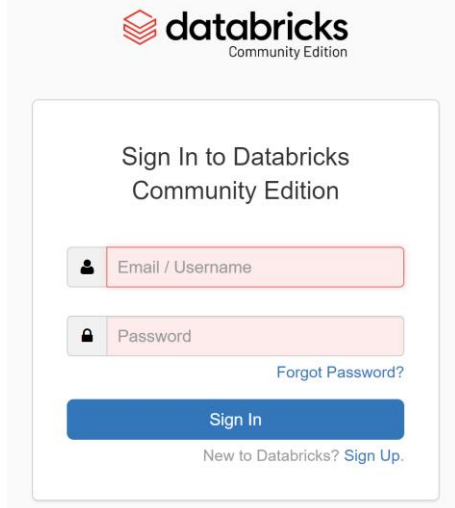


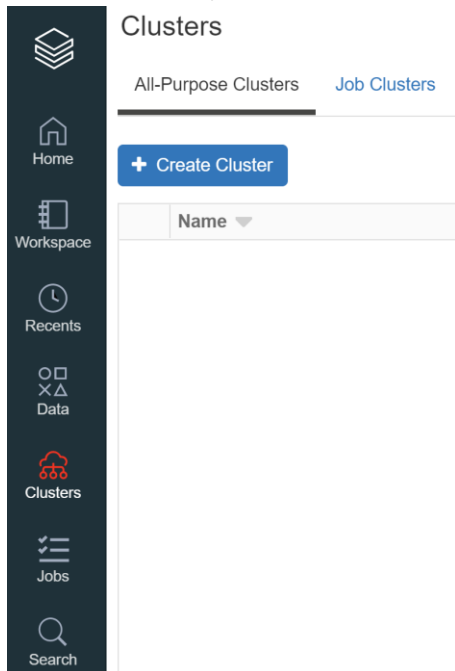
Using Databricks for Apache Spark

1. Go to <https://community.cloud.databricks.com/>



The image shows the Databricks Community Edition sign-in page. At the top is the Databricks logo with 'Community Edition' underneath. The main heading is 'Sign In to Databricks Community Edition'. Below this are two input fields: 'Email / Username' and 'Password'. A link for 'Forgot Password?' is located below the password field. A blue 'Sign In' button is positioned below the input fields. At the bottom, there is a link that says 'New to Databricks? Sign Up.'

2. Sign Up and create your account
3. Login with your account
4. On the left menu, click Clusters



5. Click Create Cluster
6. Fill the required options. For Runtime, choose 8.2 ML Beta Scala (no GPU)

Create Cluster

New Cluster

Cancel Create Cluster 0 Wc 1 Dr

Cluster Name

AIT580 Spark

Databricks Runtime Version ?

Runtime: 8.1 (Scala 2.12, Spark 3.1.1)

Databricks Runtime

8.2 Beta	Scala 2.12, Spark 3.1.1
8.2 ML Beta	GPU, Scala 2.12, Spark 3.1.1
8.2 ML Beta	Scala 2.12, Spark 3.1.1
8.1	Scala 2.12, Spark 3.1.1
8.1 ML	GPU, Scala 2.12, Spark 3.1.1

[23 more](#)

Availability Zone ?

us-west-2c

7. Click Create Cluster
8. You will see the screen below. It will take some time to get the cluster ready.

Clusters /

AIT580 Spark

Edit Clone

Configuration Notebooks Libraries Event Log Spark UI

Databricks Runtime Version

8.1 (includes Apache Spark 3.1.1, Scala 2.12)

Driver Type

Community Optimized 15.3 GB Memory, 2 Cores, 1 DBU

Instance

Free 15GB Memory: As a Community Edition user, your cluster will au
For [more configuration options](#), please [upgrade your Databricks subs](#)

Instances Spark JDBC/ODBC Permissions

Availability Zone ?

us-west-2c

9. You will see green dot when cluster is ready. Click on the cluster again and make sure the status is ready

Clusters

All-Purpose Clusters Job Clusters

[+ Create Cluster](#)

Name ▼	State ▼
AIT580 Spark	Running

10. On the left, click on Data, and then click on Create Table

Create New Table

Data source ?

Upload File S3 DBFS Other Data Sources Partner Integrations

DBFS Target Directory ?

/FileStore/tables/ (optional) [Select](#)

Files uploaded to DBFS are accessible by everyone who has access to this workspace

Files ?

Drop files to upload, or [browse](#).

11. Click Browse and upload your data file. Make sure to note the File Uploaded Path. This path will be used in your Spark Notebook for reading data

Learn more'. At the bottom is a 'Files' section with a large grey box. Inside the box, a file named 'SalesData.txt' is shown with a green checkmark, a size of '0.1 MB', and a 'Remove file' link. Below the box, a message states: '✓ File uploaded to /FileStore/tables/SalesData-1.txt'. At the bottom are two buttons: 'Create Table with UI' and 'Create Table in Notebook'."/>

Create New Table

Data source ?

Upload File S3 DBFS Other Data Sources Partner Integrations

DBFS Target Directory ?

/FileStore/tables/ (optional) [Select](#)

Files uploaded to DBFS are accessible by everyone who has access to this workspace. [Learn more](#)

Files ?

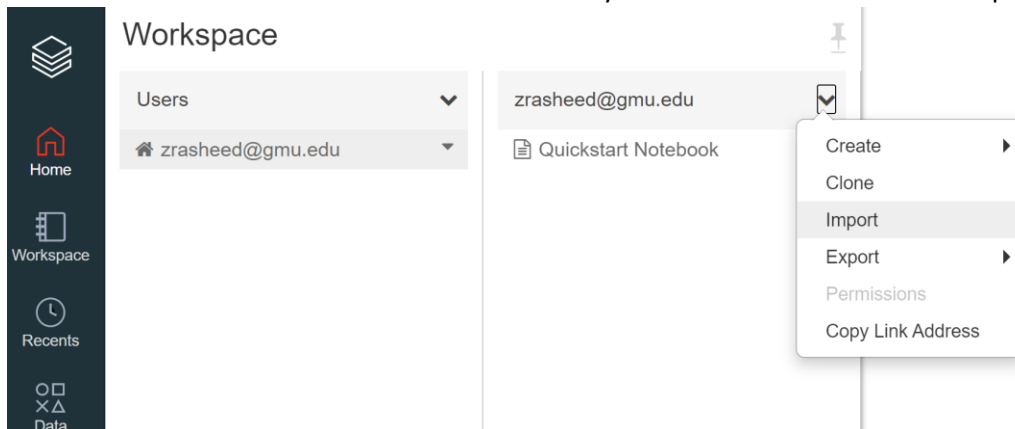
SalesData.txt ✓

0.1 MB
[Remove file](#)

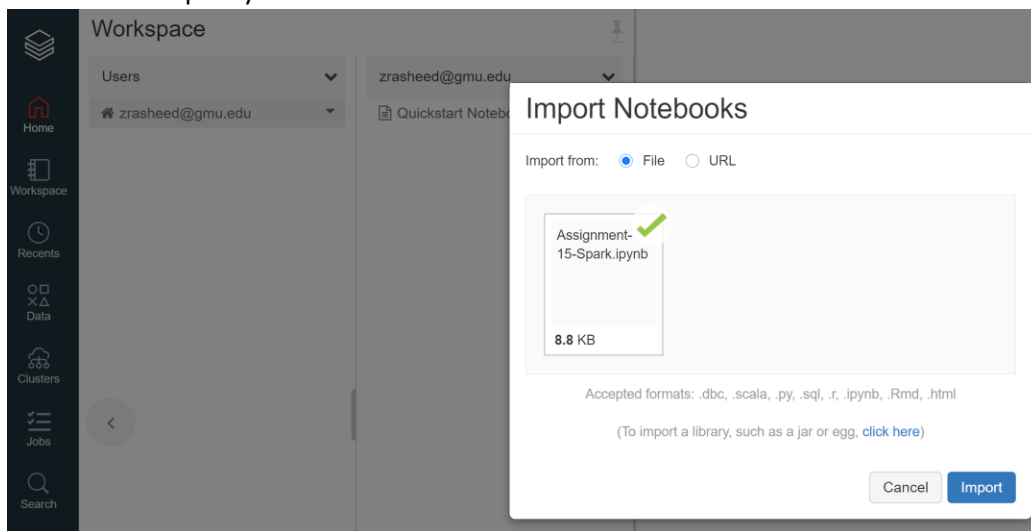
✓ File uploaded to /FileStore/tables/SalesData-1.txt

[Create Table with UI](#) [Create Table in Notebook](#) ?

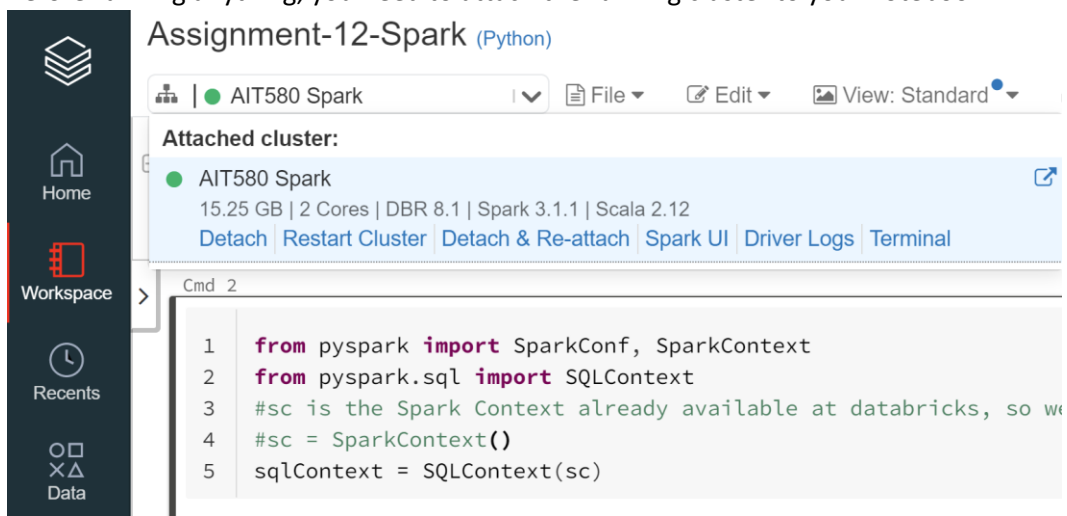
12. Click on Home. Click on down arrow at the end of your username and then click Import



13. Browse and Import your Notebook.



14. Your notebook will open. If not, then you can always click on Workspace on left menu and open the notebook.
15. Before running anything, you need to attach the running cluster to your notebook



16. Happy Coding

