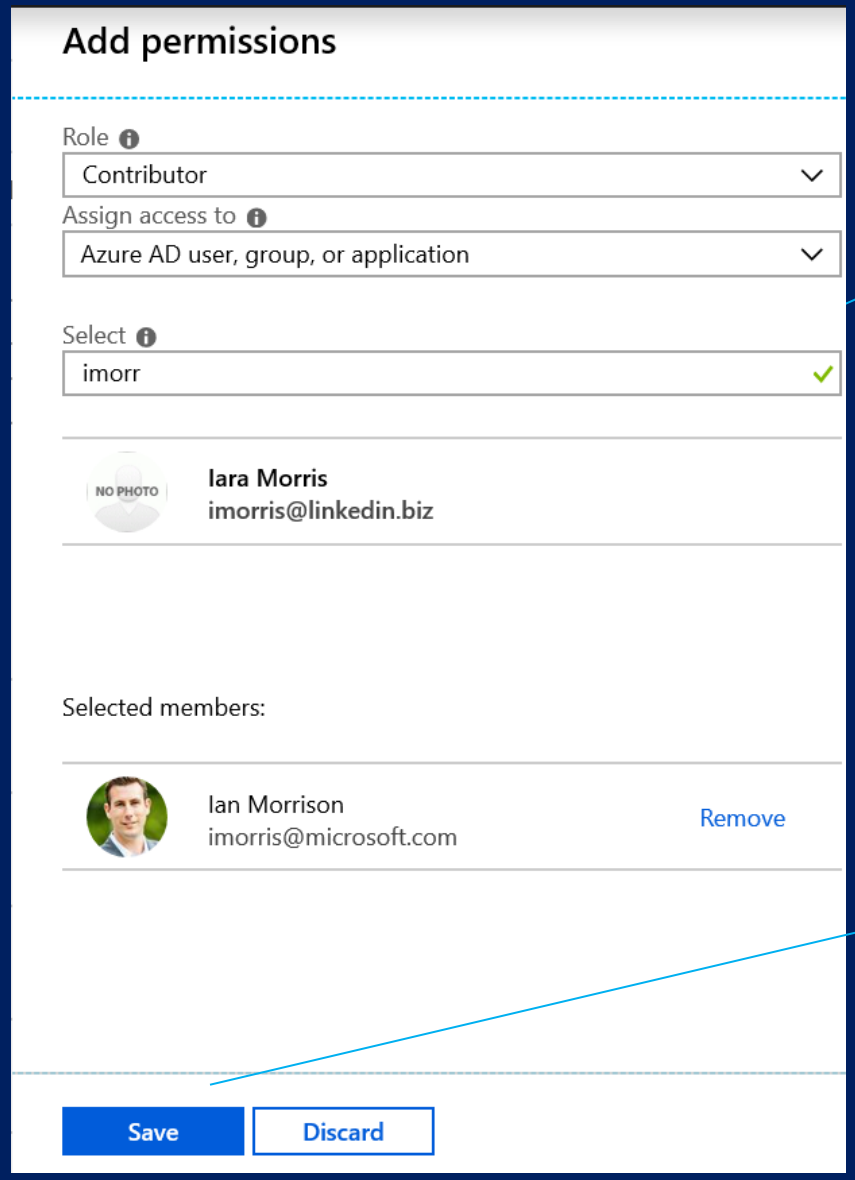
Securing Azure Databricks



Securing Azure Databricks



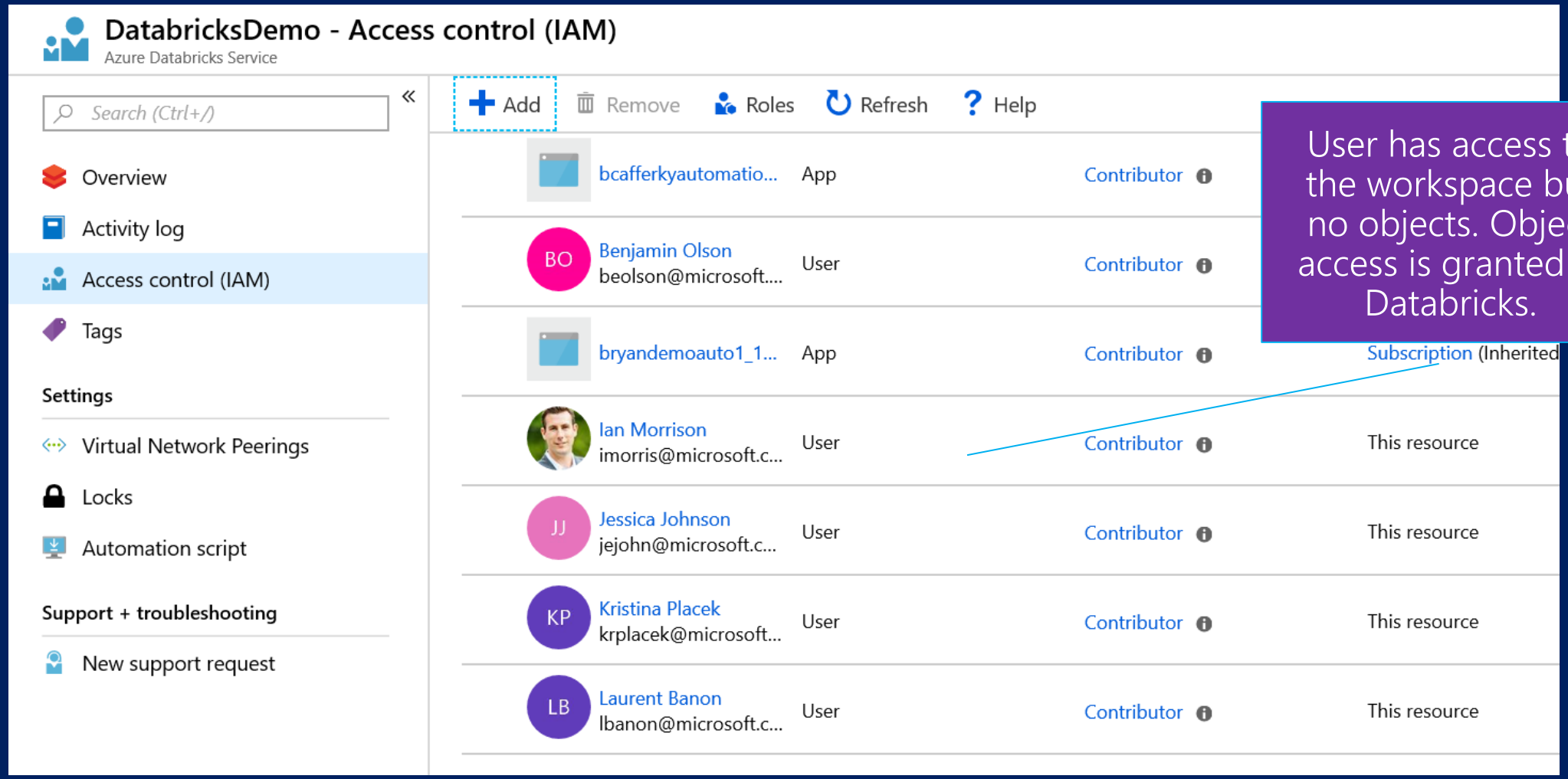
Securing Azure Databricks



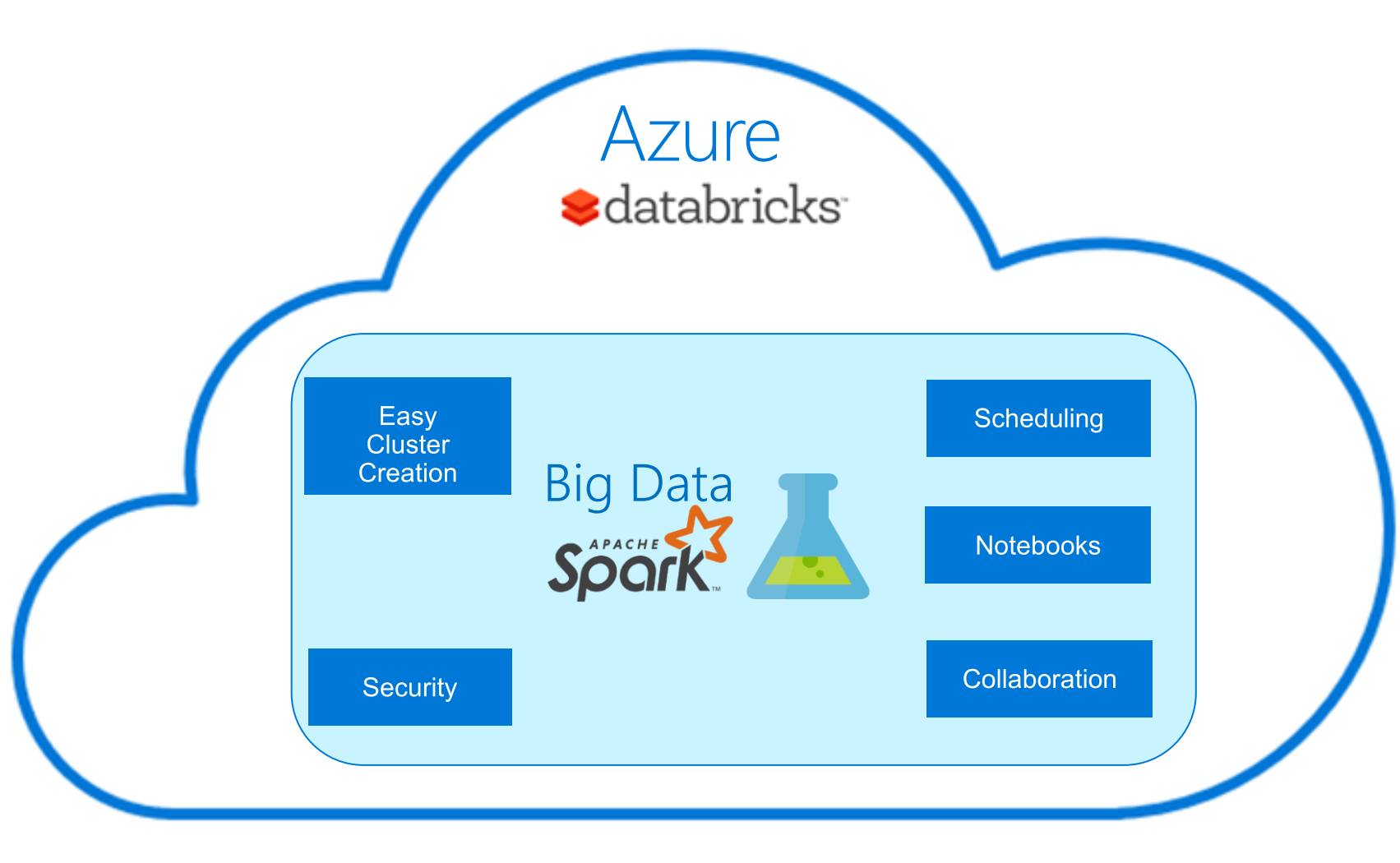
Search for the group, user, or application

Save

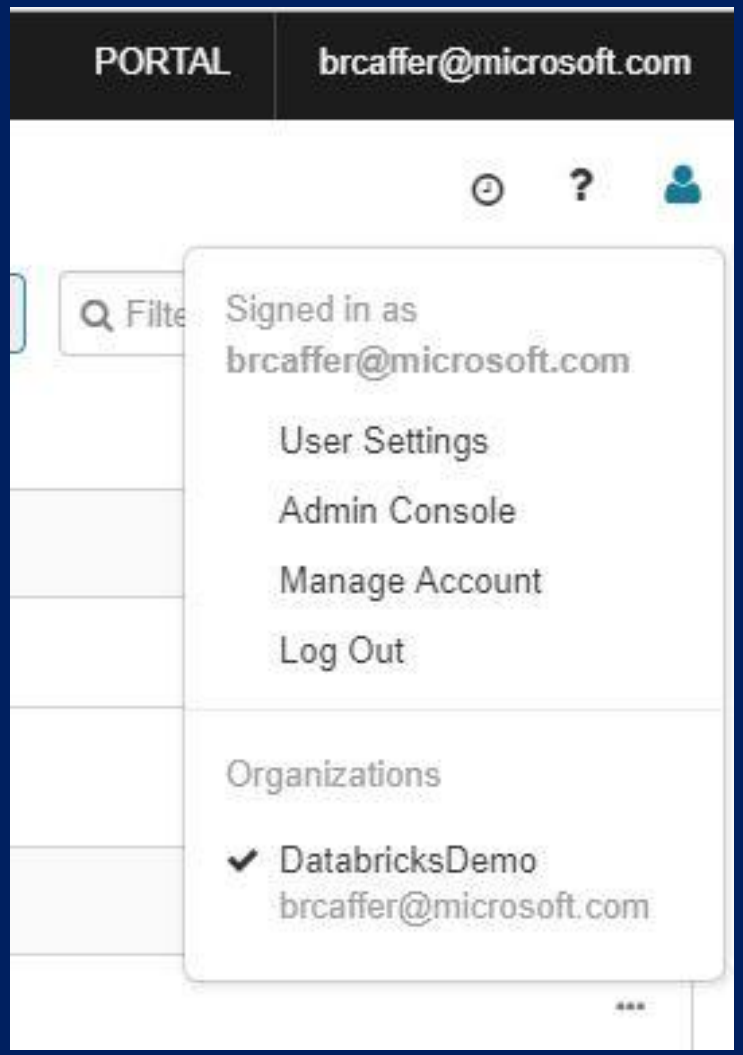
Securing Azure Databricks



Securing Azure Databricks: Workspace Objects

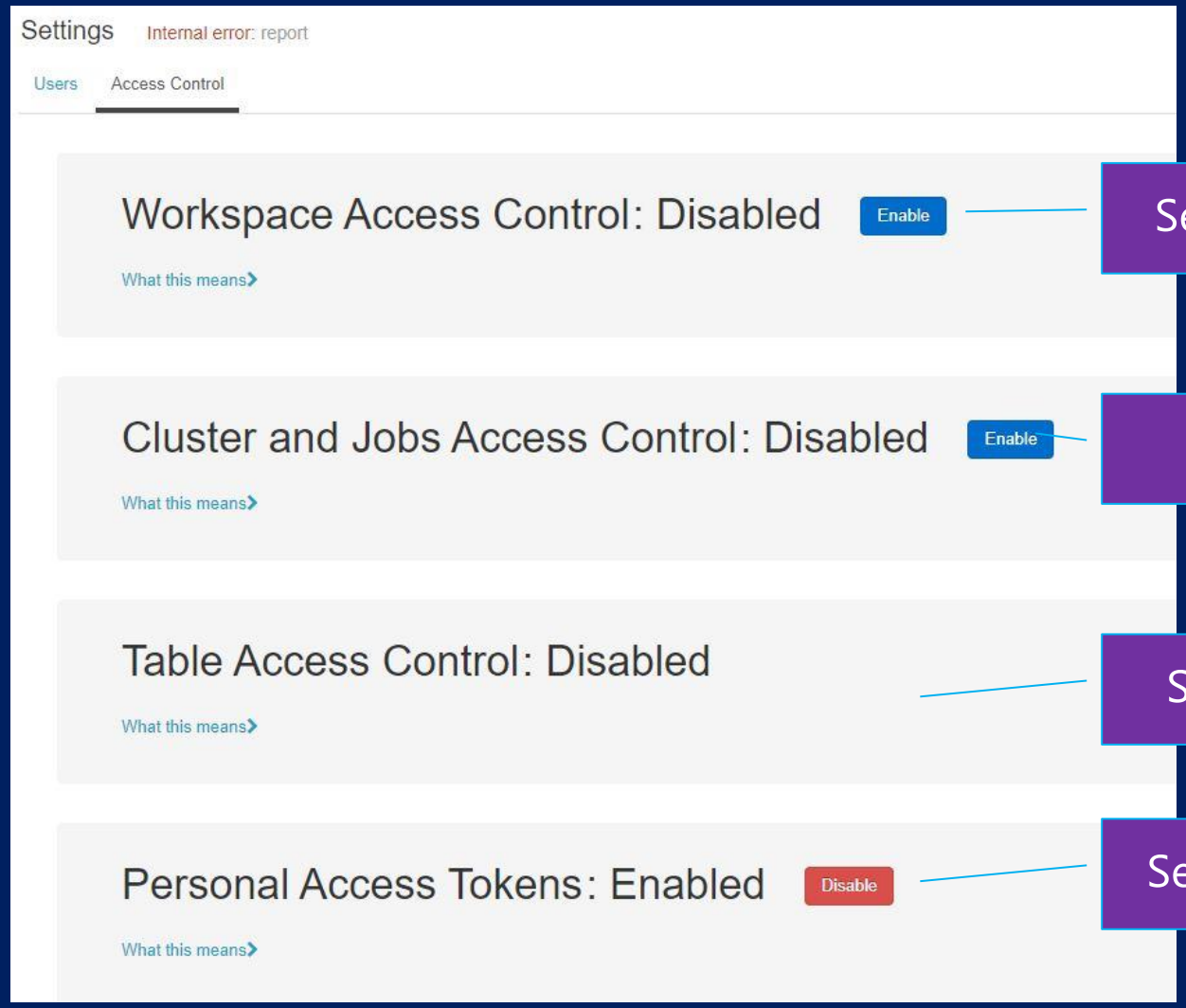


Administration



Administration Settings

What do you want to restrict?



Secure Workspaces

Secure Clusters

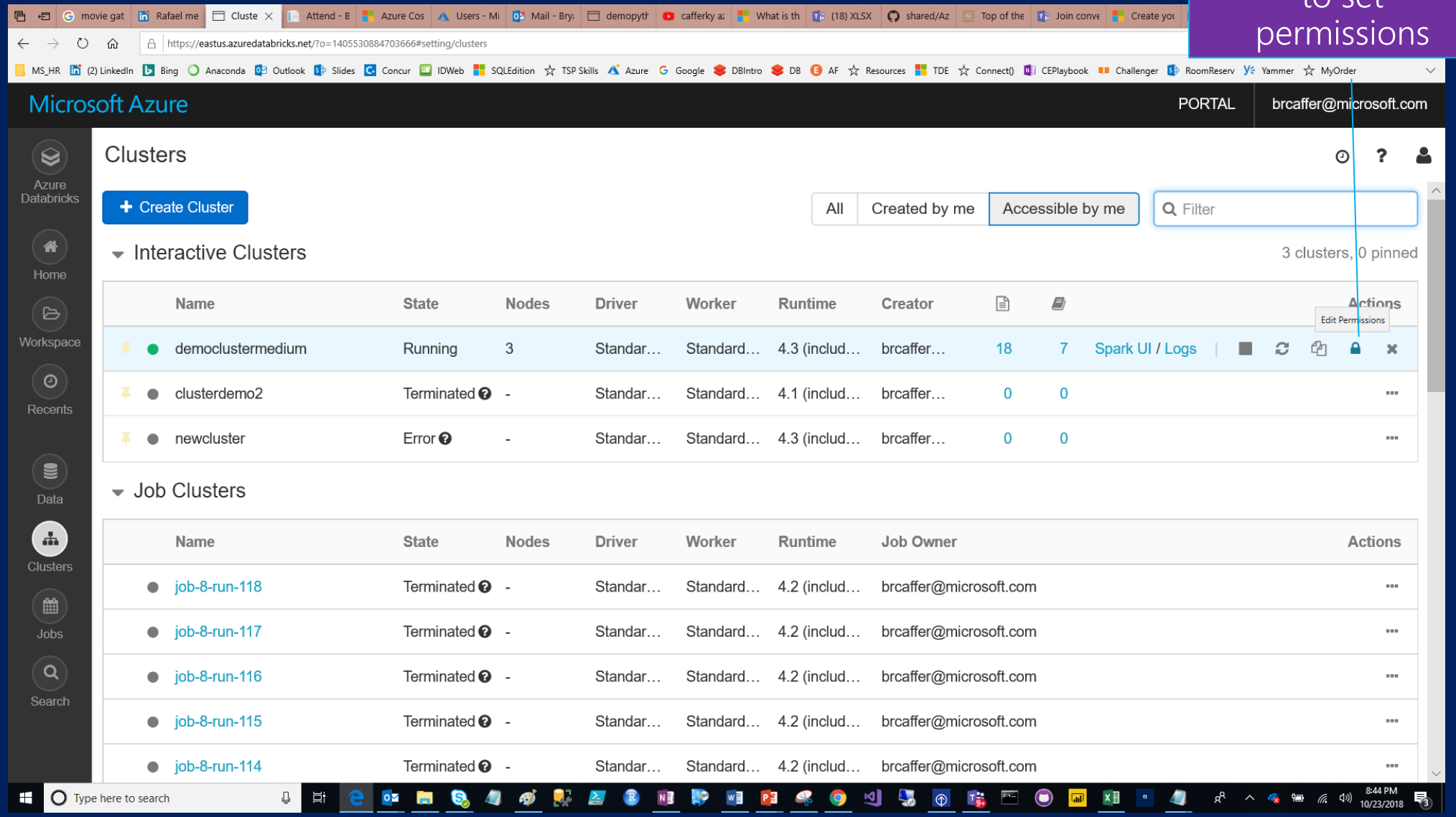
Secure SQL Tables

Secure the REST API

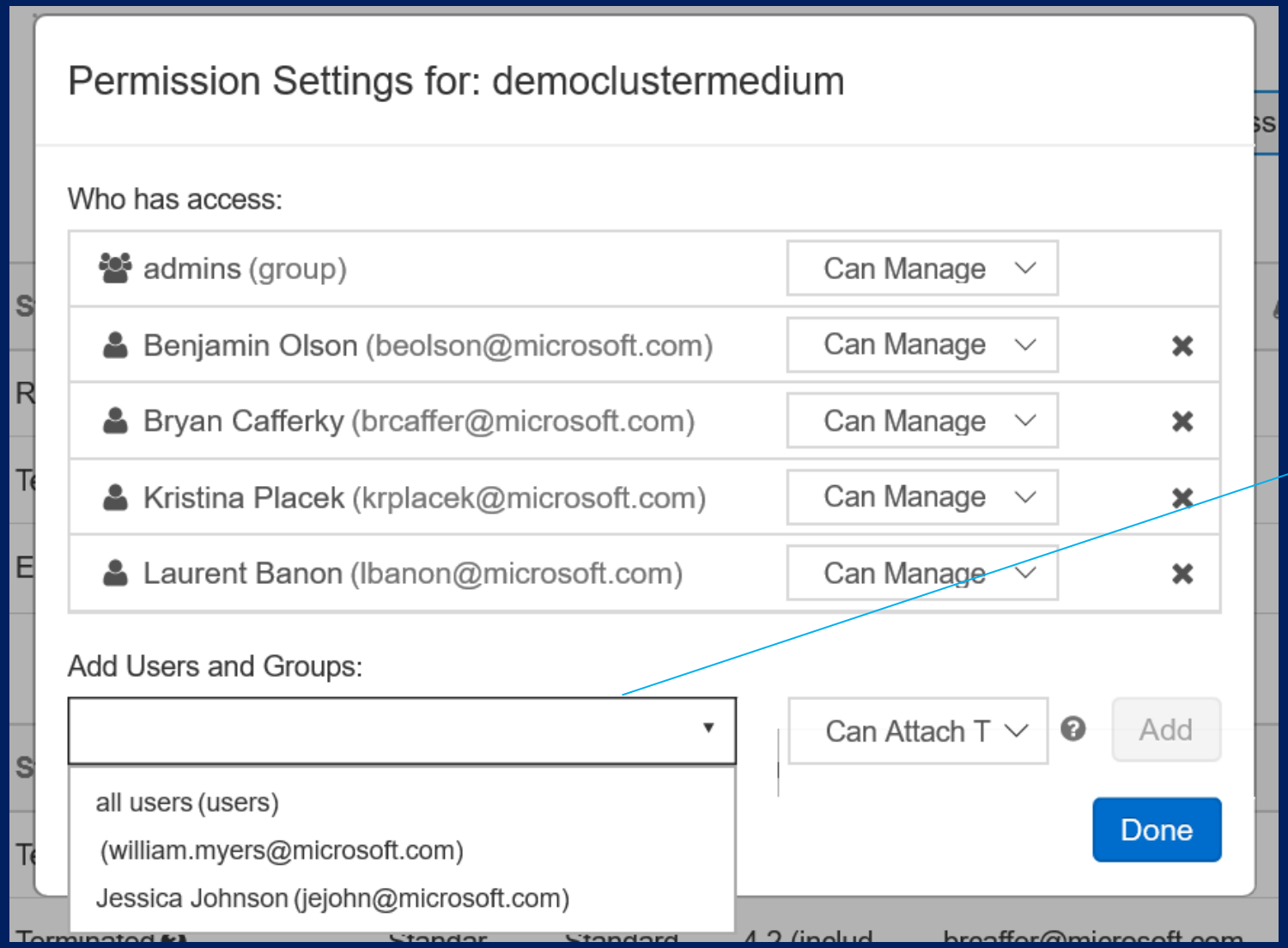
Detailed Steps on Securing Other Objects

Securing Azure Databricks: Clusters

Click on the lock to set permissions



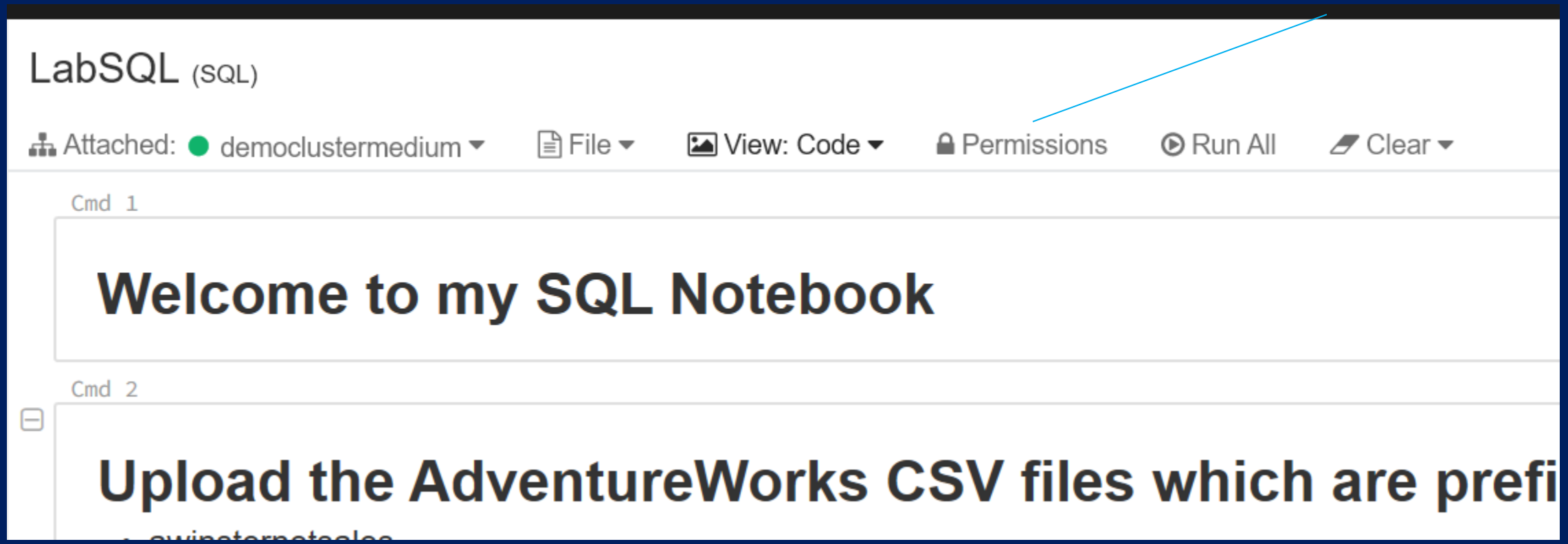
Securing Azure Databricks: Cluster Access



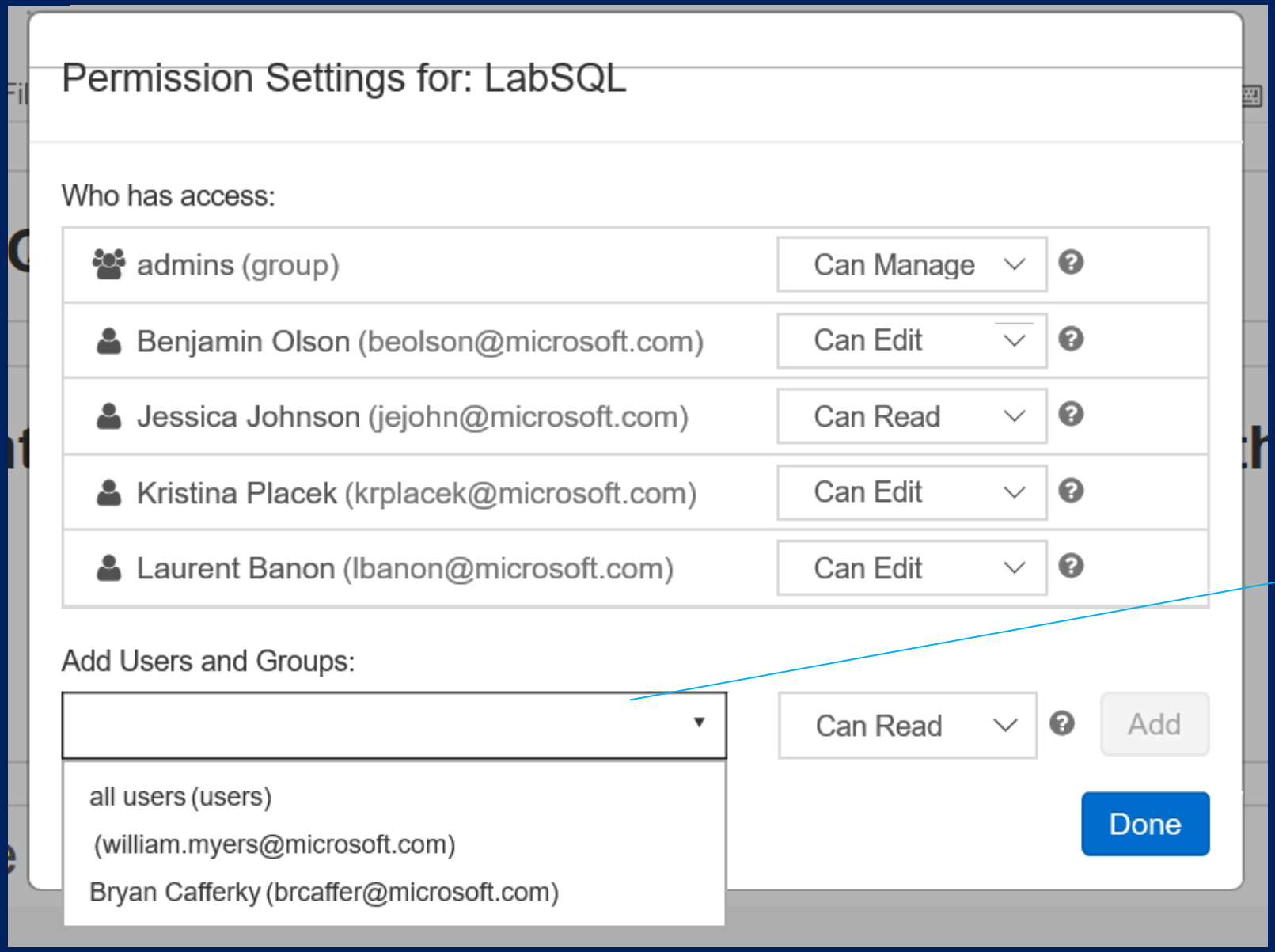
Assign Permissions

Securing Azure Databricks: Notebooks

Click to set permissions

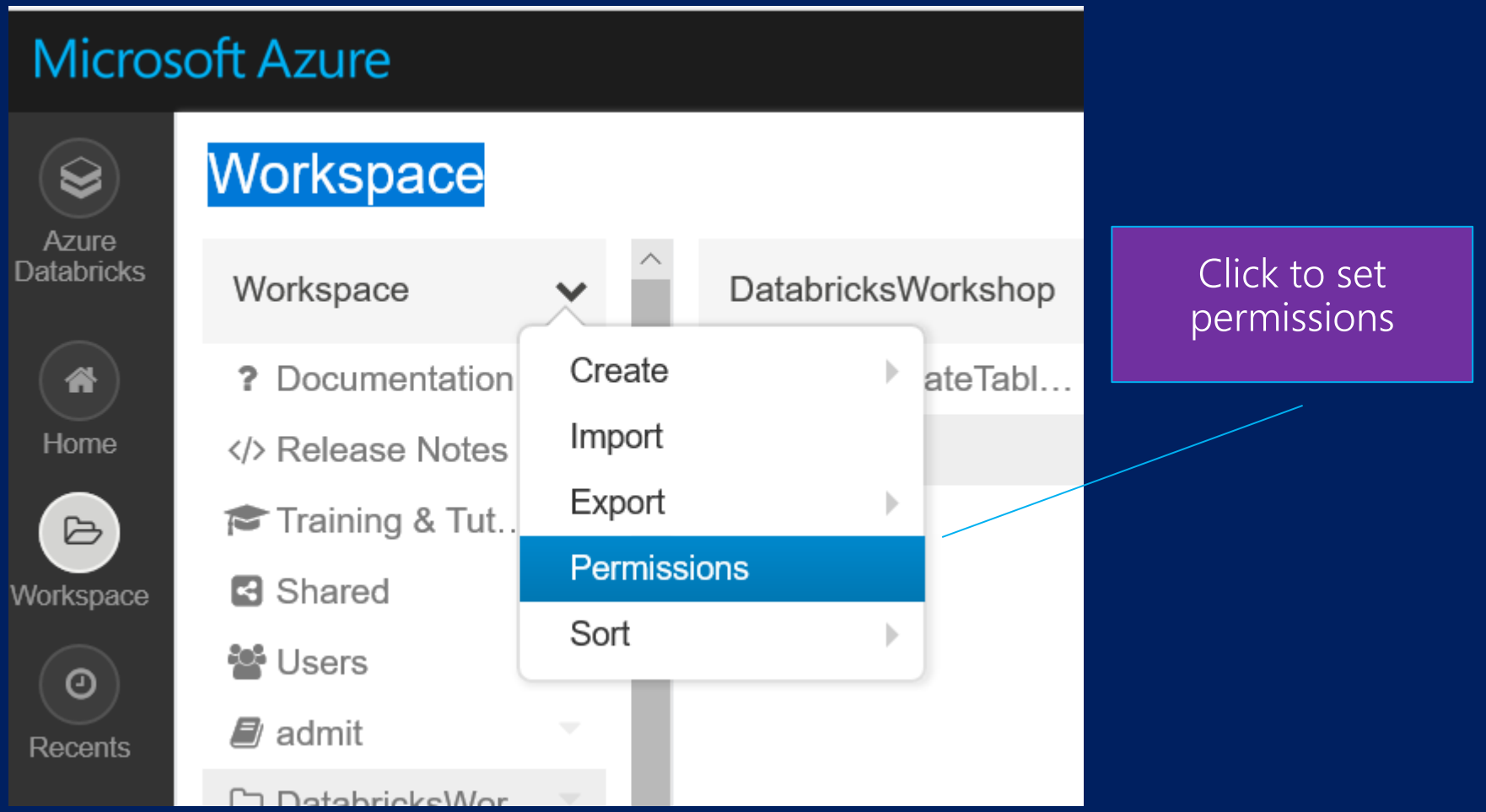


Securing Azure Databricks



Click to set permissions

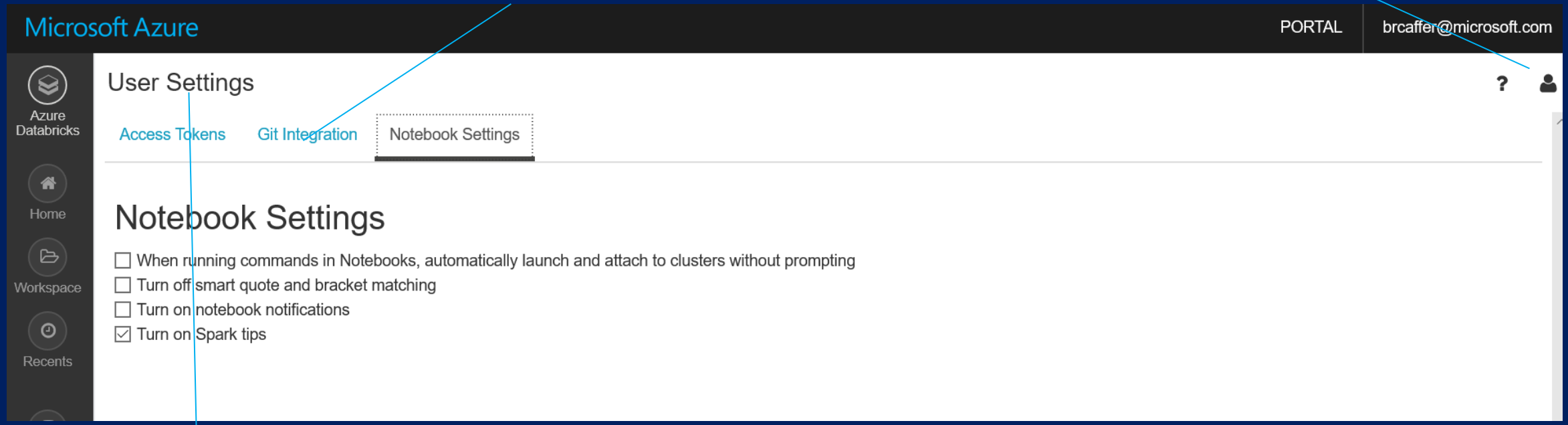
Securing Azure Databricks: Workspace



Securing Azure Databricks: User Settings

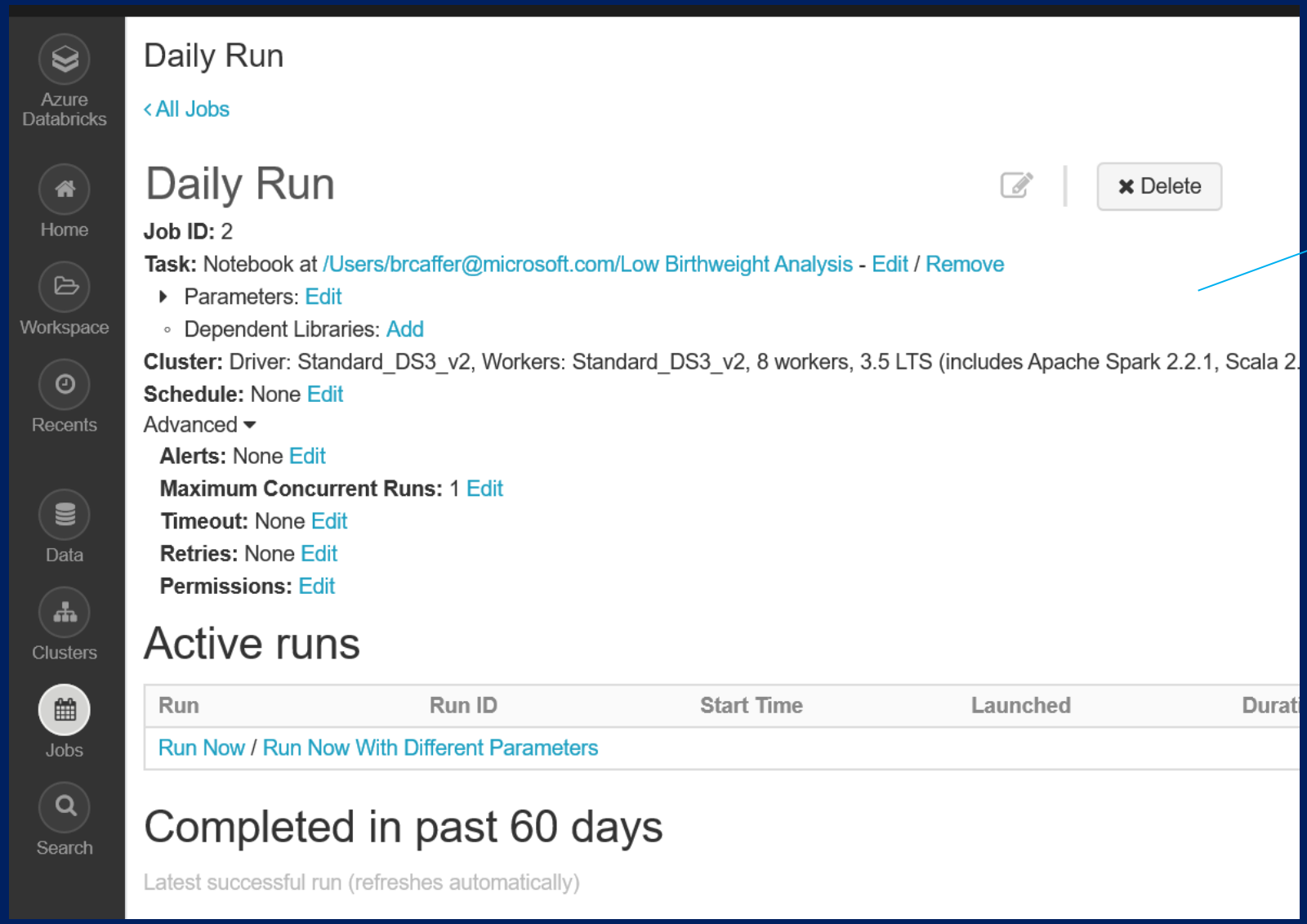
Source Code Control

Click to set configurations



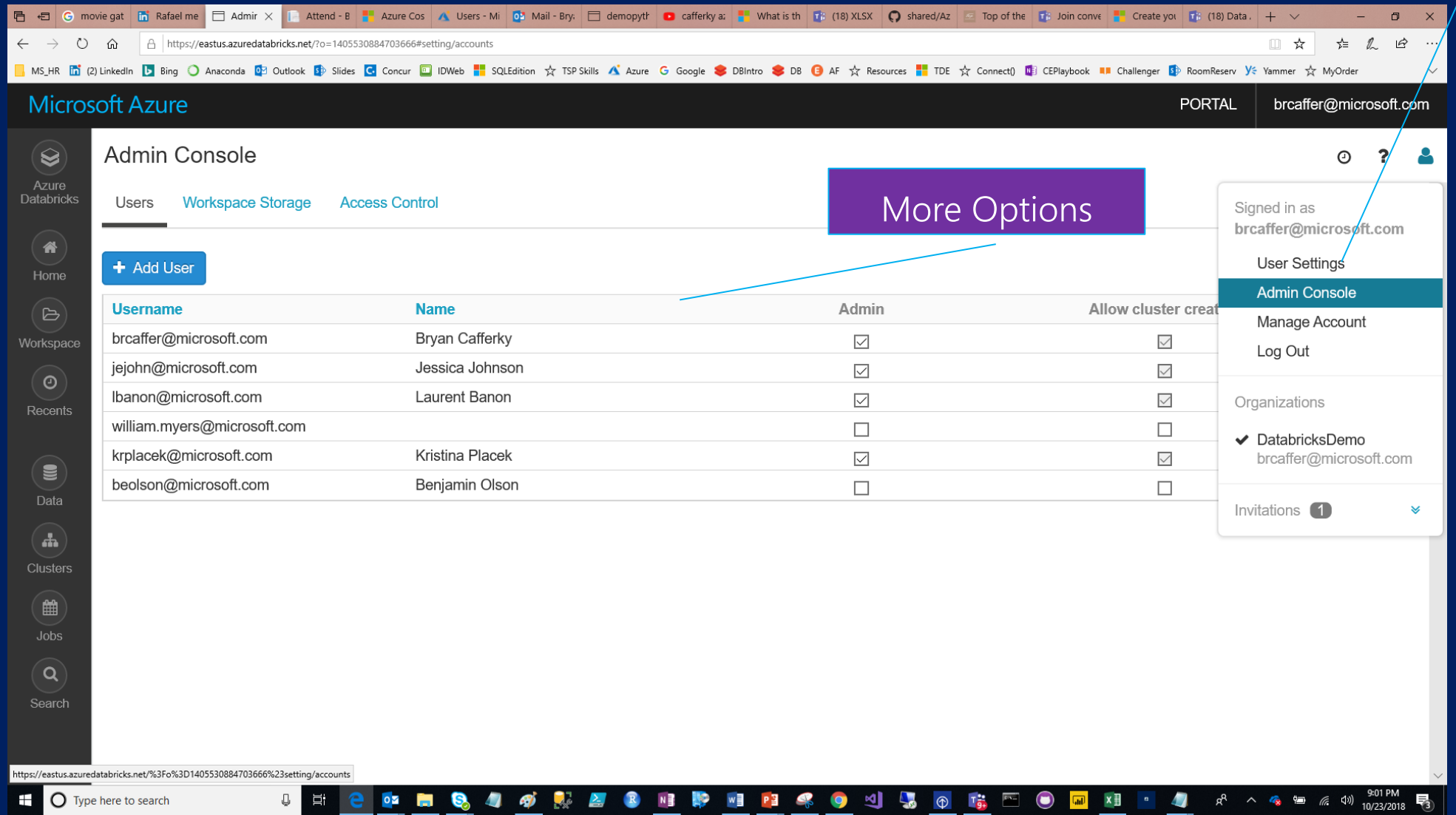
Generate Tokens for External Access

Securing Azure Databricks: Workspace

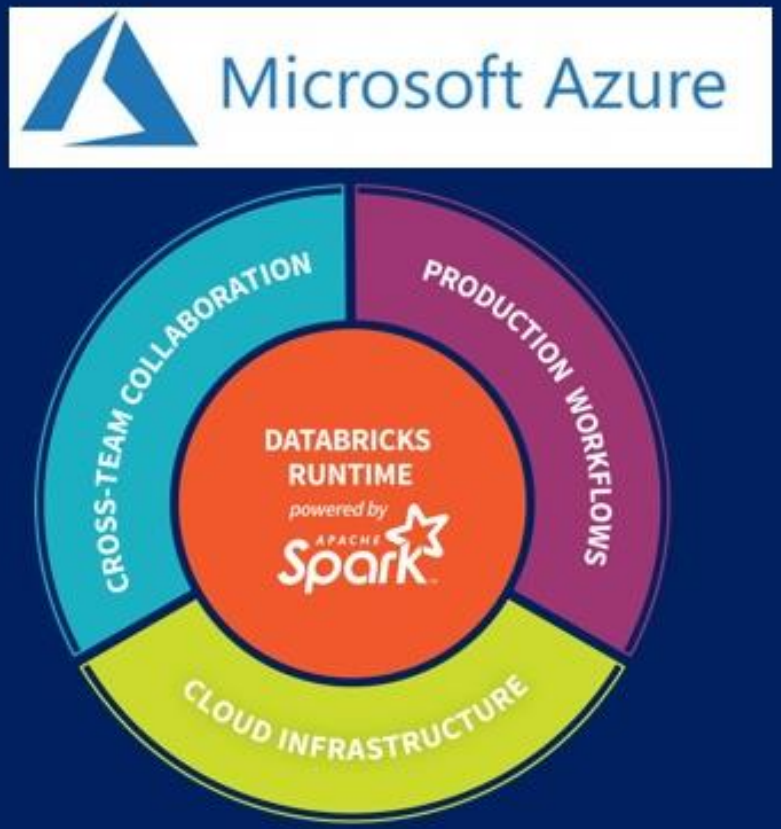


Many Configuration Options

Securing Azure Databricks: Workspace



Wrapping Up


Microsoft Azure
CROSS-TEAM COLLABORATION
PRODUCTION WORKFLOWS
DATABRICKS RUNTIME
powered by
APACHE Spark
CLOUD INFRASTRUCTURE