

Databricks

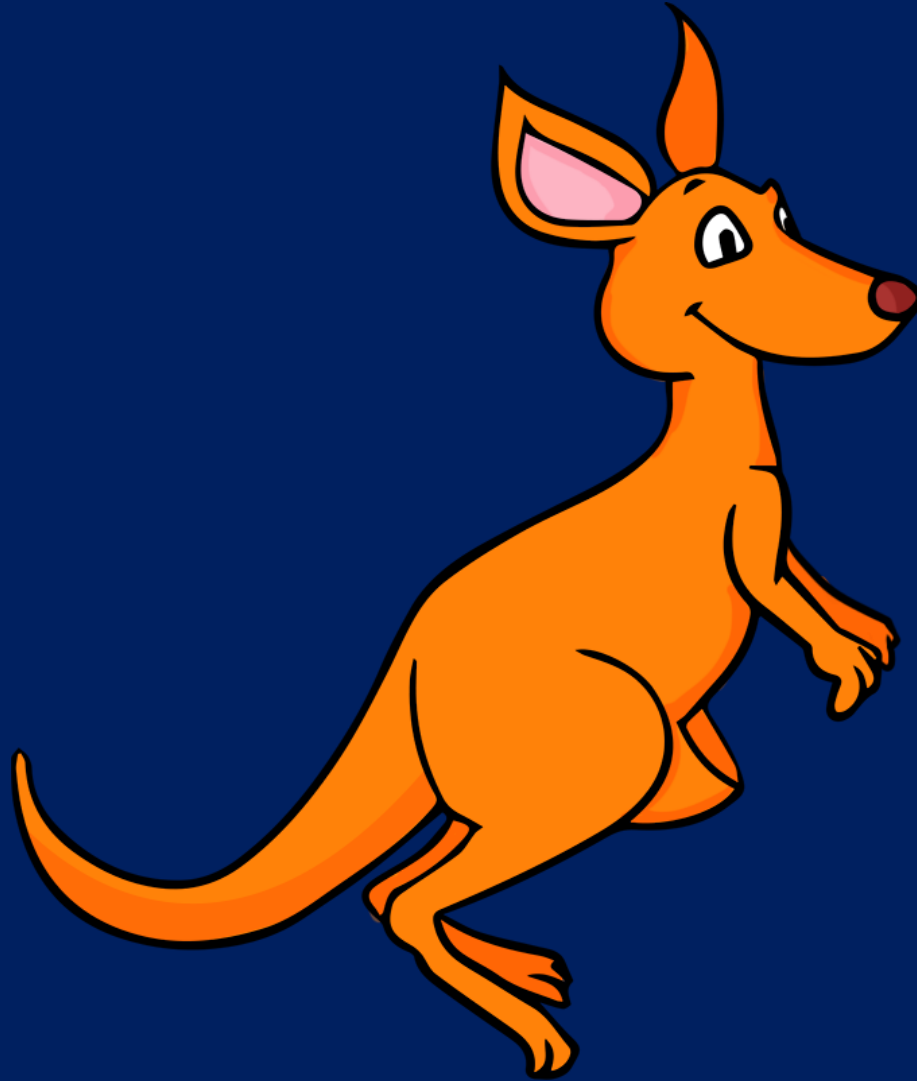
Security & Topics

Bryan Cafferky

<https://github.com/bcafferky/shared>



Jumpstart with Sample Notebooks



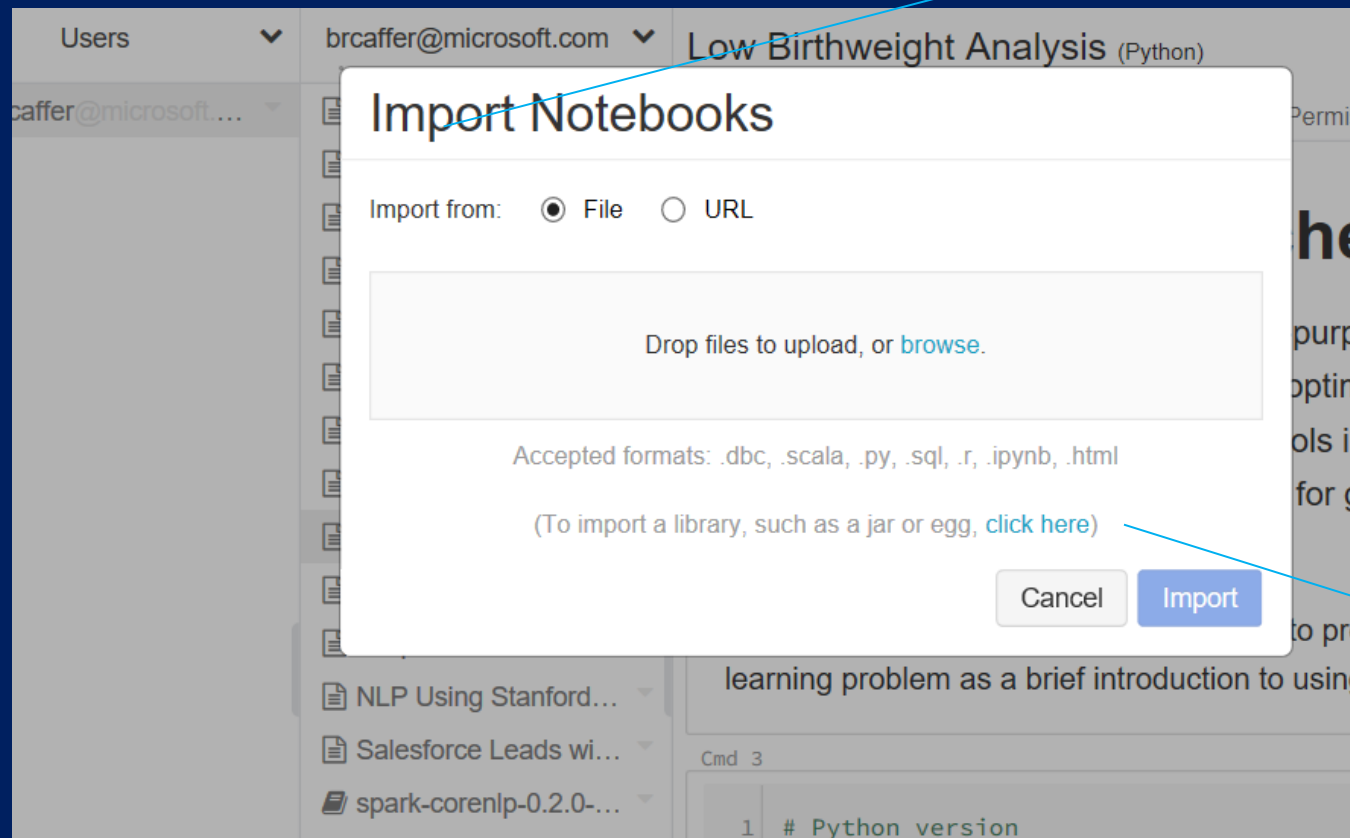
<https://clipartion.com/free-clipart-12075/>

Importing Notebooks

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

The screenshot displays the Microsoft Azure Databricks web interface. The top navigation bar includes the 'Microsoft Azure' logo and a sidebar with icons for 'Azure Databricks', 'Home', and 'Workspace'. The main content area is divided into three sections: 'Workspace', 'Users', and 'brcaffer@microsoft.com'. The 'Users' section shows a list of users, with 'brcaffer@microsoft.com' selected. A dropdown menu is open for this user, showing options: 'Create', 'Clone', 'Import' (highlighted in blue), 'Export', and 'Permissions'. The 'Import' option is the target of the instruction. The background shows a list of notebooks, including '2018-03-06 - DBFS...', '2018-03-15 - DBFS...', '2018-03-17 - DBFS...', 'Databricks for Data...', and 'Databricks for Data...'. The title of the selected notebook is 'Pop. vs. Price LR 2'.

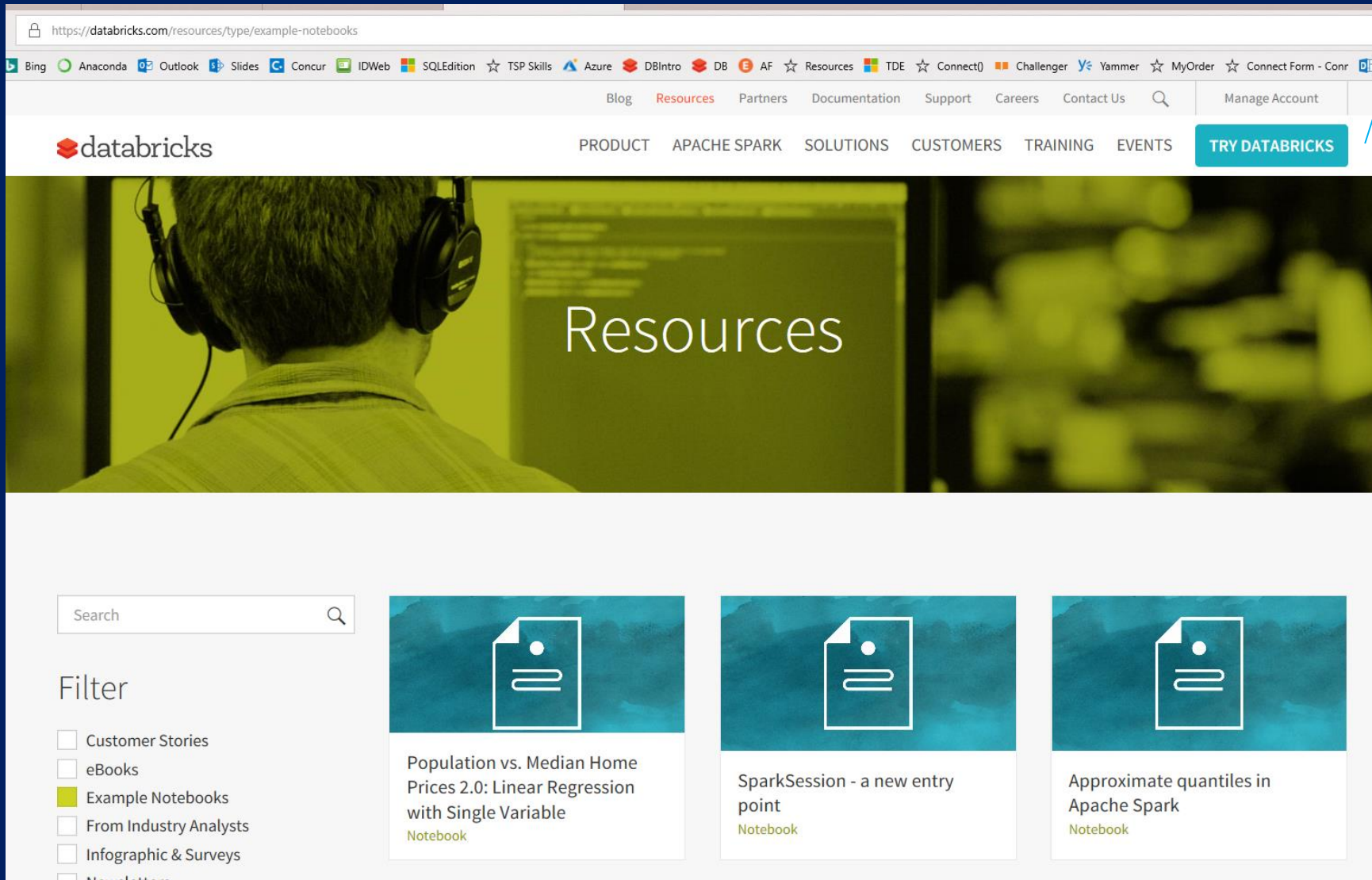
Importing Notebooks



From file or
URL

Import library

Importing Notebooks



The screenshot shows the Databricks website's 'Resources' page, specifically filtered for 'Example Notebooks'. The page features a search bar, a filter sidebar, and three notebook cards. The browser's address bar shows the URL: <https://databricks.com/resources/type/example-notebooks>. The browser's tab bar includes various links like Bing, Anaconda, Outlook, Slides, Concur, IDWeb, SQLEdition, TSP Skills, Azure, DBIntro, DB, AF, Resources, TDE, Connect(), Challenger, Yammer, MyOrder, and Connect Form - Conr. The Databricks logo is visible in the top left, and a 'TRY DATABRICKS' button is in the top right. The main heading 'Resources' is displayed over a background image of a person wearing headphones. The filter sidebar on the left includes a search bar and a list of categories: Customer Stories, eBooks, Example Notebooks (selected), From Industry Analysts, Infographic & Surveys, and Newsletters. The three notebook cards shown are: 'Population vs. Median Home Prices 2.0: Linear Regression with Single Variable', 'SparkSession - a new entry point', and 'Approximate quantiles in Apache Spark'. Each card includes a notebook icon and the word 'Notebook' in green text.

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

Search

Filter

- ☐ Customer Stories
- ☐ eBooks
- ☒ Example Notebooks
- ☐ From Industry Analysts
- ☐ Infographic & Surveys
- ☐ Newsletters

Population vs. Median Home Prices 2.0: Linear Regression with Single Variable
Notebook

SparkSession - a new entry point
Notebook

Approximate quantiles in Apache Spark
Notebook

Lot of sample notebooks to try

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

Importing Notebooks

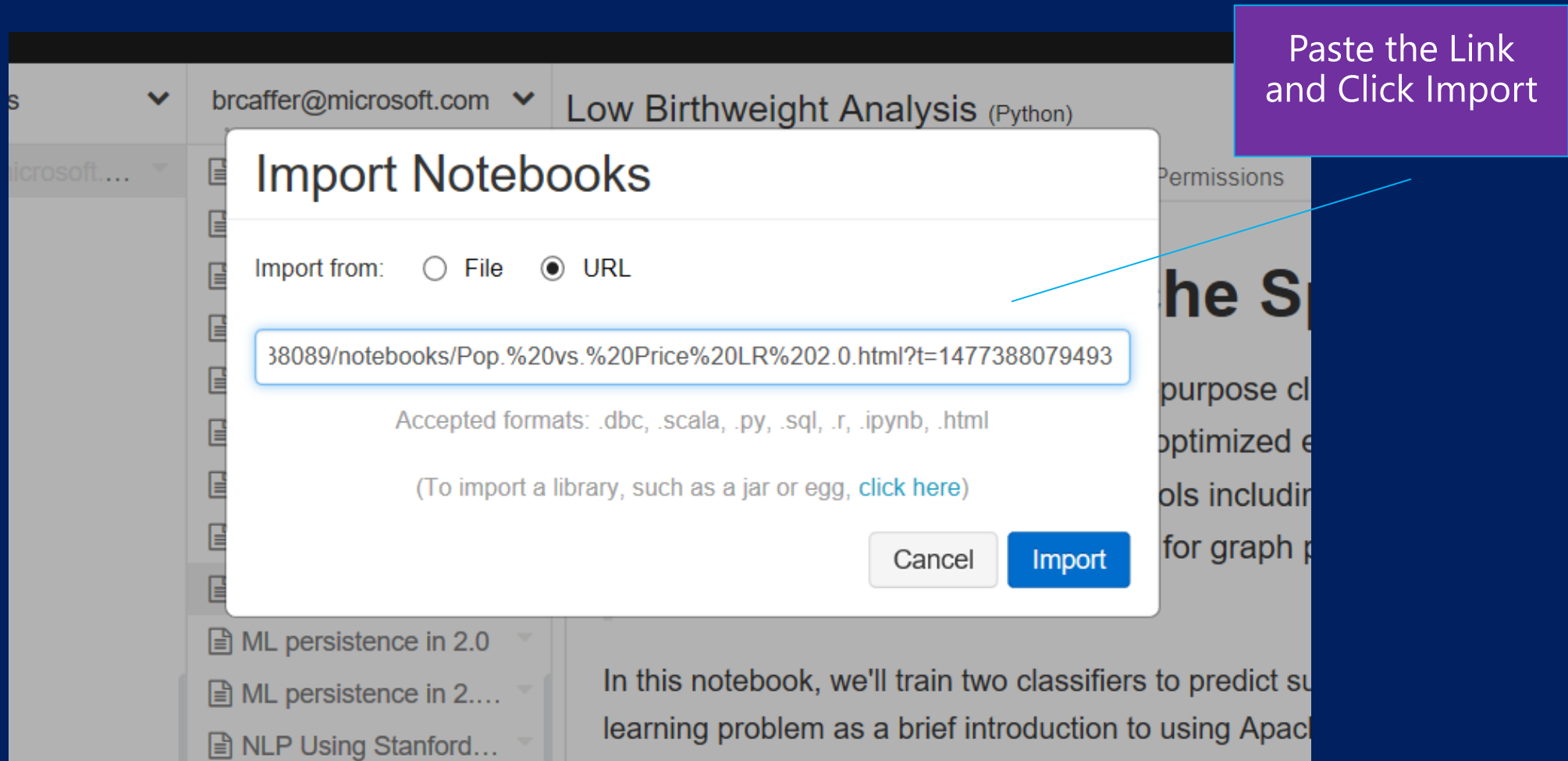
Open and Click
on the Import
Notebook
Button

The screenshot shows the Databricks web interface. At the top, the Databricks logo is on the left, and the notebook title 'Pop. vs. Price LR 2.0 (Python)' is in the center. On the right, there is an 'Import Notebook' button. Below the title, the notebook content is displayed, starting with 'Population vs. Median Home Value' and 'Linear Regression with Single Variable'. A note states 'Note, this notebook requires Spark 2.0+'. Below this, there is a code cell with Scala code: `%scala if (org.apache.spark.BuildInfo.sparkBranch < "2.0")`. An 'Import Notebook' dialog box is open in the center. It contains the text: 'You can edit and run this notebook by importing it into your Databricks account. Select **Import** from any folder's menu and paste the URL below.' Below this text is a text input field containing the URL: `https://cdn2.hubspot.net/hubfs/438089/notebooks/Pop.%20vs.%20Price%20LR`. Below the input field is a link: 'New to Databricks? [Try it now.](#)'. At the bottom right of the dialog box is a blue 'Done' button. A callout box with a blue arrow points to the URL in the input field, containing the text: 'Copy the Link to the Clipboard'.

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

Microsoft Azure

Azure Databricks

Home

Workspace

Recent

Data

Clusters

Jobs

Pop. vs. Price LR 2.0 (Python)

Detached File View: Code Permissions Run All Clear

Cmd 1

Population vs. Median Home Prices

Linear Regression with Single Variable

Cmd 2

Note, this notebook requires Spark 2.0+

Cmd 3

```
1 %scala if (org.apache.spark.BuildInfo.sparkBranch < "2.0") sys.error("Attach this notebook to a cluster with Spark 2.0+")
```

Command took 5.45 seconds -- by a user at 10/25/2016 5:32:42 AM on unknown cluster

Cmd 4

Load and parse the data

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

Securing Azure Databricks



<https://clipartion.com/free-clipart-12073/>

SECURE COLLABORATION

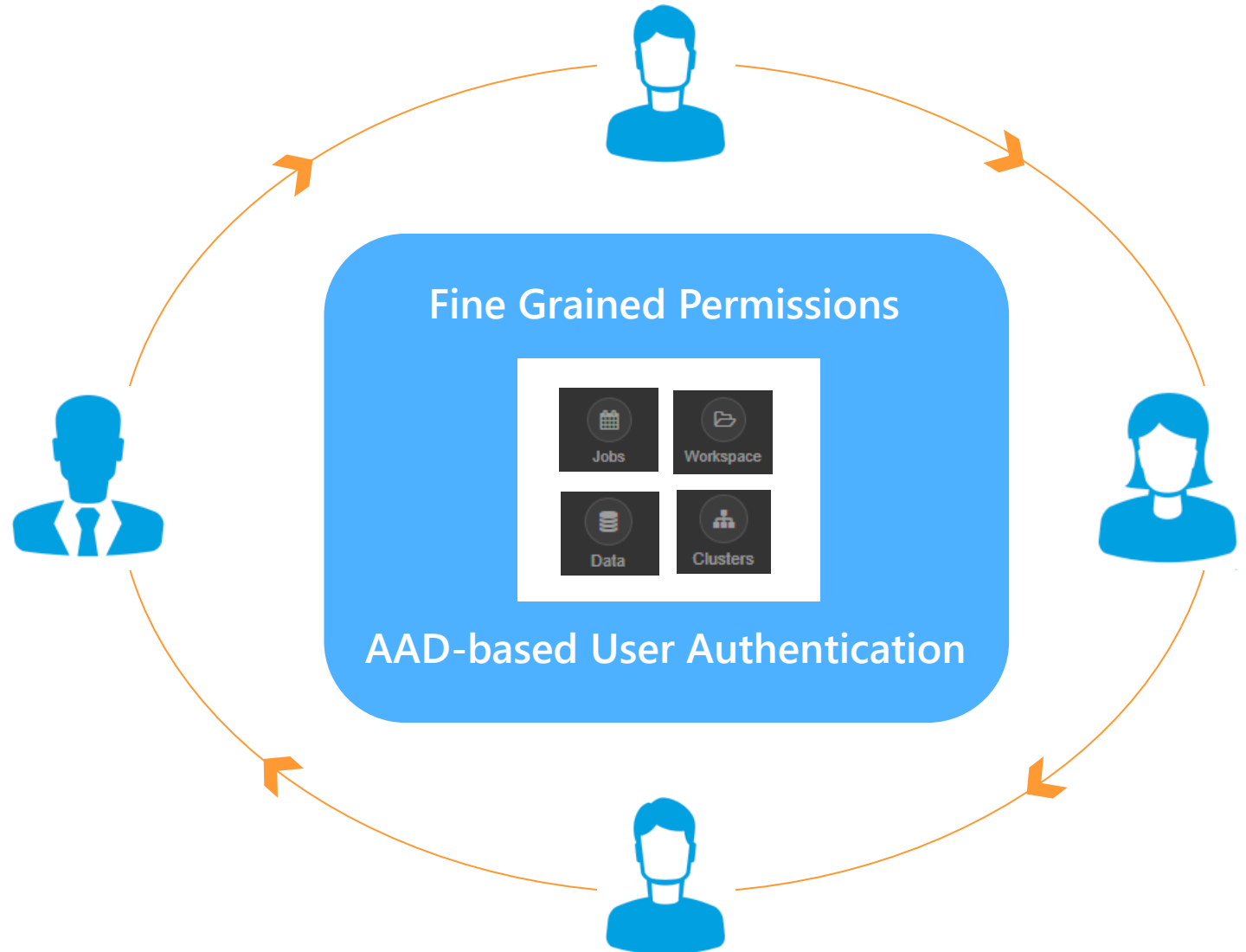
Azure Databricks enables *secure* collaboration between colleagues

- 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)



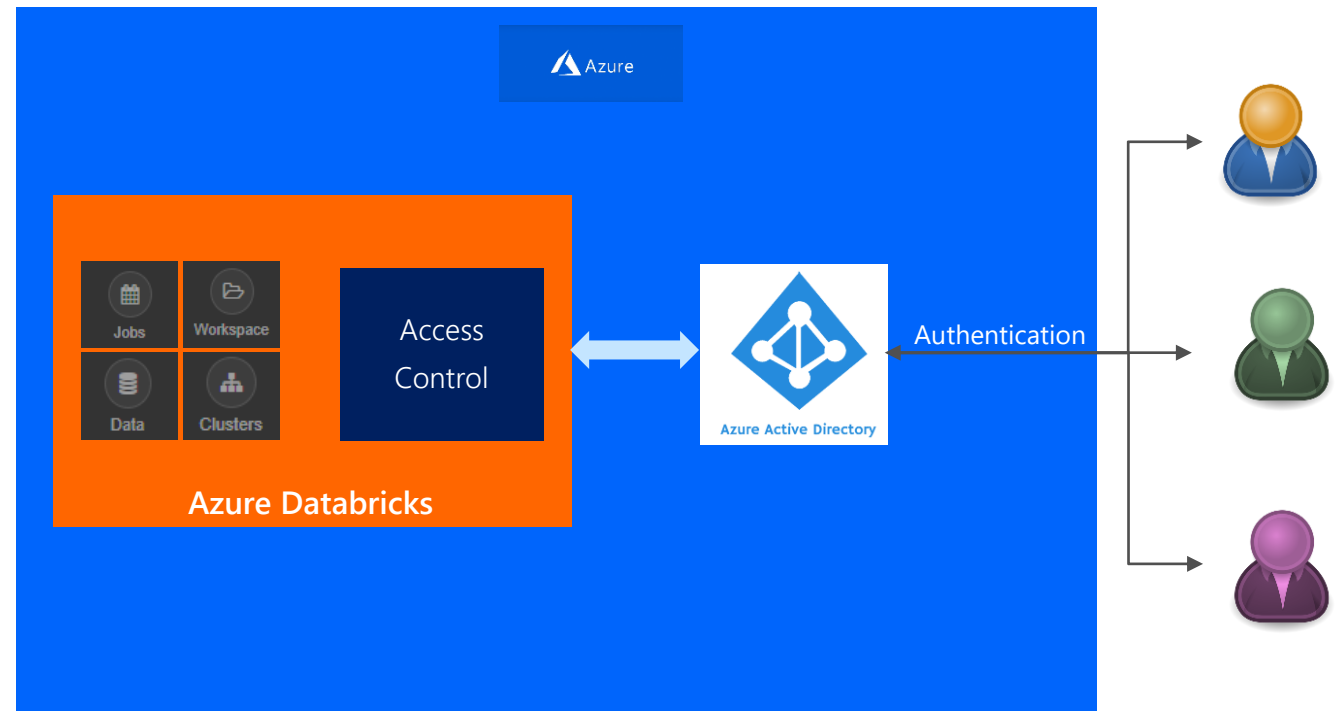
AAD-based authentication: Ensures that users are actually who they claim to be



A Z U R E D A T A B R I C K S I N T E G R A T I O N W I T H A A D

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

The screenshot shows the Microsoft Azure portal interface. At the top, there are navigation links: 'Preview', 'Microsoft Azure', and 'Report a bug'. A search bar contains the text 'Search resources, services, and docs'. On the left sidebar, there are options to 'Create a resource', view 'All services', and a 'FAVORITES' section with links to 'Dashboard', 'All resources', 'Virtual machines', 'Resource groups', 'App Services', and 'Function Apps'. The main content area is titled 'Bryan's Dashboard' and displays a list of 'All resources' under 'All subscriptions'. A 'Refresh' button is in the top right of this section. The resource list includes:

Resource Name	Resource Type
DatabricksDemo	Azure Databricks Service
blobstoragedev1	Storage account
AzureLogin	Runbook
ListVMs	Runbook
bcafferkyautomation	Automation Account
bcafferkystorage1	Storage account

A purple callout box with the text 'Open your Azure Databricks Service' has a line pointing to the 'DatabricksDemo' resource in the list.

Click to Add Users

Authorized Users

Securing Azure Databricks

The screenshot shows the 'DatabricksDemo - Access control (IAM)' page. The left sidebar contains navigation links: Overview, Activity log, Access control (IAM) (selected), Tags, Settings, Virtual Network Peerings, Locks, Automation script, Support + troubleshooting, and New support request. The main area displays a list of 9 items (5 Users, 4 Service Principals) under the 'CONTRIBUTOR' column. The list includes entries like 'bcafferkyauto1_KH9...' and 'bcafferkyautomatio...'. A modal window titled 'Add permissions' is open on the right, showing fields for 'Role' (set to 'Select a role'), 'Assign access to' (set to 'Azure AD user, group, or application'), and 'Select' (set to 'Search by name or email address'). Below these fields, a list of users is shown, including '_DataAnalytics_danalyt@microsoft.com' and '_dirsync passsync_dirsync@ptwijayakarya.onmicrosoft.com'. A 'Selected members' section at the bottom of the modal states 'No members selected. Search for and add one or more members you want to assign to the role for this resource.' and provides a link to 'learn more on our docs site'.

Home > DatabricksDemo - Access control (IAM)

DatabricksDemo - Access control (IAM)
Azure Databricks Service

Search (Ctrl+/,)

Overview
Activity log
Access control (IAM)
Tags
Settings
Virtual Network Peerings
Locks
Automation script
Support + troubleshooting
New support request

+ Add Remove Roles Refresh ?

Name *Search by name or email*
Scope *All scopes*
Type *All*
Group by *Role*

9 items (5 Users, 4 Service Principals)

☐ NAME

CONTRIBUTOR

bcafferkyauto1_KH9... App
 bcafferkyautomatio... App
 Benjamin Olson

Add permissions

Role *Select a role*
Assign access to *Azure AD user, group, or application*
Select *Search by name or email address*

_DataAnalytics_danalyt@microsoft.com
_dirsync passsync_dirsync@ptwijayakarya.onmicrosoft.com

Selected members:
No members selected. Search for and add one or more members you want to assign to the role for this resource.
[If you are new to RBAC, learn more on our docs site.](#)

Assign Role

Select User

Type of account

Securing Azure Databricks

Add permissions

Role ⓘ

Select a role ^

- Owner ⓘ
- Contributor ⓘ
- Reader ⓘ
- Azure Service Deploy Release Management Contributor ⓘ
- Log Analytics Contributor ⓘ
- Log Analytics Reader ⓘ
- Managed Applications Reader ⓘ
- masterreader ⓘ
- Monitoring Contributor ⓘ
- Monitoring Metrics Publisher ⓘ
- Monitoring Reader ⓘ
- Resource Policy Contributor (Preview) ⓘ
- User Access Administrator ⓘ

[If you are new to RBAC, learn more on our docs site.](#)

Select the Role

Securing Azure Databricks

Add permissions

Role ⓘ

Contributor

▼

Assign access to ⓘ


Azure AD user, group, or application

▼

Select ⓘ


imorr

✓

 NO PHOTO

Lara Morris
imorris@linkedin.biz

Selected members:

 Ian Morrison
imorris@microsoft.com

Remove


Save

Discard

Search for the group, user, or application

Save

Securing Azure Databricks

 **DatabricksDemo - Access control (IAM)**
Azure Databricks Service

Search (Ctrl+ /)

Overview

Activity log

Access control (IAM)

Tags

Settings

Virtual Network Peerings

Locks

Automation script

Support + troubleshooting

New support request








+ Add

Remove

Roles

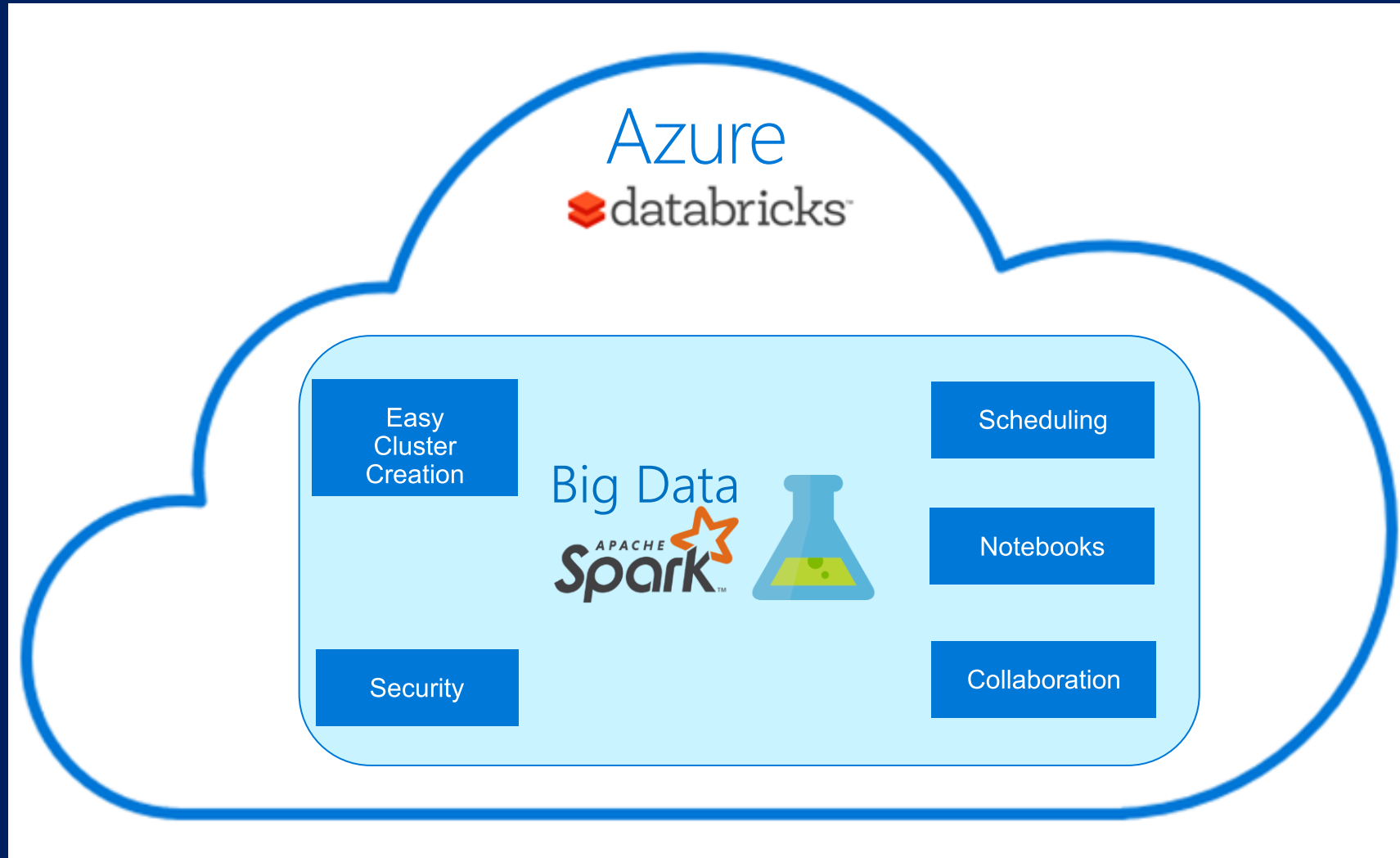
Refresh

Help

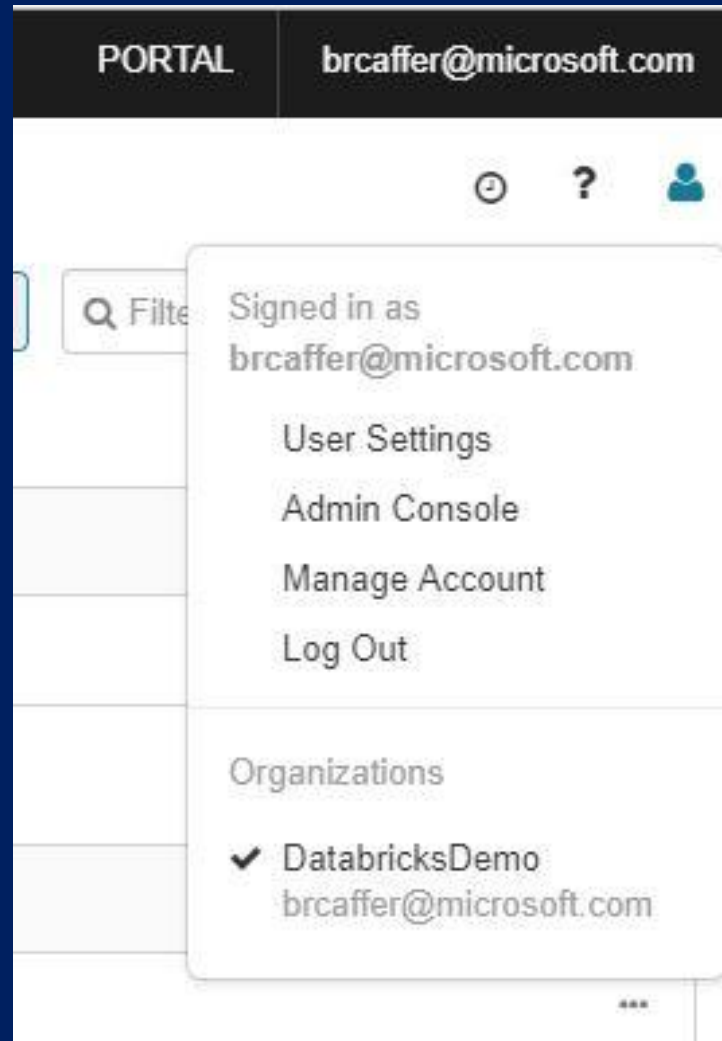
	bcafferkyautomatio...	App	Contributor	
	Benjamin Olson beolson@microsoft...	User	Contributor	
	bryandemoauto1_1...	App	Contributor	Subscription (Inherited)
	Ian Morrison imorris@microsoft.c...	User	Contributor	This resource
	Jessica Johnson jejohn@microsoft.c...	User	Contributor	This resource
	Kristina Placek krplacek@microsoft...	User	Contributor	This resource
	Laurent Banon lbanon@microsoft.c...	User	Contributor	This resource

User has access to the workspace but no objects. Object access is granted in Databricks.

Securing Azure Databricks: Workspace Objects



Administration



Administration Settings

What do
you
want to
restrict?

Settings [Internal error: report](#)

[Users](#) [Access Control](#)

Workspace Access Control: Disabled [Enable](#)

[What this means>](#)

Cluster and Jobs Access Control: Disabled [Enable](#)

[What this means>](#)

Table Access Control: Disabled

[What this means>](#)

Personal Access Tokens: Enabled [Disable](#)

[What this means>](#)

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

Microsoft Azure

PORTAL brcaffer@microsoft.com

Clusters

+ Create Cluster

All Created by me Accessible by me Filter

3 clusters, 0 pinned

▼ Interactive Clusters

Name	State	Nodes	Driver	Worker	Runtime	Creator			Actions
democlustermedium	Running	3	Standar...	Standard...	4.3 (includ...	brcaffer...	18	7	Spark UI / Logs [Stop] [Refresh] [Copy] [Lock] [Close]
clusterdemo2	Terminated	-	Standar...	Standard...	4.1 (includ...	brcaffer...	0	0	...
newcluster	Error	-	Standar...	Standard...	4.3 (includ...	brcaffer...	0	0	...






▼ Job Clusters

Name	State	Nodes	Driver	Worker	Runtime	Job Owner	Actions
job-8-run-118	Terminated	-	Standar...	Standard...	4.2 (includ...	brcaffer@microsoft.com	...
job-8-run-117	Terminated	-	Standar...	Standard...	4.2 (includ...	brcaffer@microsoft.com	...
job-8-run-116	Terminated	-	Standar...	Standard...	4.2 (includ...	brcaffer@microsoft.com	...
job-8-run-115	Terminated	-	Standar...	Standard...	4.2 (includ...	brcaffer@microsoft.com	...
job-8-run-114	Terminated	-	Standar...	Standard...	4.2 (includ...	brcaffer@microsoft.com	...

Securing Azure Databricks: Cluster Access

Permission Settings for: democlustermedium

Who has access:

 admins (group)	Can Manage ▾	
 Benjamin Olson (beolson@microsoft.com)	Can Manage ▾	✕
 Bryan Cafferky (brcaffer@microsoft.com)	Can Manage ▾	✕
 Kristina Placek (krplacek@microsoft.com)	Can Manage ▾	✕
 Laurent Banon (lbanon@microsoft.com)	Can Manage ▾	✕

Add Users and Groups:

all users (users)
(william.myers@microsoft.com)
Jessica Johnson (jejohn@microsoft.com)

Can Attach T ▾ ?

Add

Done

Assign
Permissions

Securing Azure Databricks: Notebooks

Click to set permissions

The screenshot displays the LabSQL (SQL) interface. At the top, the title 'LabSQL (SQL)' is shown. Below it, a toolbar contains several icons and labels: 'Attached: democlustermedium' with a green status indicator, 'File', 'View: Code', 'Permissions' (highlighted by a blue arrow from the 'Click to set permissions' text box), 'Run All', and 'Clear'. The notebook content is organized into two command blocks. 'Cmd 1' contains the text 'Welcome to my SQL Notebook'. 'Cmd 2' contains the text 'Upload the AdventureWorks CSV files which are pref' followed by a partially visible URL 'wineternotes.com'.

LabSQL (SQL)

Attached: democlustermedium File View: Code Permissions Run All Clear

Cmd 1

Welcome to my SQL Notebook
















Cmd 2

Upload the AdventureWorks CSV files which are pref
wineternotes.com


Securing Azure Databricks

Permission Settings for: LabSQL



Who has access:

 admins (group)	Can Manage  
 Benjamin Olson (beolson@microsoft.com)	Can Edit  
 Jessica Johnson (jejohn@microsoft.com)	Can Read  
 Kristina Placek (krplacek@microsoft.com)	Can Edit  
 Laurent Banon (lbanon@microsoft.com)	Can Edit  

Add Users and Groups:



all users (users)
(william.myers@microsoft.com)
Bryan Cafferky (brcaffer@microsoft.com)

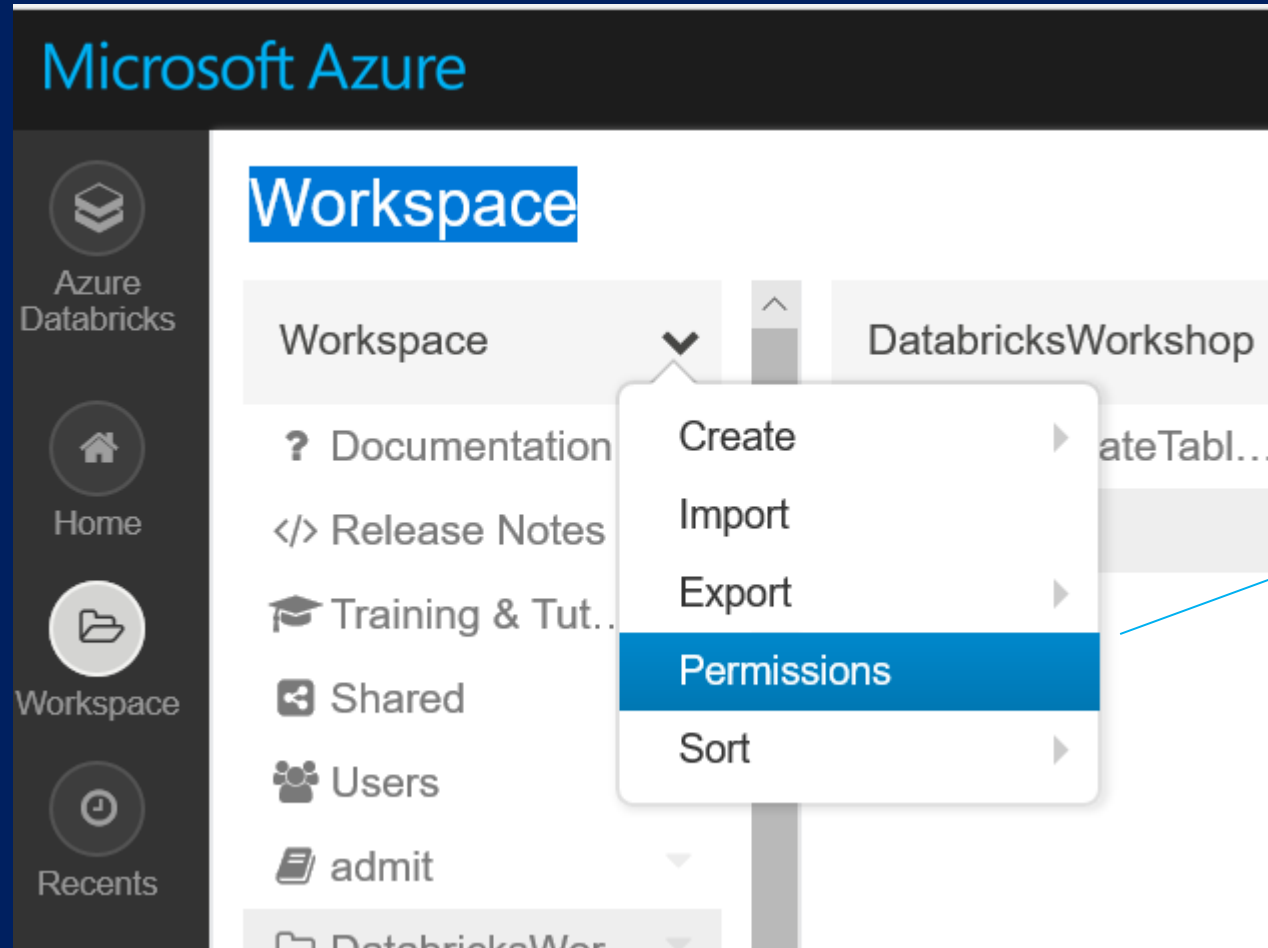
Can Read  

Add

Done

Click to set permissions

Securing Azure Databricks: Workspace



Securing Azure Databricks: User Settings

Source Code
Control

Click to set
configurations

The screenshot shows the Microsoft Azure Databricks portal interface. At the top, the 'Microsoft Azure' logo is on the left, and 'PORTAL' and the user email 'brcaffer@microsoft.com' are on the right. A left-hand navigation menu includes icons and labels for 'Azure Databricks', 'Home', 'Workspace', and 'Recents'. The main content area is titled 'User Settings' and contains three tabs: 'Access Tokens', 'Git Integration', and 'Notebook Settings'. The 'Notebook Settings' tab is selected and highlighted with a dashed border. Below the tabs, the 'Notebook Settings' section lists four configuration options with checkboxes:

- ☐ When running commands in Notebooks, automatically launch and attach to clusters without prompting
- ☐ Turn off smart quote and bracket matching
- ☐ Turn on notebook notifications
- ☒ Turn on Spark tips

Generate Tokens
for External
Access

Securing Azure Databricks: Workspace

The screenshot displays the Azure Databricks Workspace interface. On the left is a dark sidebar with navigation icons and labels: Azure Databricks, Home, Workspace, Recents, Data, Clusters, Jobs, and Search. The main content area is titled 'Daily Run' and includes a '< All Jobs' link. Below the title, there are edit and delete icons. The 'Job ID' is 2. The 'Task' is a Notebook at a specific path, with links to 'Edit' and 'Remove'. Under 'Task', there are links for 'Parameters: Edit' and 'Dependent Libraries: Add'. The 'Cluster' information is shown: Driver: Standard_DS3_v2, Workers: Standard_DS3_v2, 8 workers, 3.5 LTS (includes Apache Spark 2.2.1, Scala 2.10.6). The 'Schedule' is set to 'None' with an 'Edit' link. An 'Advanced' dropdown menu is expanded, showing several configuration options, each with an 'Edit' link: Alerts: None, Maximum Concurrent Runs: 1, Timeout: None, Retries: None, and Permissions. Below this is the 'Active runs' section, which contains a table with columns: Run, Run ID, Start Time, Launched, and Duration. The table has one row with a link 'Run Now / Run Now With Different Parameters'. At the bottom is the 'Completed in past 60 days' section, with a note 'Latest successful run (refreshes automatically)'.

Daily Run
< All Jobs

Daily Run [Edit] [Delete]

Job ID: 2

Task: Notebook at [/Users/brcaffer@microsoft.com/Low Birthweight Analysis](#) - [Edit](#) / [Remove](#)

- Parameters: [Edit](#)
- Dependent Libraries: [Add](#)

Cluster: Driver: Standard_DS3_v2, Workers: Standard_DS3_v2, 8 workers, 3.5 LTS (includes Apache Spark 2.2.1, Scala 2.10.6)

Schedule: None [Edit](#)

Advanced ▾

- Alerts: None [Edit](#)
- Maximum Concurrent Runs: 1 [Edit](#)
- Timeout: None [Edit](#)
- Retries: None [Edit](#)
- Permissions: [Edit](#)

Active runs

Run	Run ID	Start Time	Launched	Duration
Run Now / Run Now With Different Parameters				

Completed in past 60 days
Latest successful run (refreshes automatically)

Many
Configuration
Options

Securing Azure Databricks: Workspace

Admin Console

Microsoft Azure

Admin Console

Users Workspace Storage Access Control

+ Add User

Username	Name	Admin	Allow cluster creation
brcaffer@microsoft.com	Bryan Cafferky	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
jejohn@microsoft.com	Jessica Johnson	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
lbanon@microsoft.com	Laurent Banon	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
william.myers@microsoft.com		<input type="checkbox"/>	<input type="checkbox"/>
krplacek@microsoft.com	Kristina Placek	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
beolson@microsoft.com	Benjamin Olson	<input type="checkbox"/>	<input type="checkbox"/>

Signed in as brcaffer@microsoft.com

- User Settings
- Admin Console
- Manage Account
- Log Out

Organizations

- ✓ DatabricksDemo
brcaffer@microsoft.com

Invitations 1

Wrapping Up

