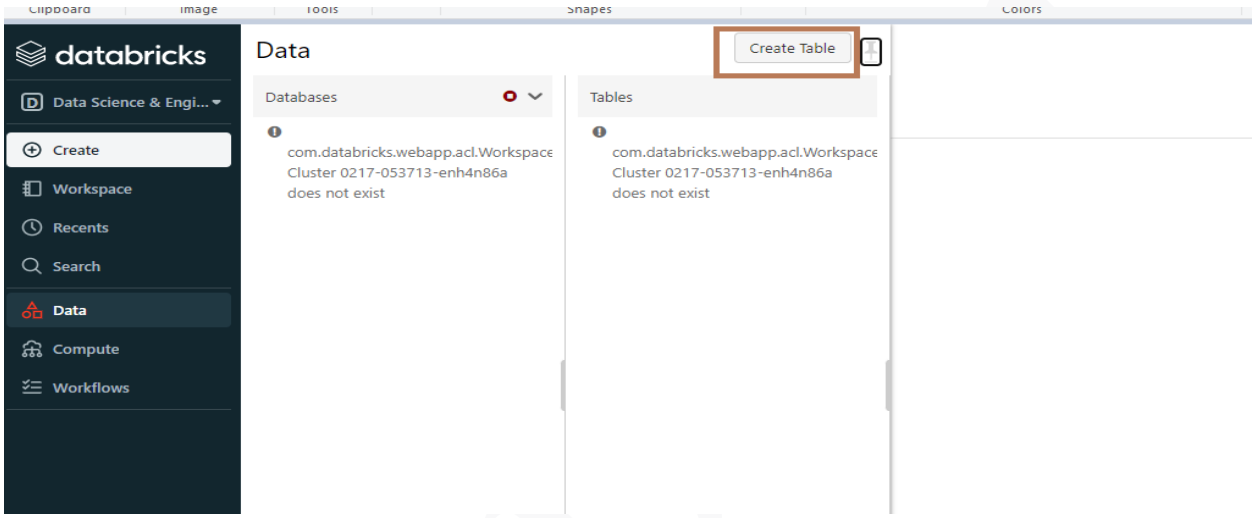
Working with Databricks

Data Loading To DBFS (Databricks File System)

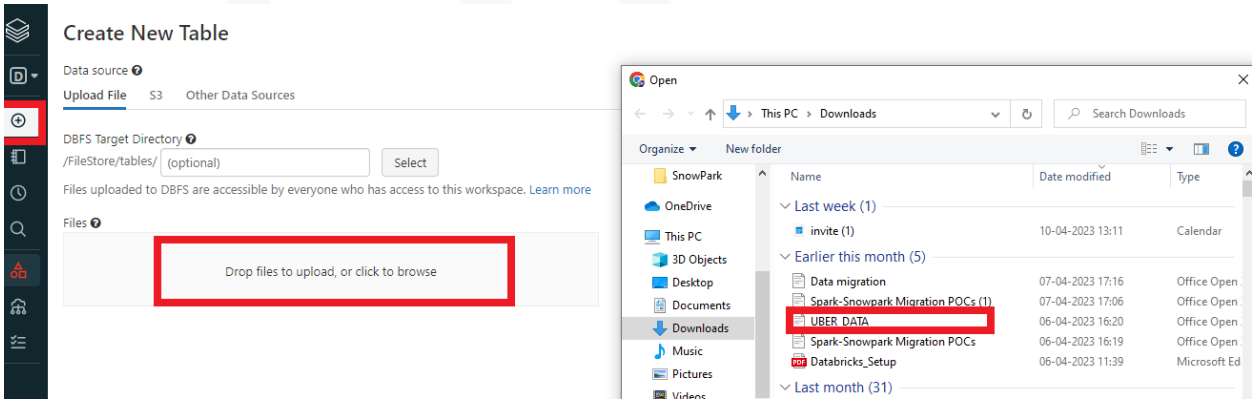
The Databricks File System (DBFS) is a distributed file system mounted into a Databricks workspace and available on Databricks clusters.

Data loading from local system to DBFS

1. Select data tab on the left panel and click on *Create Table*



2. click on *Upload File* tab and upload file from local to Databricks

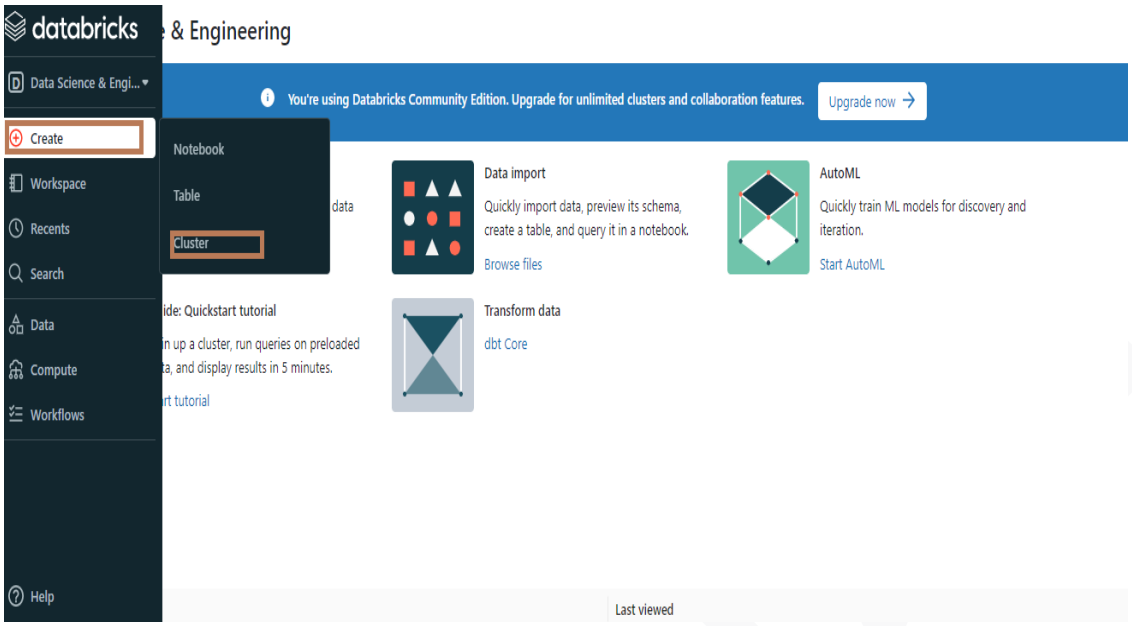


Cluster creation

Create a new cluster in Databricks or use an existing cluster.

Before creating a new cluster, check for existing clusters in the **Clusters** tab of the Databricks portal. If there is an existing cluster, you can restart the cluster.

1. Click **Create** and choose **cluster** as shown in the image below:

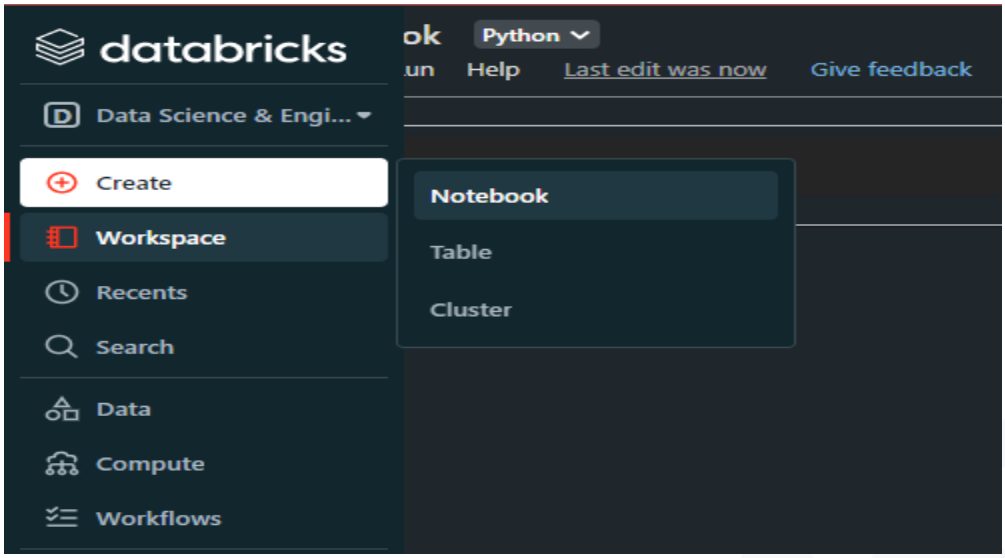


2. Enter a name for the cluster.
   You can continue with the default values for Worker type and Driver type.
3. Click **Create Cluster** and wait for the cluster to be up.

NOTE: you are using an existing cluster, make sure that the cluster is up and running.

Notebook Creation

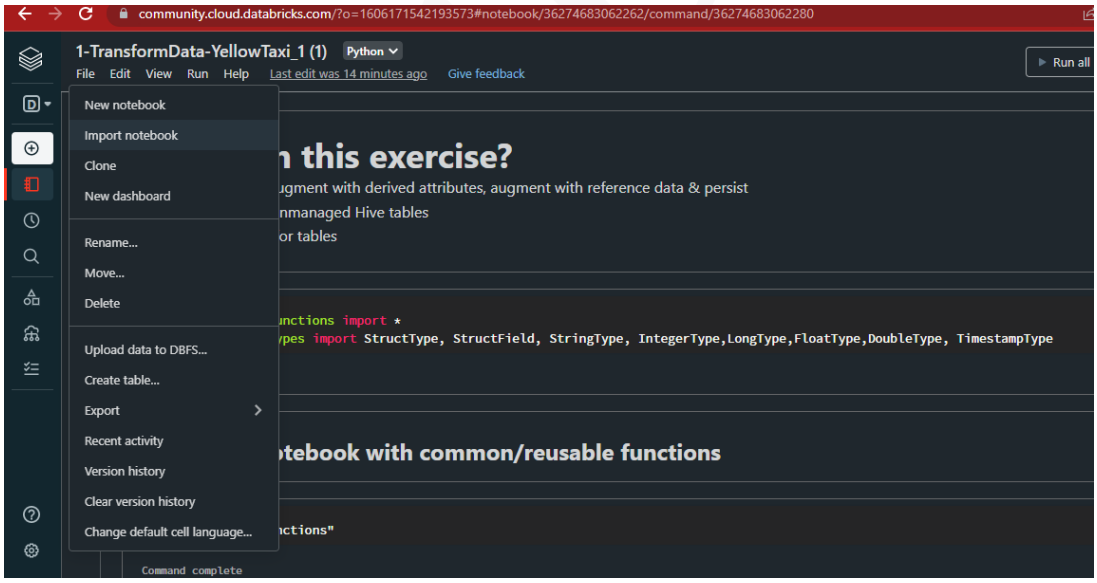
1. Click ⊕ (Create) in the sidebar and select Notebook from the menu. The Create Notebook dialog appears.

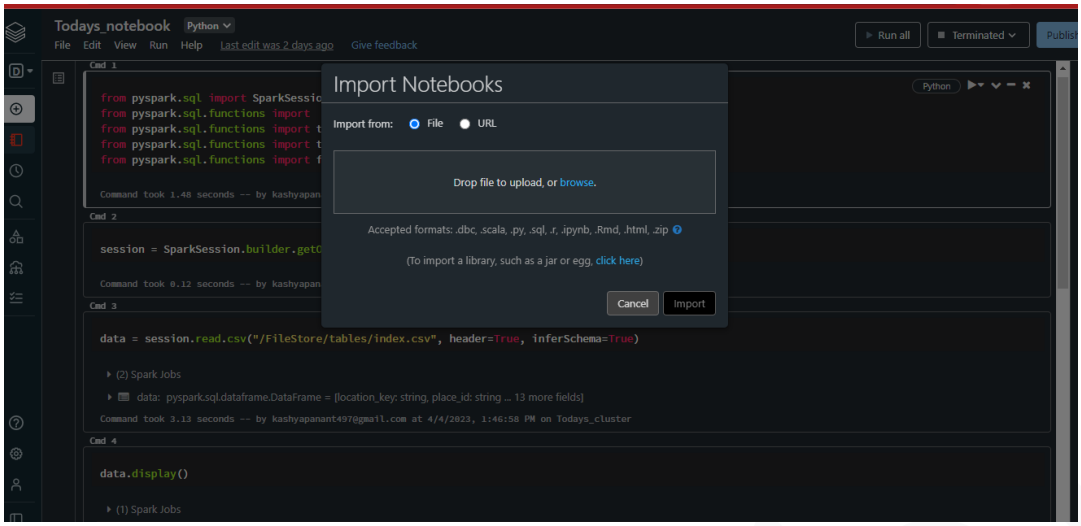


2. Enter a name and select the notebook's default language.
3. If there are running clusters, the **Cluster** drop-down displays. Select the cluster you want to attach the notebook to.
4. Click Create.

IMPORTING SAMPLE CODE FROM LOCAL

1. To import code from local you can go to FILE→IMPORT NOTEBOOK→FILE UPLOAD
And you are ready to execute the code

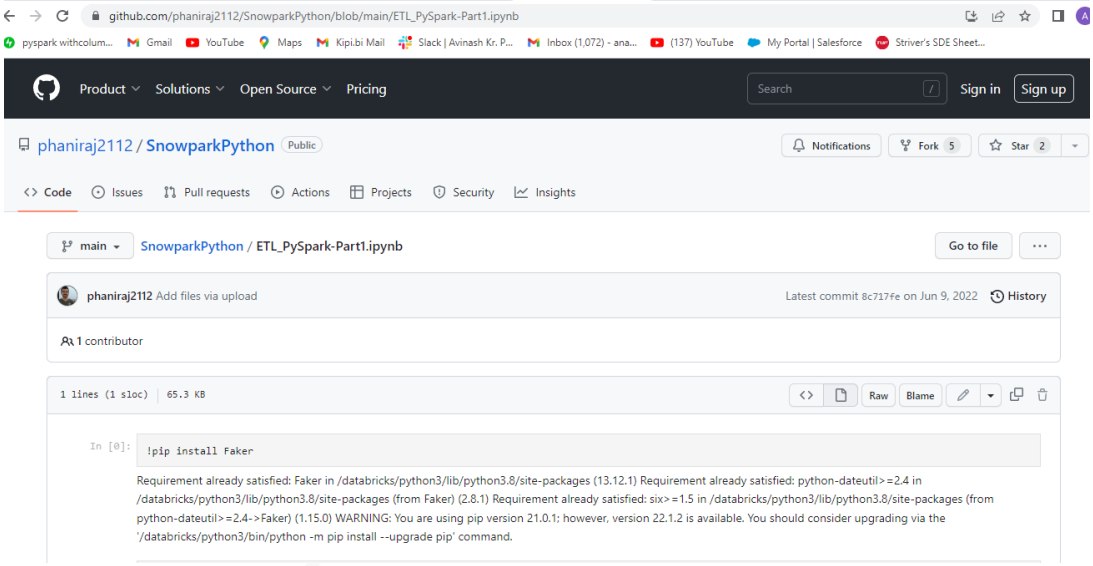




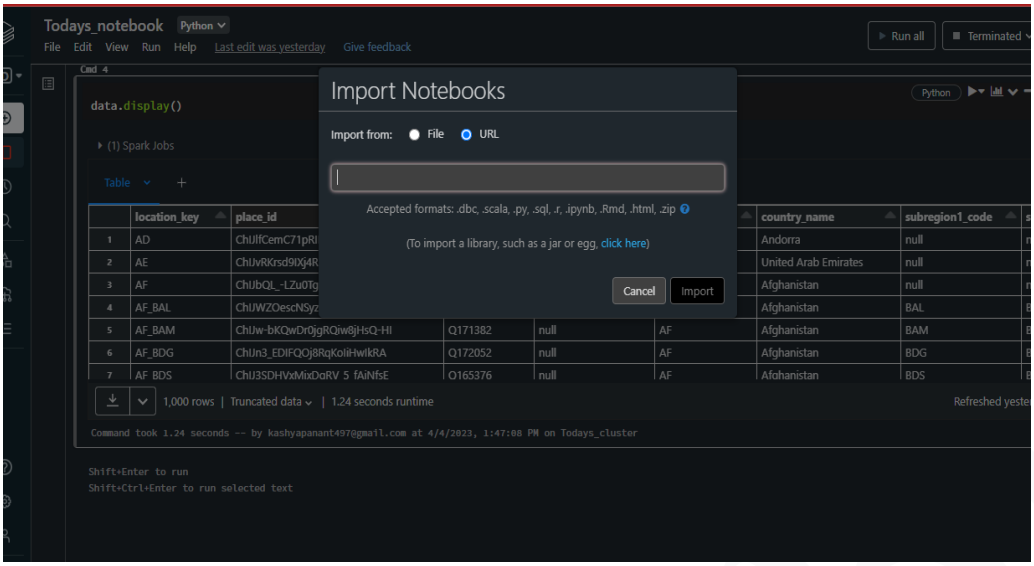
Now you just need to press the run button in each cell and execute code.

IMPORTING SAMPLE CODE FROM GITHUB

1. Go to github repository and copy the URL of the notebook containing the code



2. To import code directly from github, go to FILE→Import notebook→URL (put the URL of notebook) and click ***import***.


Todays_notebook
Python
File Edit View Run Help Last edit was yesterday Give feedback
Run all
Terminated
Cmd 4
data.display()
(1) Spark Jobs
Table
Import Notebooks
Import from: File URL
Accepted formats: .dbc, .scala, .py, .sql, .r, .ipynb, .Rmd, .html, .zip
(To import a library, such as a jar or egg, click here)
Cancel
Import
location_key
place_id
country_name
subregion1_code
1 AD Andorra null
2 AE United Arab Emirates null
3 AF Afghanistan null
4 AF_BAL Afghanistan BAL
5 AF_BAM ChIJw-bKQwDr0jgRQiw8jHsQ-HI Q171382 null AF Afghanistan BAM
6 AF_BDG ChIJn3_EDlFQOj8RqKoliHwlkRA Q172052 null AF Afghanistan BDG
7 AF_BDS ChIJ3SDHVxMixDoRV_5_fAiNfsE Q165376 null AF Afghanistan BDS
1,000 rows | Truncated data | 1.24 seconds runtime
Refreshed yeste
Command took 1.24 seconds -- by kashyapanant497@gmail.com at 4/4/2023, 1:47:08 PM on Todays_cluster
Shift+Enter to run
Shift+Ctrl+Enter to run selected text