CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Big Data & Analytics

Databricks: A Tutorial

Dr. Ignacio Castineiras
Department of Computer Science

Databricks: An Online Platform for Data Engineers

- Databricks is a spin-off company leveraging Spark as a **cloud-based** big data processing tool https://databricks.com/

Databricks: An Online Platform for Data Engineers

- The Databricks Community Edition is free to use.
  It will be the approach we will follow.

Databricks: An Online Platform for Data Engineers

- The Databricks Community Edition is free to use.
  It will be the approach we will follow.

- The Community Edition provides us with a **local mode**-based
  Spark configuration (similar to using Spark in our own local machine).

Databricks: An Online Platform for Data Engineers

- The Databricks Community Edition is free to use.
  It will be the approach we will follow.

- The Community Edition provides us with a **local mode**-based
  Spark configuration (similar to using Spark in our own local machine).

- Specifically, it provides us with **A Cluster of 1 Databricks Unit (DBU)**:
  "A unit of processing capability per hour"

  - 1 physical machine with
  - 2 cores and
  - 8GB of memory storage.

Databricks: An Online Platform for Data Engineers

- On the other hand, the Databricks Enterprise Edition allows us to rent a cluster in a price-per-usage basis, and configure it to target the Business Unit specific needs.

Databricks: An Online Platform for Data Engineers

- On the other hand, the Databricks Enterprise Edition allows us to rent a cluster in a price-per-usage basis, and configure it to target the Business Unit specific needs.

- The cluster is indeed hosted by Microsoft Azure or Amazon Web Services, with Databricks providing the interface to automate the cluster setup.

Databricks: An Online Platform for Data Engineers

- On the other hand, the Databricks Enterprise Edition allows us to rent a cluster in a price-per-usage basis, and configure it to target the Business Unit specific needs.

- The cluster is indeed hosted by Microsoft Azure or Amazon Web Services, with Databricks providing the interface to automate the cluster setup.

- The range of prices depends on the functionality provided by Databricks, and can be consulted at: https://databricks.com/product/pricing#dbu

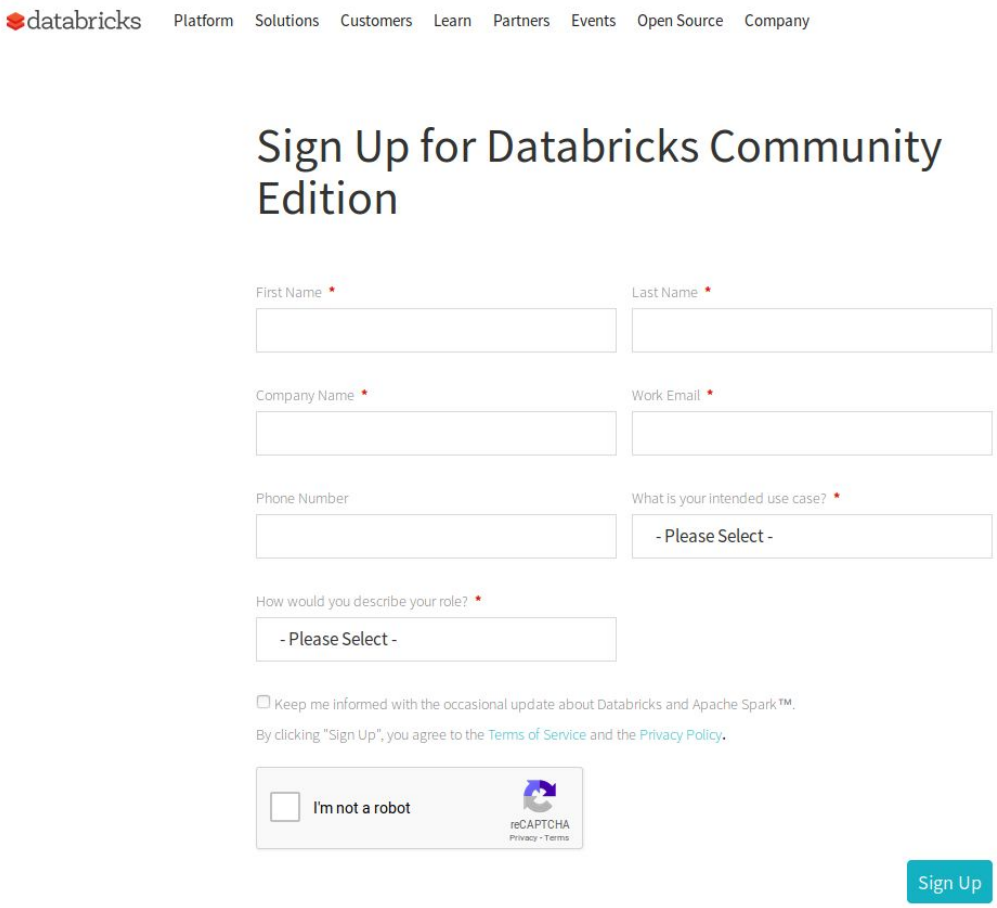
Databricks: An Online Platform for Data Engineers

How to...

Sign up / sign in.

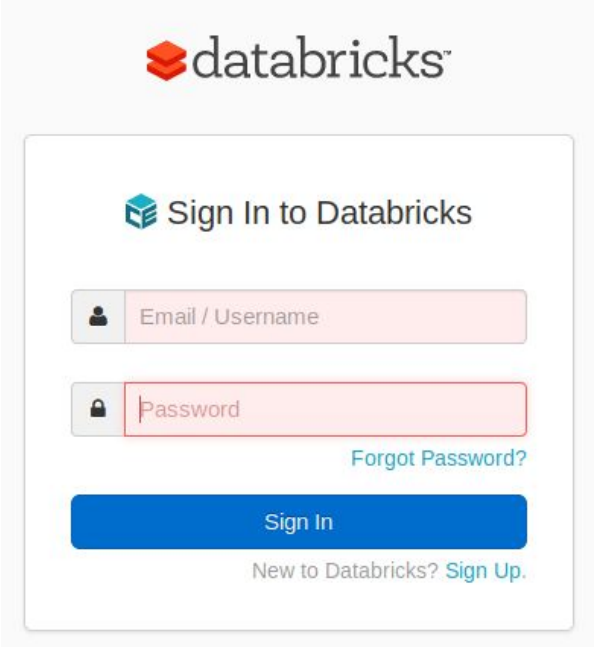
Databricks: An Online Platform for Data Engineers

1. We register/sign up to Databricks Community Edition at: https://databricks.com/signup/signup-community



Databricks: An Online Platform for Data Engineers

2. Once registered, we sign in at: https://community.cloud.databricks.com



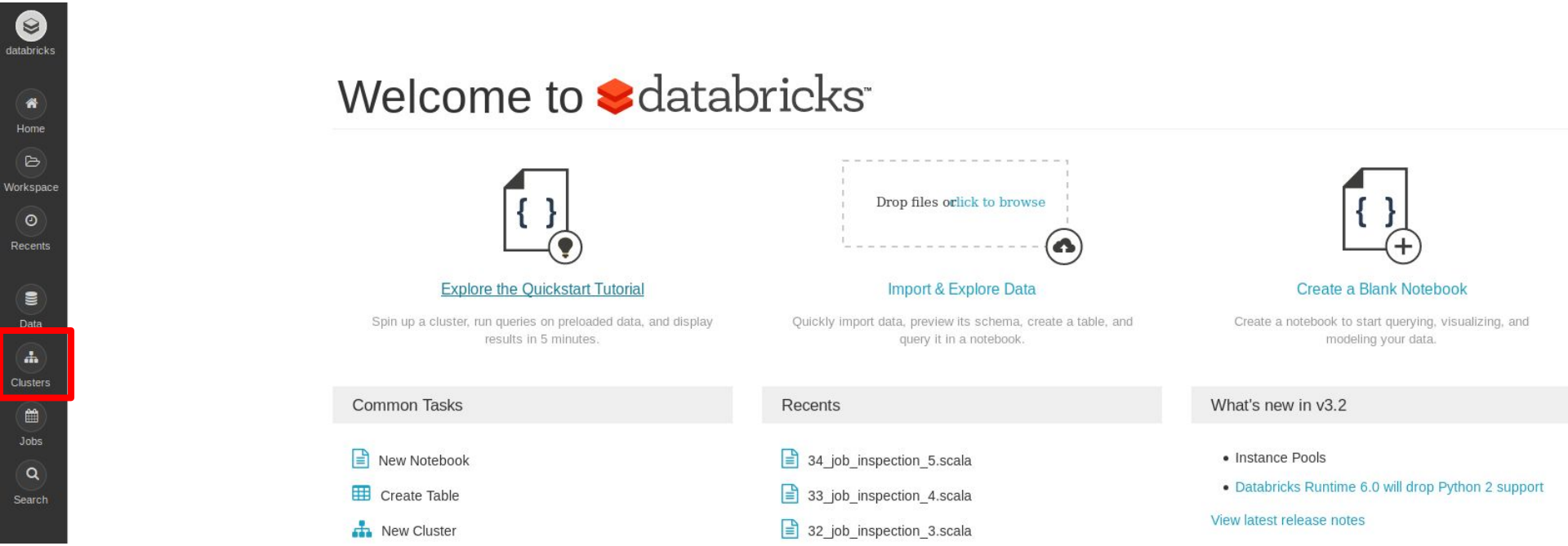
Databricks: An Online Platform for Data Engineers

How to...

Create a new cluster.

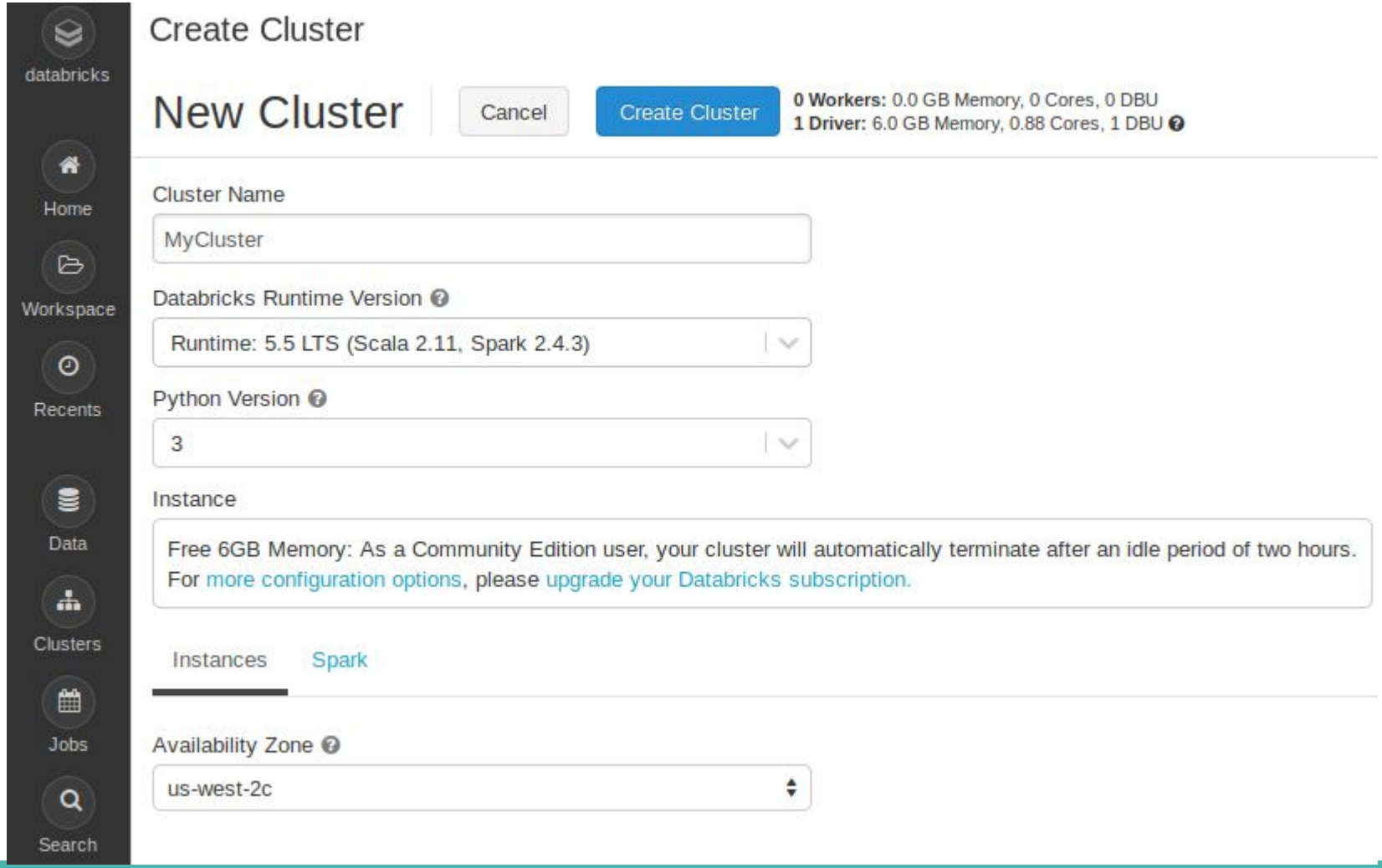
Databricks: An Online Platform for Data Engineers

3. We are redirected to the main page of the web interface.
   Our first step is to set up a new Community Edition cluster by clicking in **Clusters**.



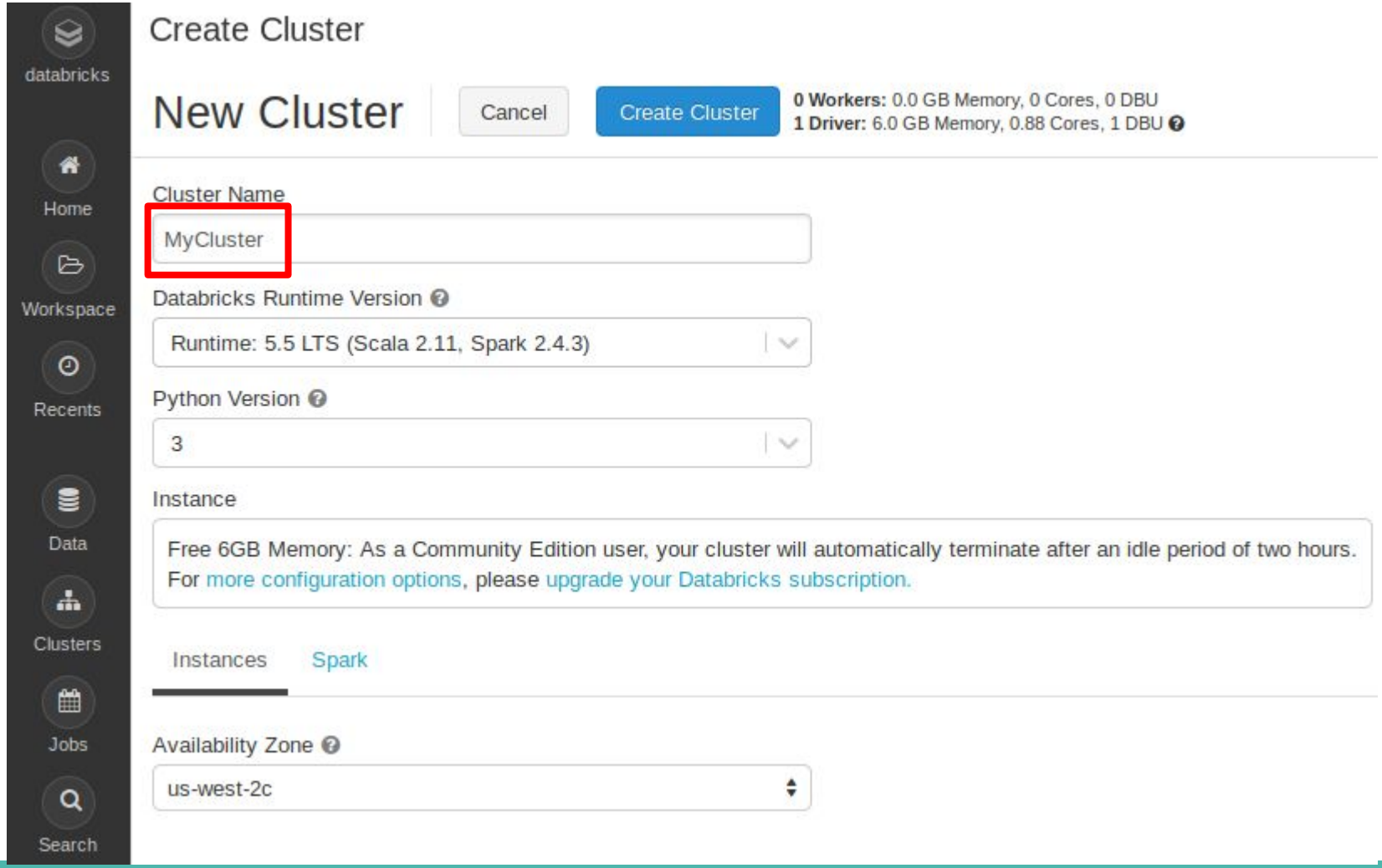
Databricks: An Online Platform for Data Engineers

3. We set up the new cluster with the following parameters:



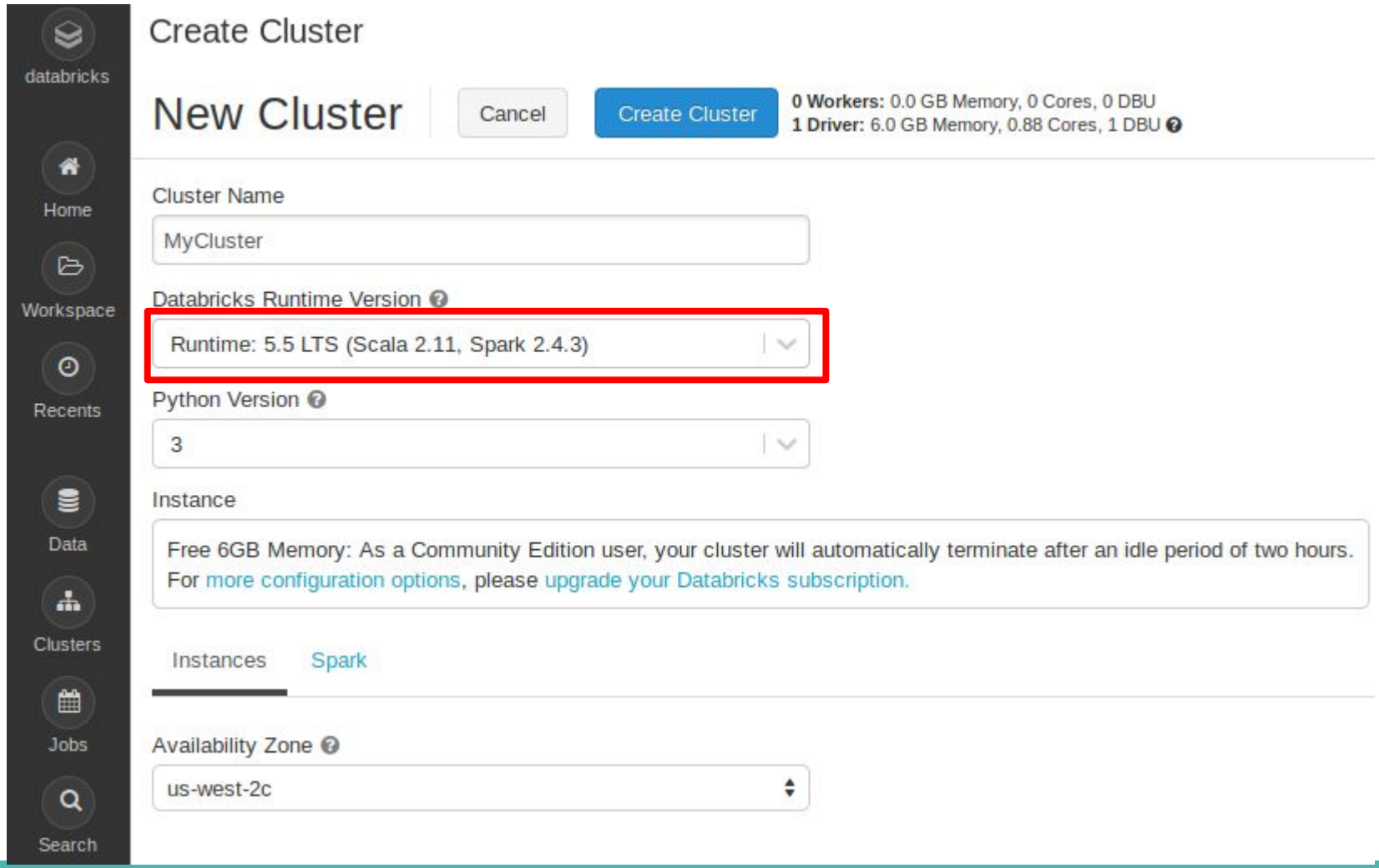
Databricks: An Online Platform for Data Engineers

3. We set up the new cluster with the following parameters:



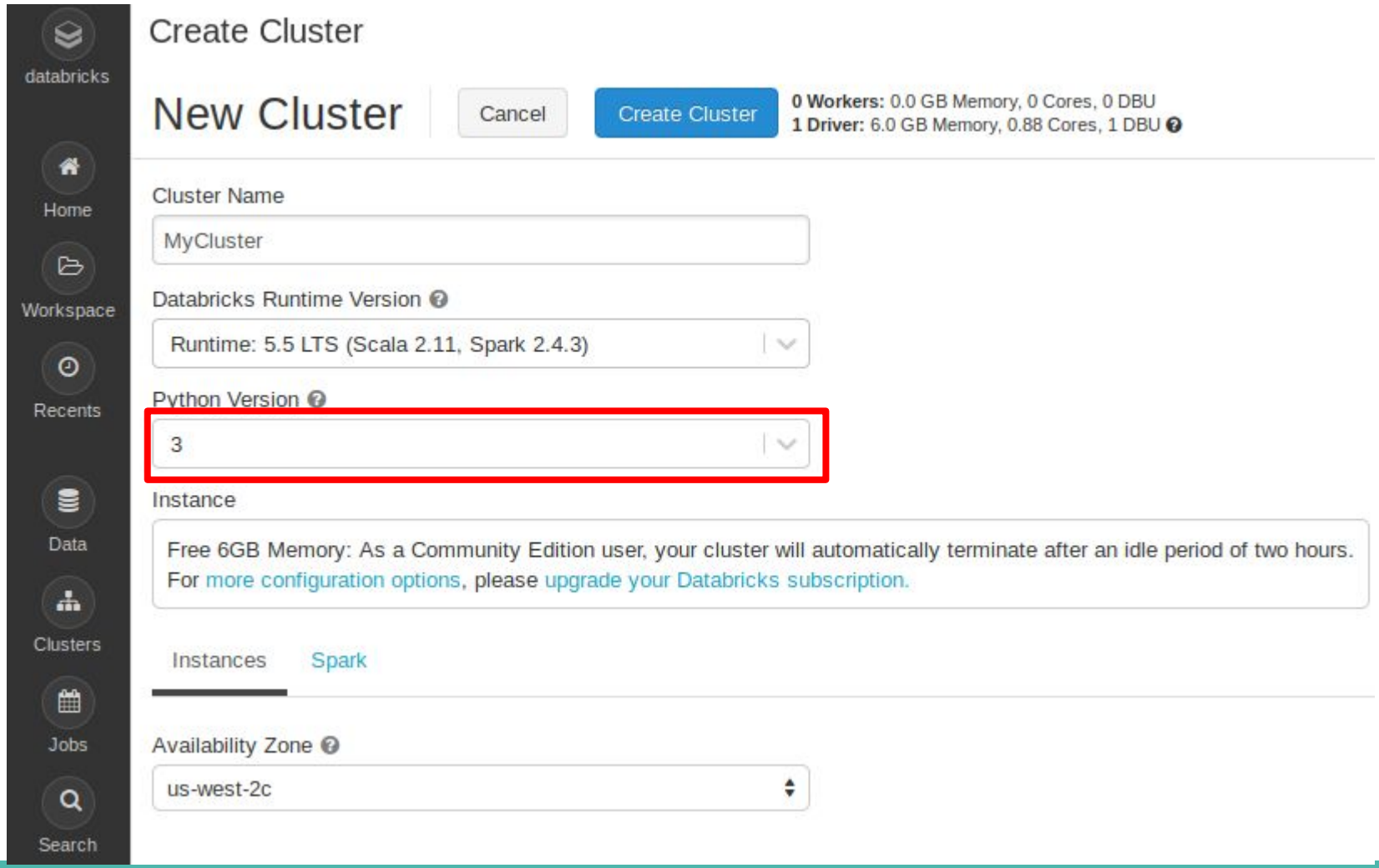
Databricks: An Online Platform for Data Engineers

3. We set up the new cluster with the following parameters:



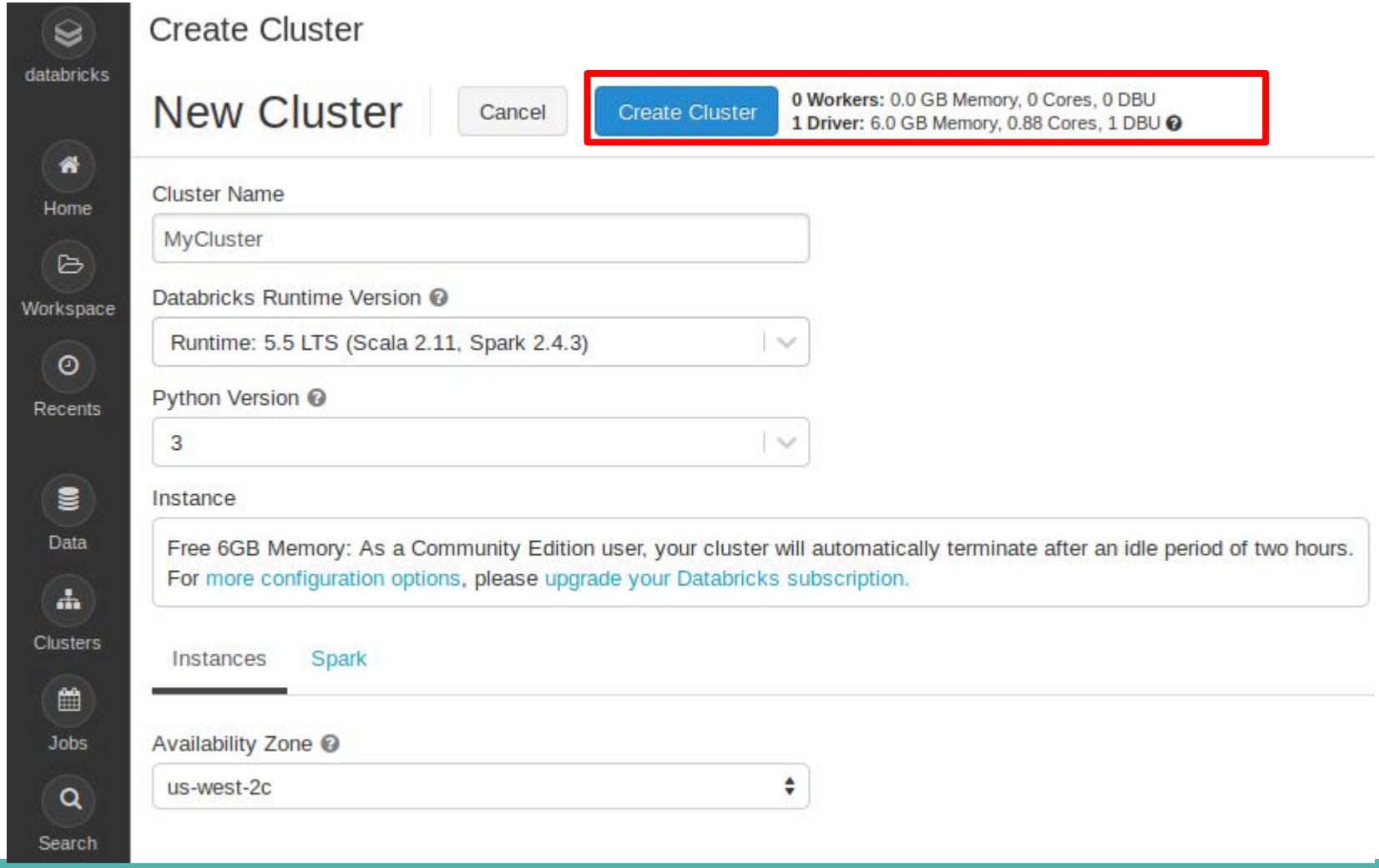
Databricks: An Online Platform for Data Engineers

3. We set up the new cluster with the following parameters:



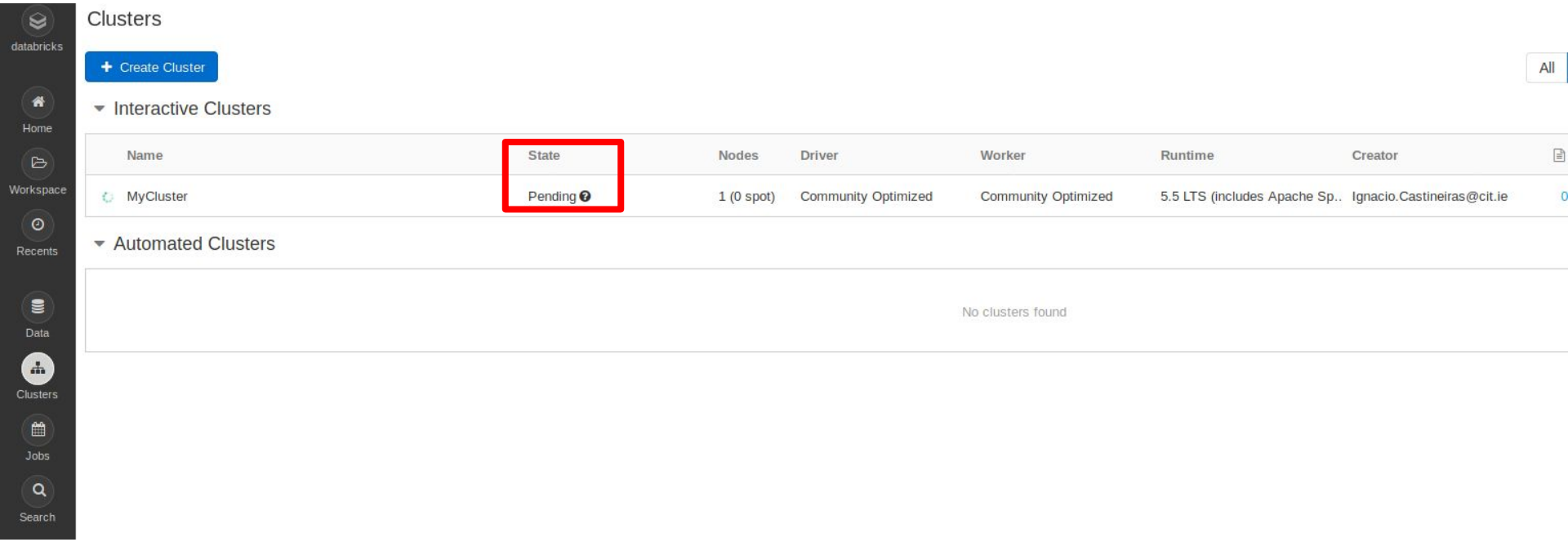
Databricks: An Online Platform for Data Engineers

3. We set up the new cluster with the following parameters:



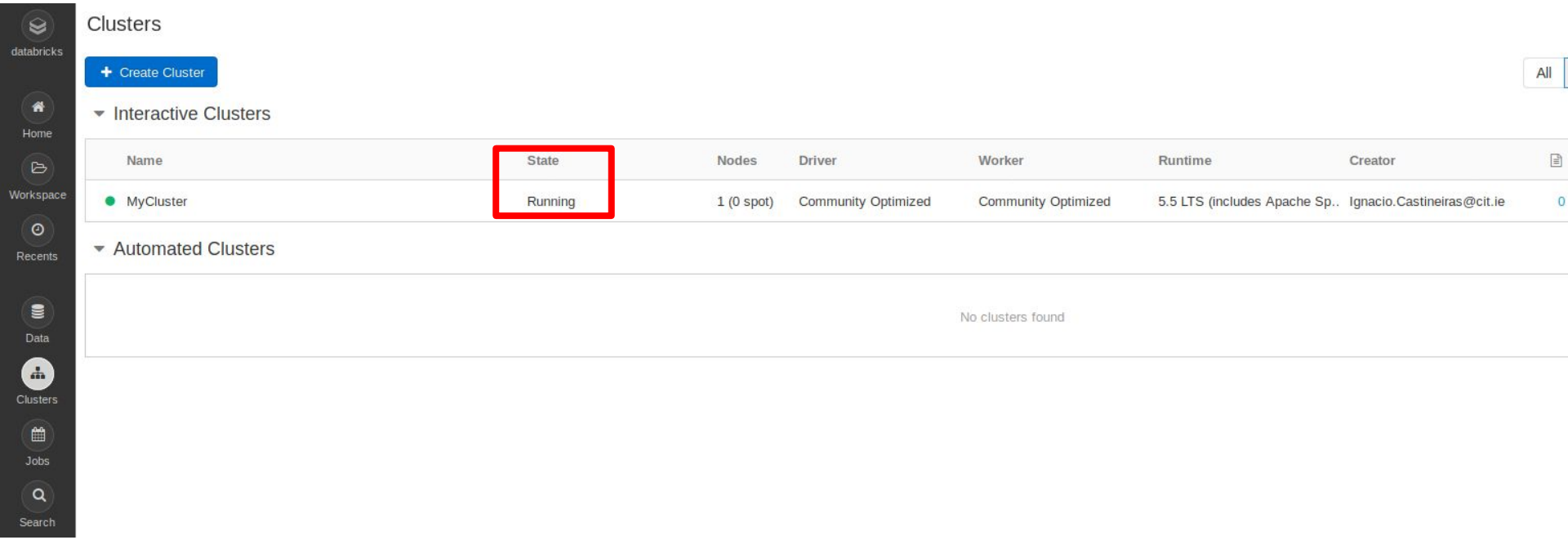
Databricks: An Online Platform for Data Engineers

3. The cluster will take a couple of minutes to be set up:



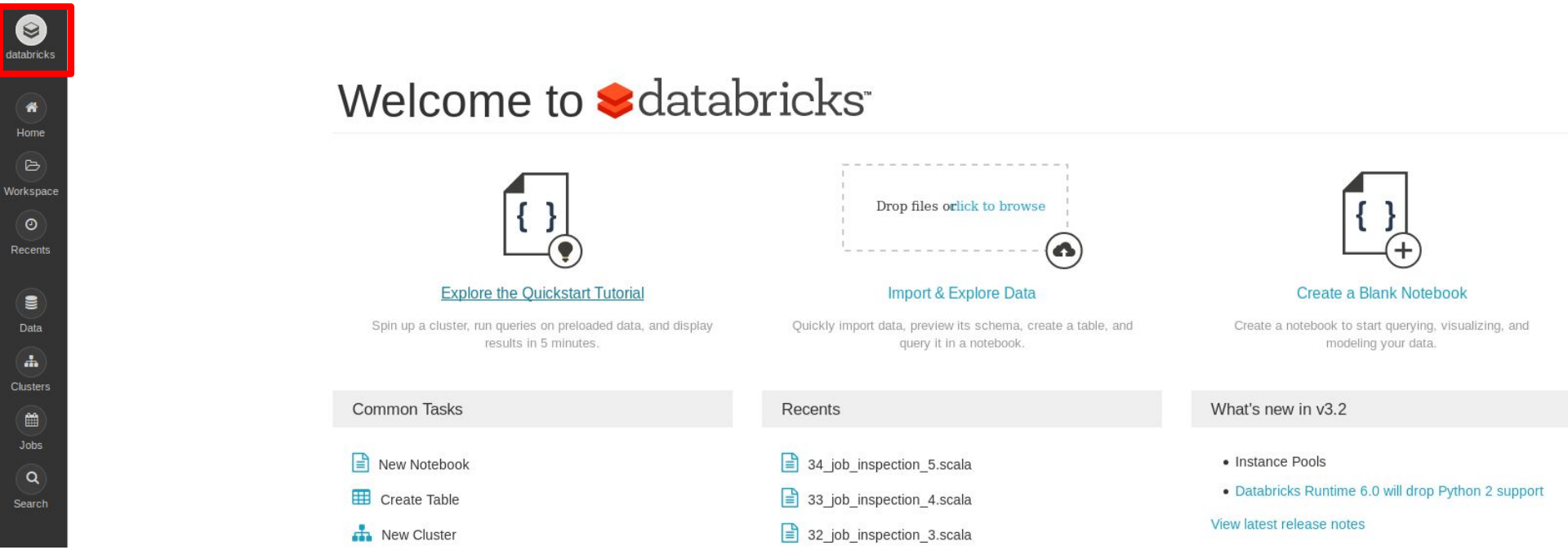
Databricks: An Online Platform for Data Engineers

3. The cluster will take a couple of minutes to be set up:



Databricks: An Online Platform for Data Engineers

3. We can come back to the main page of the web interface by clicking in **databricks**.



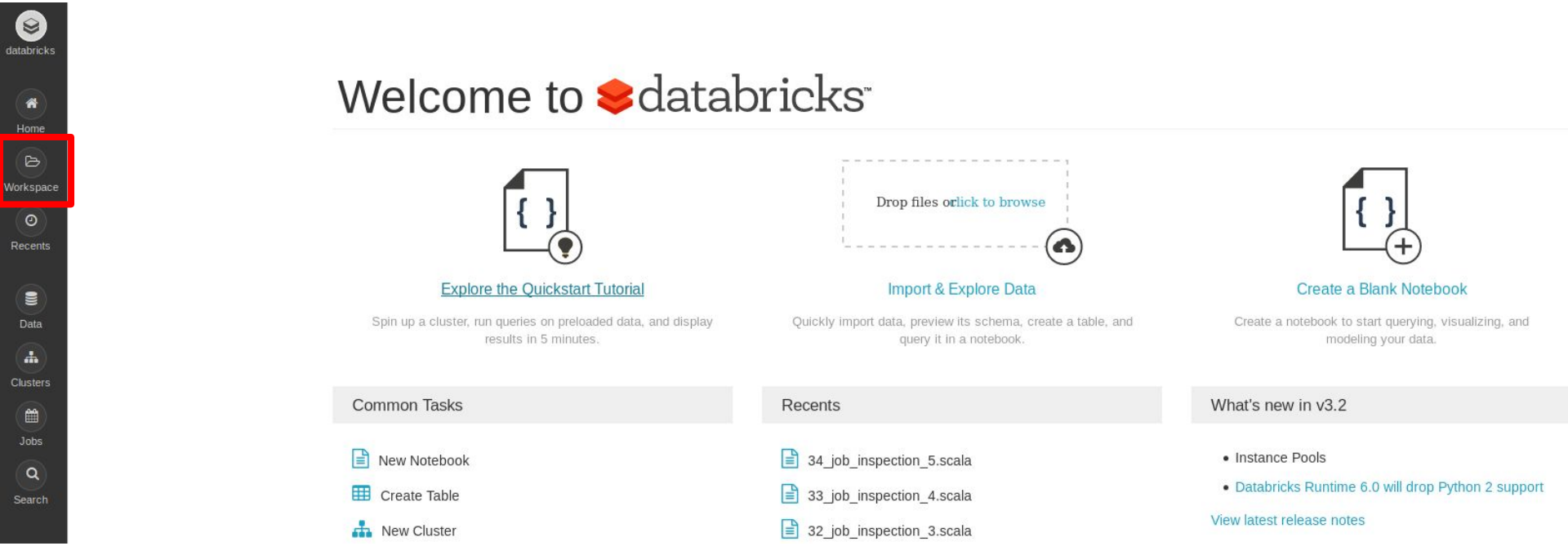
Databricks: An Online Platform for Data Engineers

How to...

Create a new program / Spark Application.

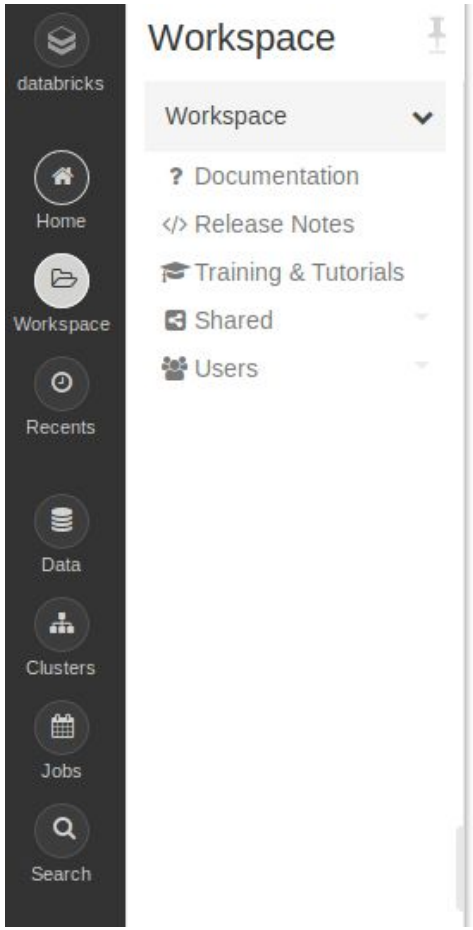
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



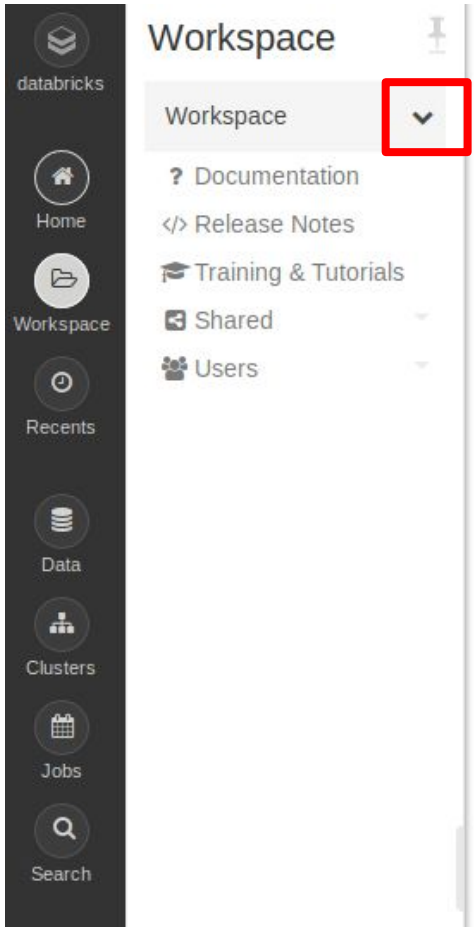
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



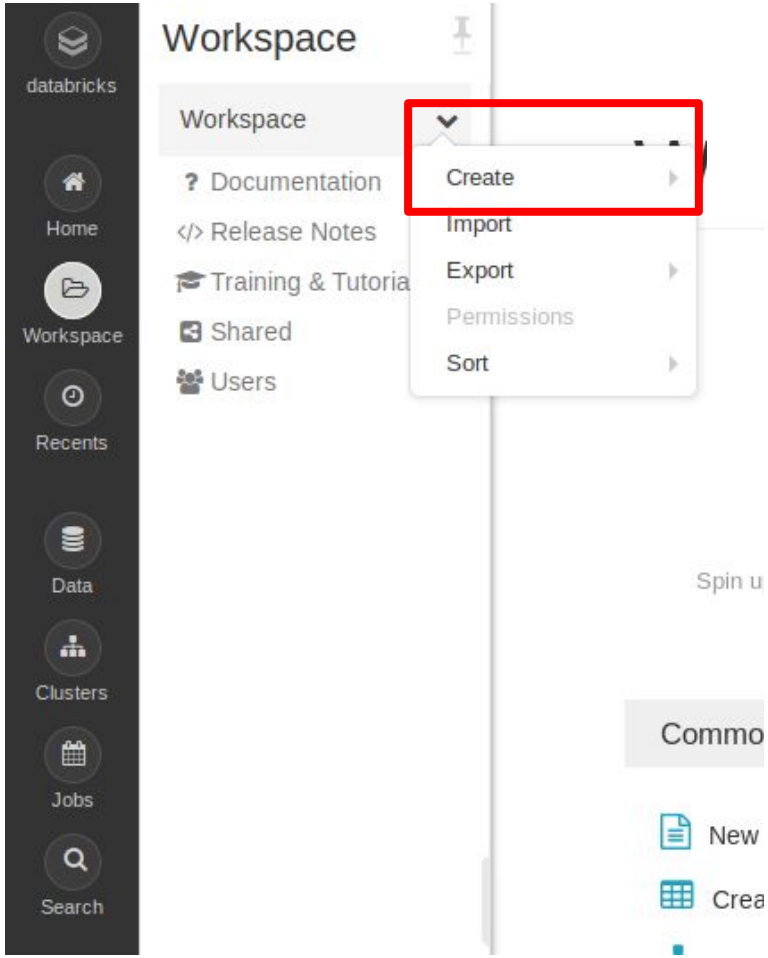
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



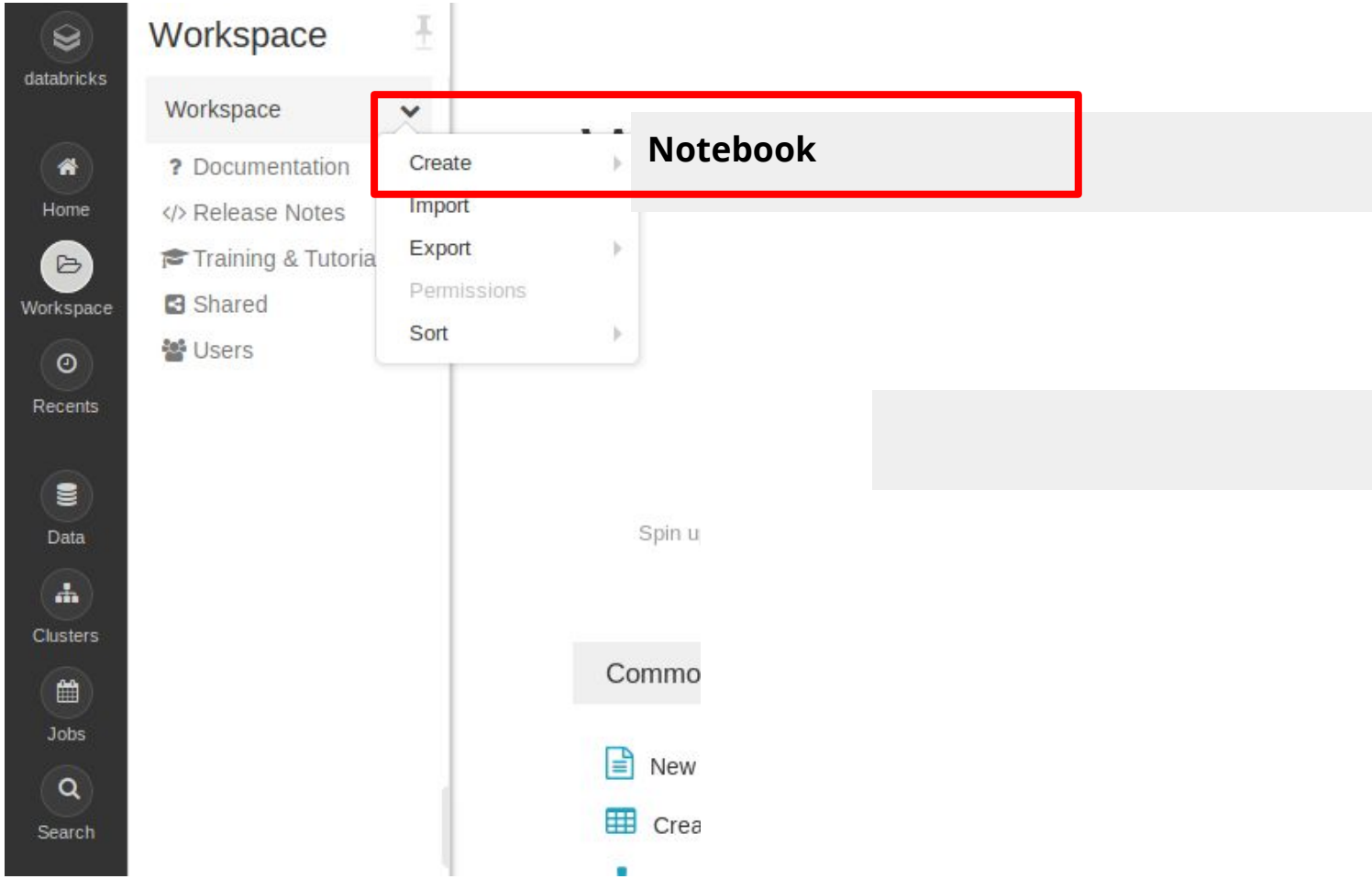
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



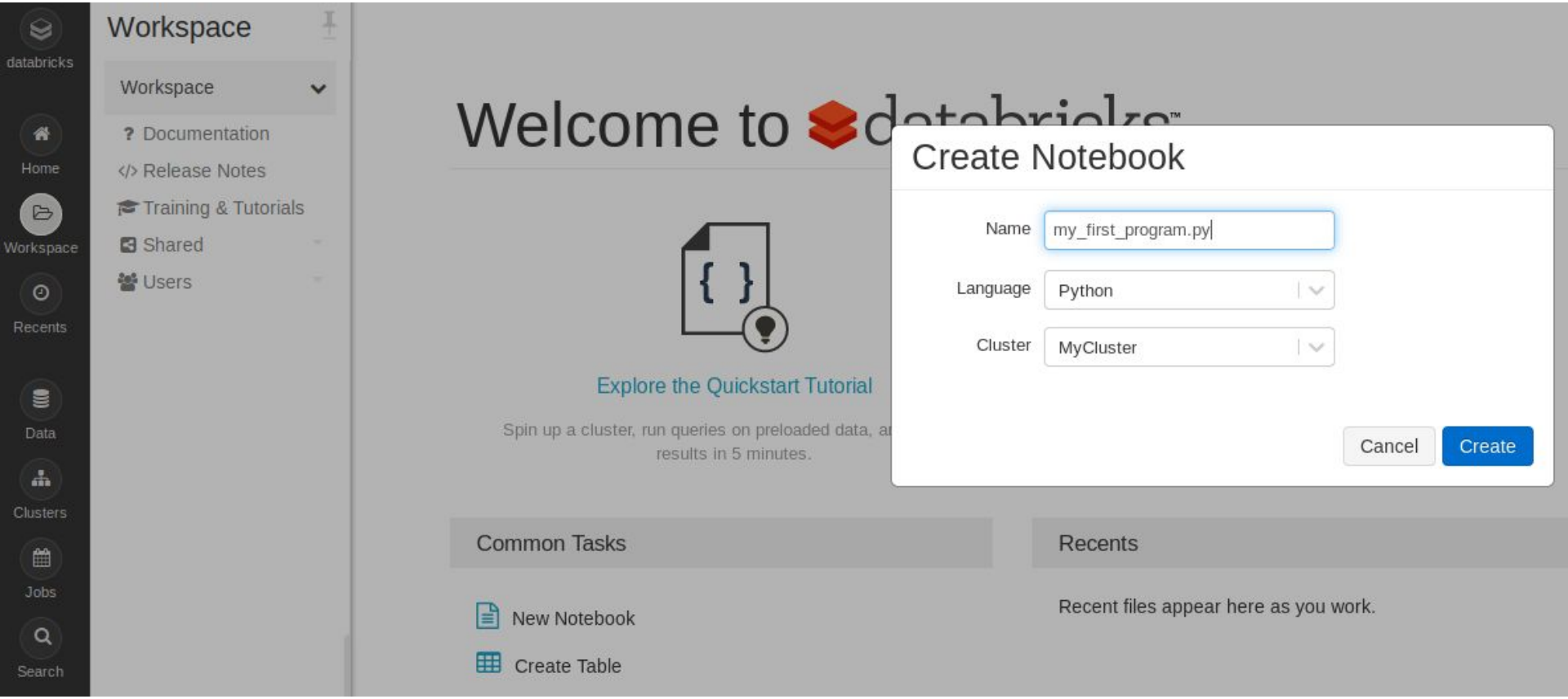
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



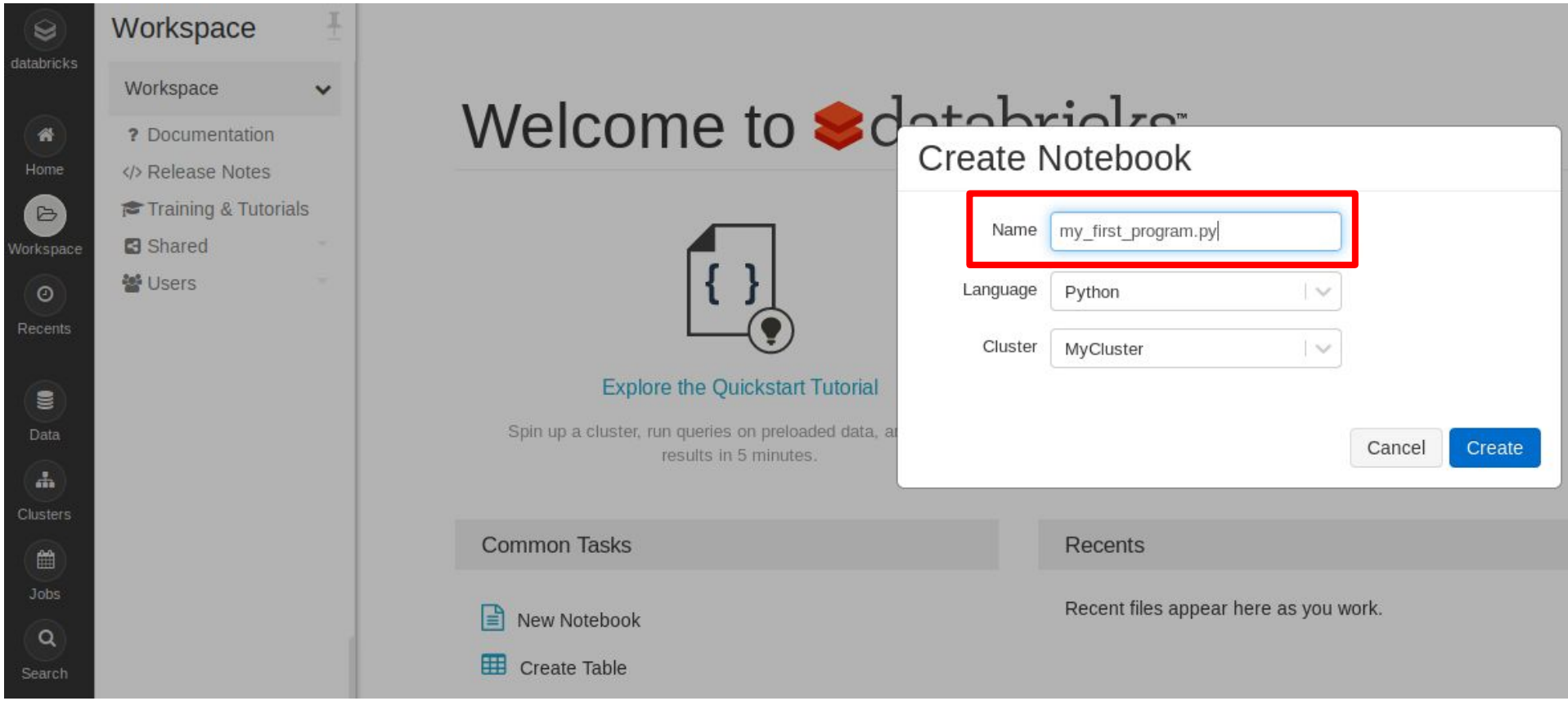
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



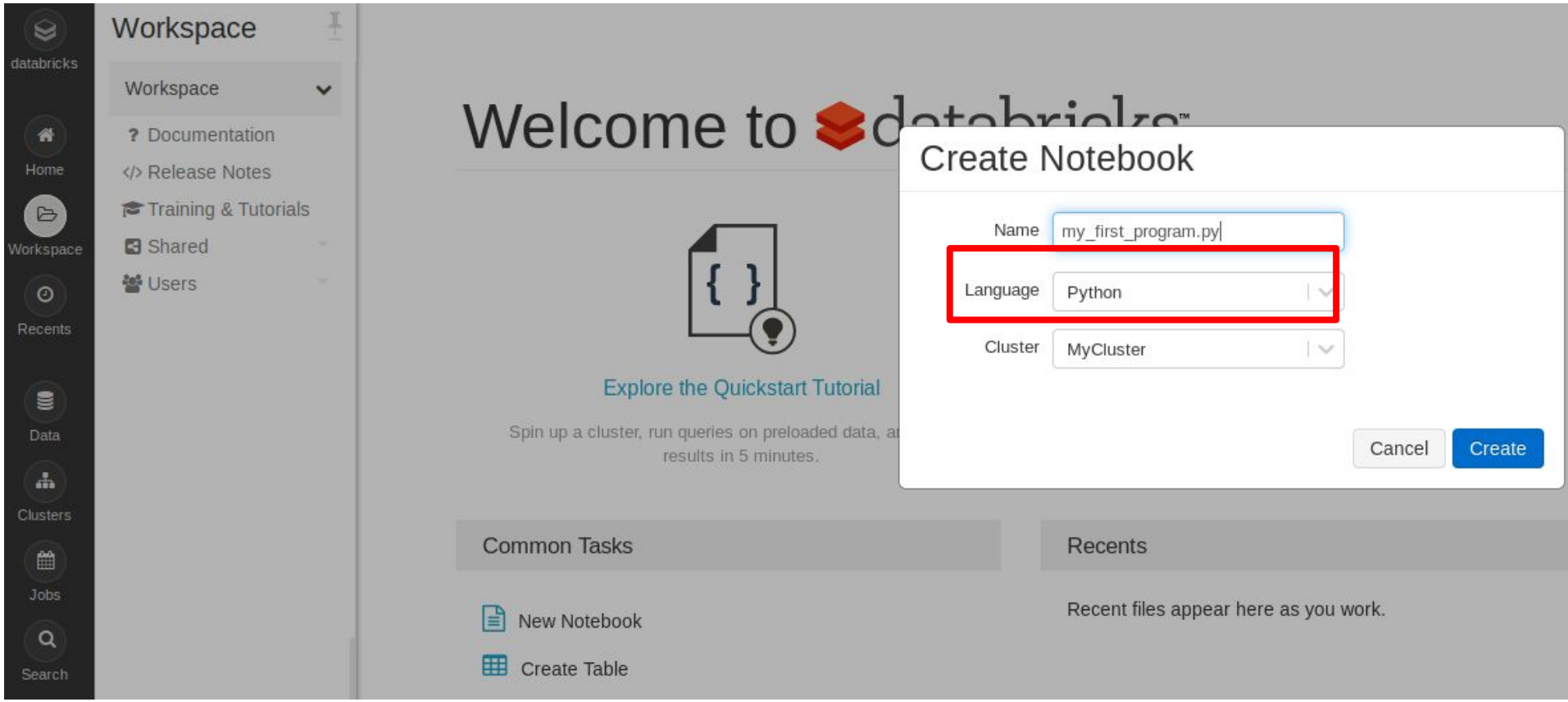
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



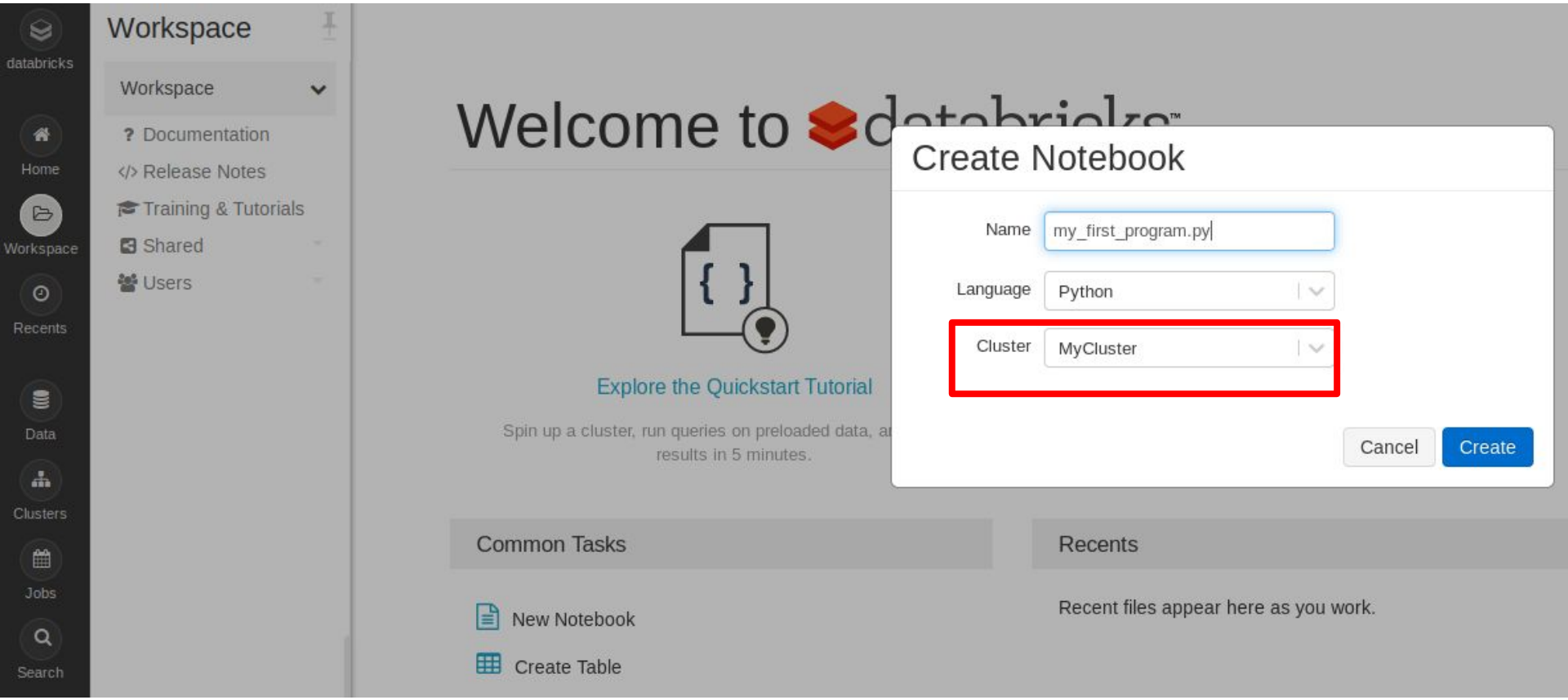
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



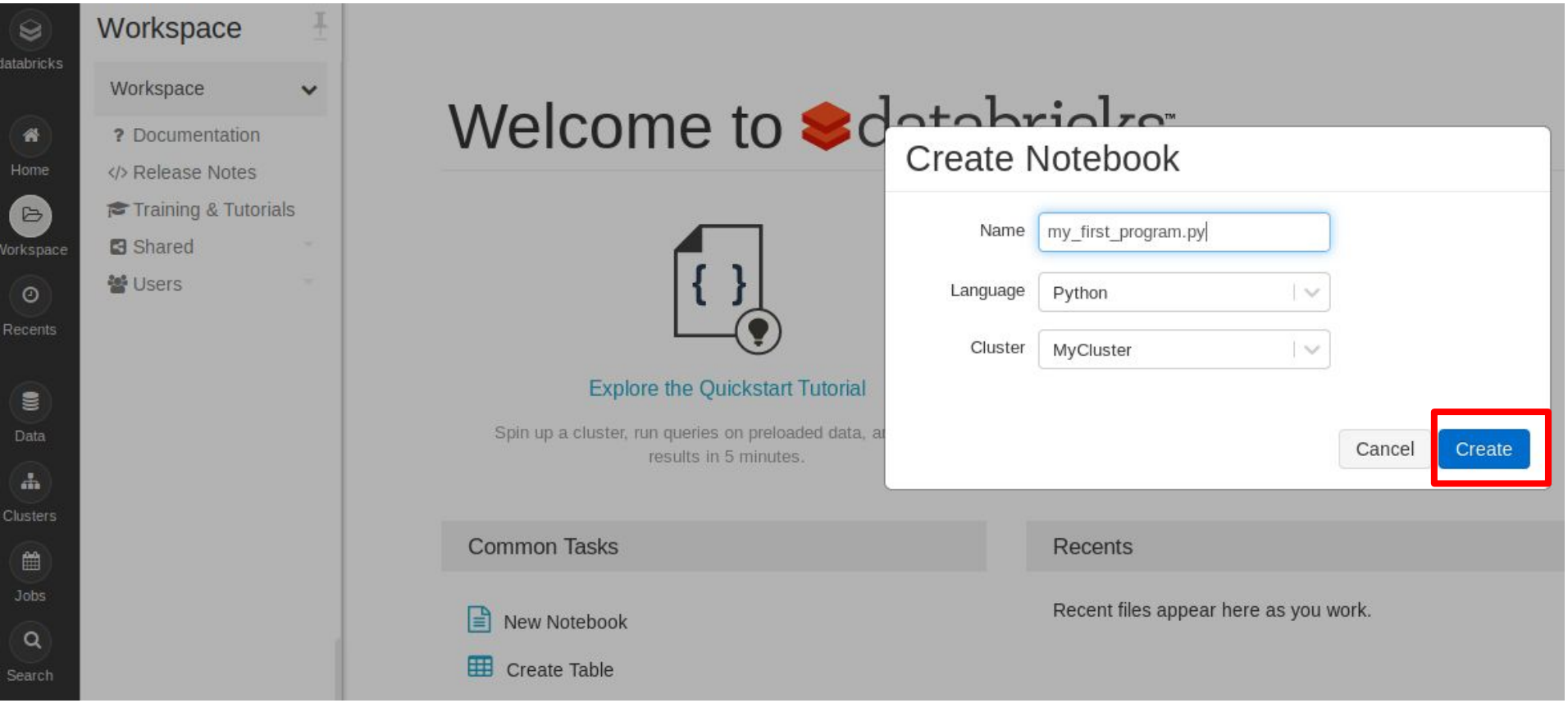
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



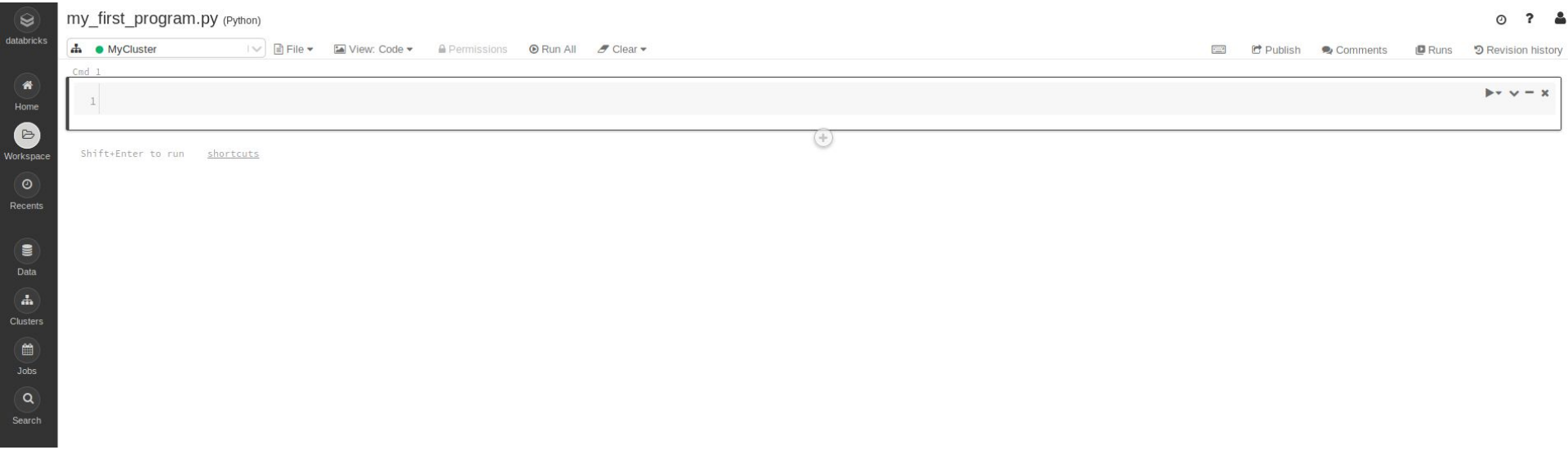
Databricks: An Online Platform for Data Engineers

4. Our next step is to create a new Spark application as a new notebook in our **Workspace**.



Databricks: An Online Platform for Data Engineers

4. We are redirected to the main web page to edit our notebook. At the beginning it is empty.

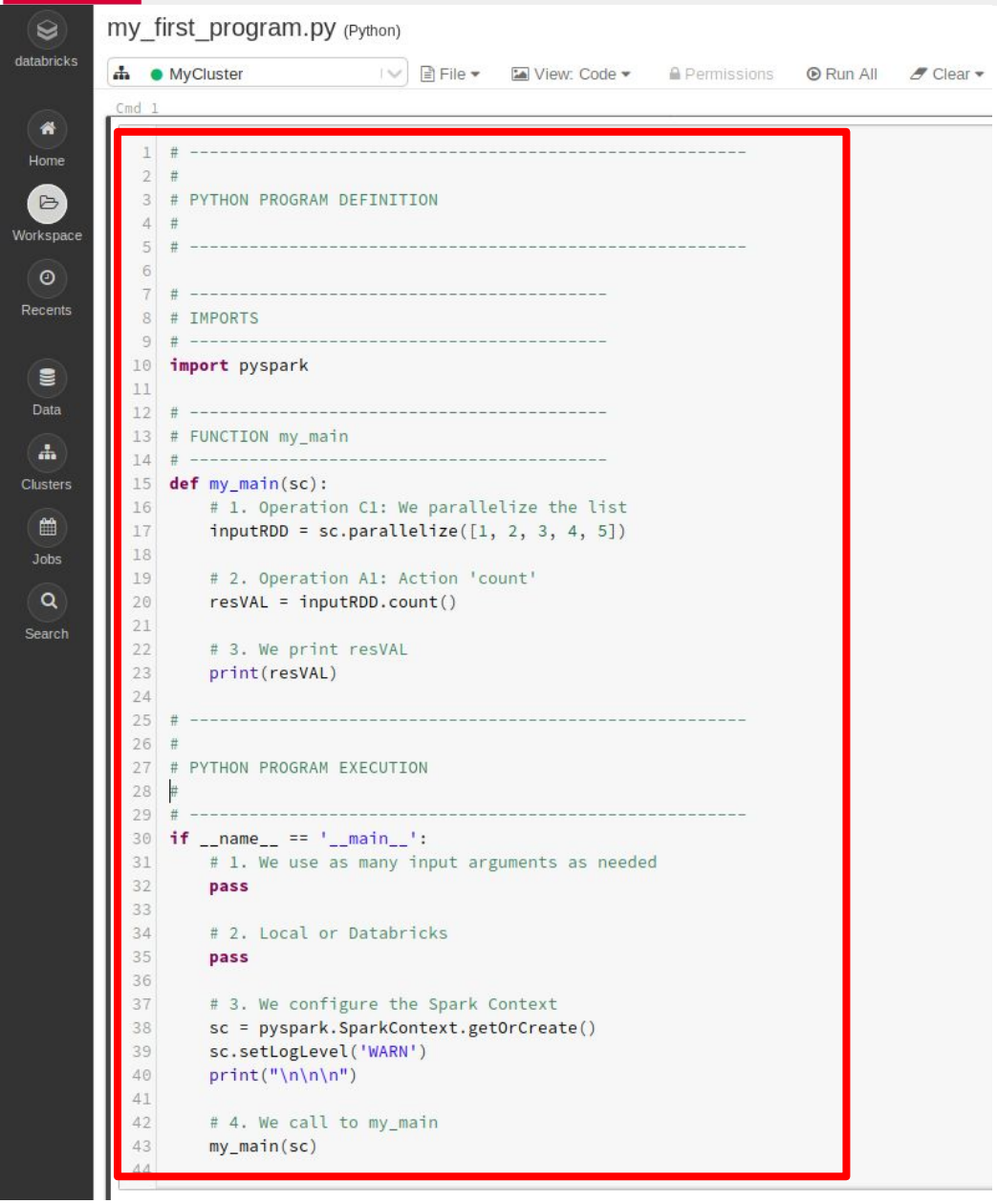


Databricks: An Online Platform for Data Engineers

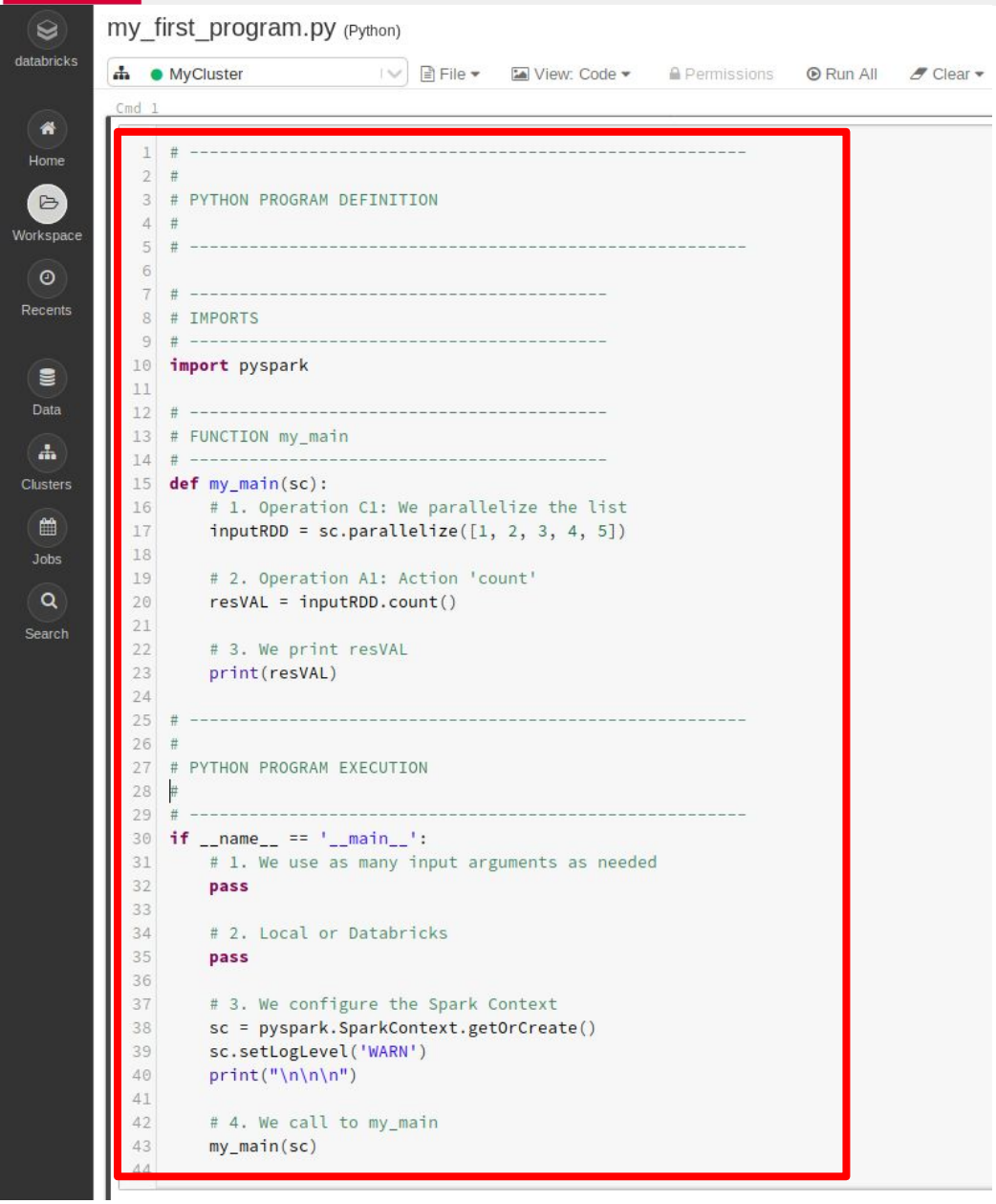
4. We are redirected to the main web page to edit our notebook.
   As we can see our program called **my_first_program.py** is automatically attached to our cluster **MyCluster**. Otherwise we can attach it by ourselves.

Databricks: An Online Platform for Data Engineers

4. We are redirected to the main web page to edit our notebook.
   As we can see our program called **my_first_program.py** is automatically attached to our cluster **MyCluster**. Otherwise we can attach it by ourselves.



4. We fill in our program in the Jupiter Notebook being created.



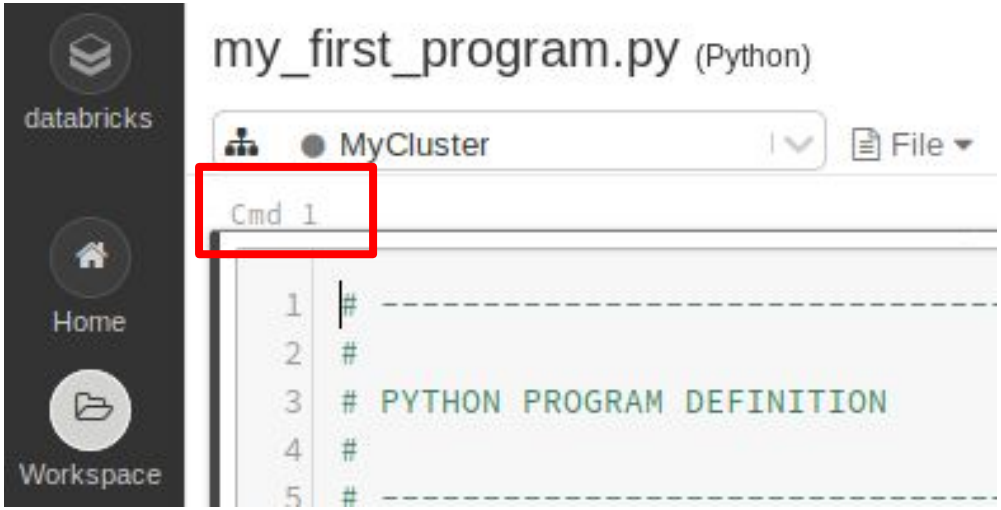
4. We fill in our program in the Jupiter Notebook being created.

Note:

As not being a big fan of Jupiter Notebooks, let's please treat them as if they were regular Python programs.

That is, let's fill the entire Python program into the **Cmd 1** cell of the notebook.

Databricks: An Online Platform for Data Engineers



4. We fill in our program in the Jupiter Notebook being created.

Note:

As not being a big fan of Jupiter Notebooks, let's please treat them as if they were regular Python programs.

That is, let's fill the entire Python program into the **Cmd 1** cell of the notebook.

Databricks: An Online Platform for Data Engineers

How to...

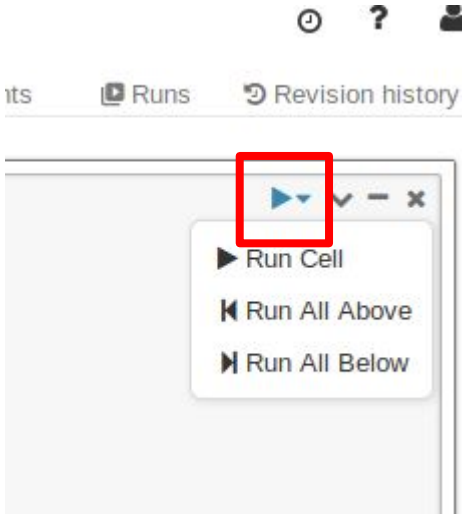
Run the new program / Spark Application.

Databricks: An Online Platform for Data Engineers

5. Once the program (Spark application) is ready, we can execute it in our cluster.

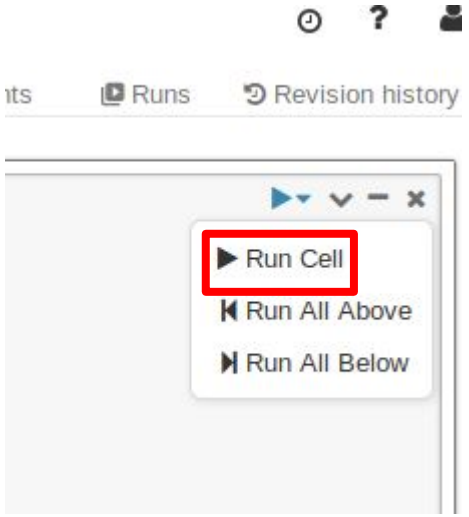
Databricks: An Online Platform for Data Engineers

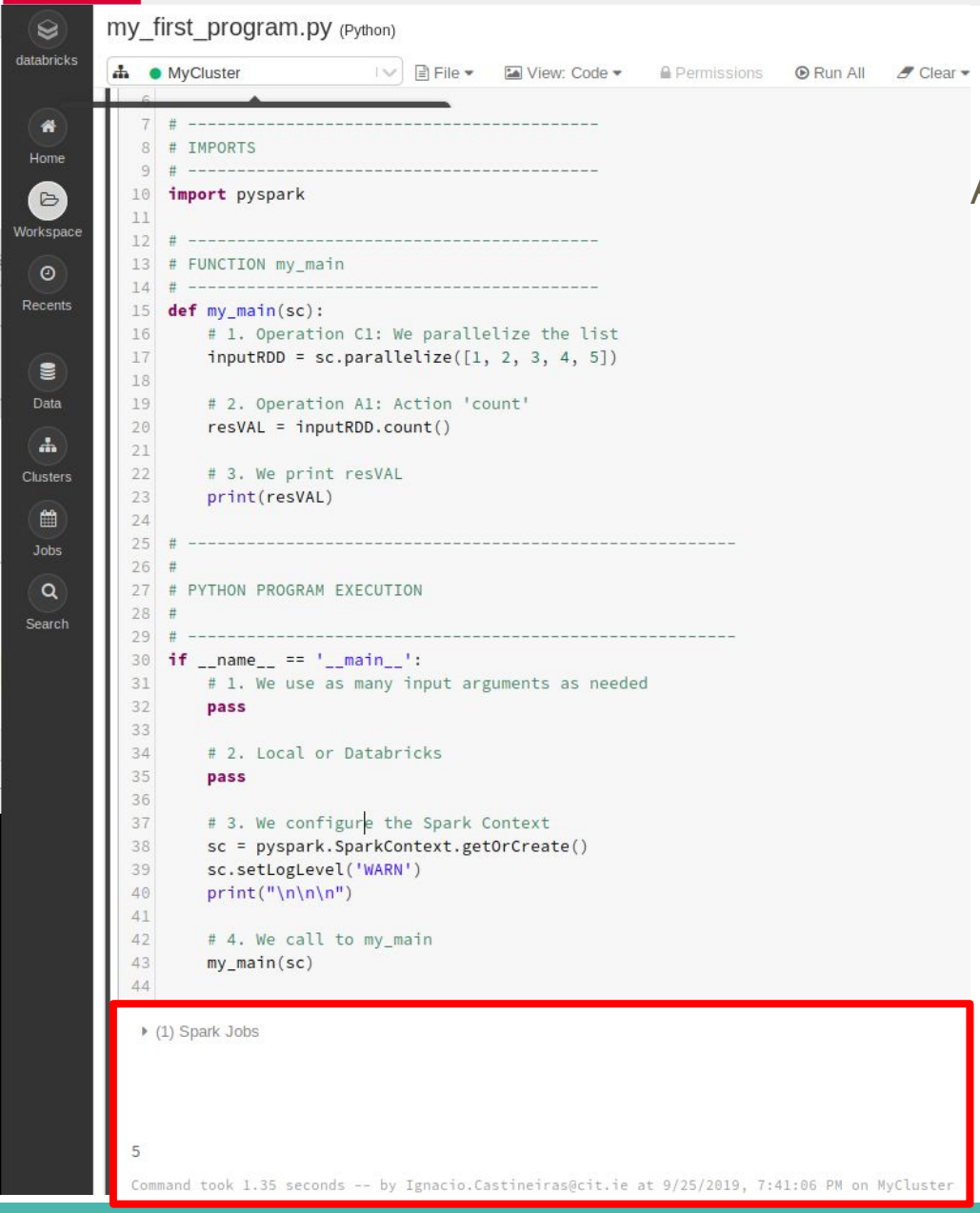
5. Once the program (Spark application) is ready, we can execute it in our cluster.



Databricks: An Online Platform for Data Engineers

5. Once the program (Spark application) is ready, we can execute it in our cluster.





And we can evaluate its result.

Databricks: An Online Platform for Data Engineers

5. And we can evaluate its result.

```
34      # 2. Local or Databricks
35      pass
36
37      # 3. We configure the Spark Context
38      sc = pyspark.SparkContext.getOrCreate()
39      sc.setLogLevel('WARN')
40      print("\n\n\n")
41
42      # 4. We call to my_main
43      my_main(sc)
44
```

▸ (1) Spark Jobs

```
5
```

Command took 1.35 seconds -- by Ignacio.Castineiras@cit.ie at 9/25/2019, 7:41:06 PM on MyCluster

Databricks: An Online Platform for Data Engineers

5. And we can evaluate its result.

```
34      # 2. Local or Databricks
35      pass
36
37      # 3. We configure the Spark Context
38      sc = pyspark.SparkContext.getOrCreate()
39      sc.setLogLevel('WARN')
40      print("\n\n\n")
41
42      # 4. We call to my_main
43      my_main(sc)
44
```

▸ (1) Spark Jobs

5

Command took 1.35 seconds -- by Ignacio.Castineiras@cit.ie at 9/25/2019, 7:41:06 PM on MyCluster

Databricks: An Online Platform for Data Engineers

5. And we can evaluate its result.

```
34      # 2. Local or Databricks
35      pass
36
37      # 3. We configure the Spark Context
38      sc = pyspark.SparkContext.getOrCreate()
39      sc.setLogLevel('WARN')
40      print("\n\n\n")
41
42      # 4. We call to my_main
43      my_main(sc)
44
```

▸ (1) Spark Jobs

```
5
```

Command took 1.35 seconds -- by Ignacio.Castineiras@cit.ie at 9/25/2019, 7:41:06 PM on MyCluster

Databricks: An Online Platform for Data Engineers

5. And we can evaluate its result.

```
34      # 2. Local or Databricks
35      pass
36
37      # 3. We configure the Spark Context
38      sc = pyspark.SparkContext.getOrCreate()
39      sc.setLogLevel('WARN')
40      print("\n\n\n")
41
42      # 4. We call to my_main
43      my_main(sc)
44
```

▾ (1) Spark Jobs
▸ Job 0 View (Stages: 1/1)

5

Command took 2.74 seconds -- by Ignacio.Castineiras@cit.ie at 9/26/2019, 3:17:32 PM on MyCluster

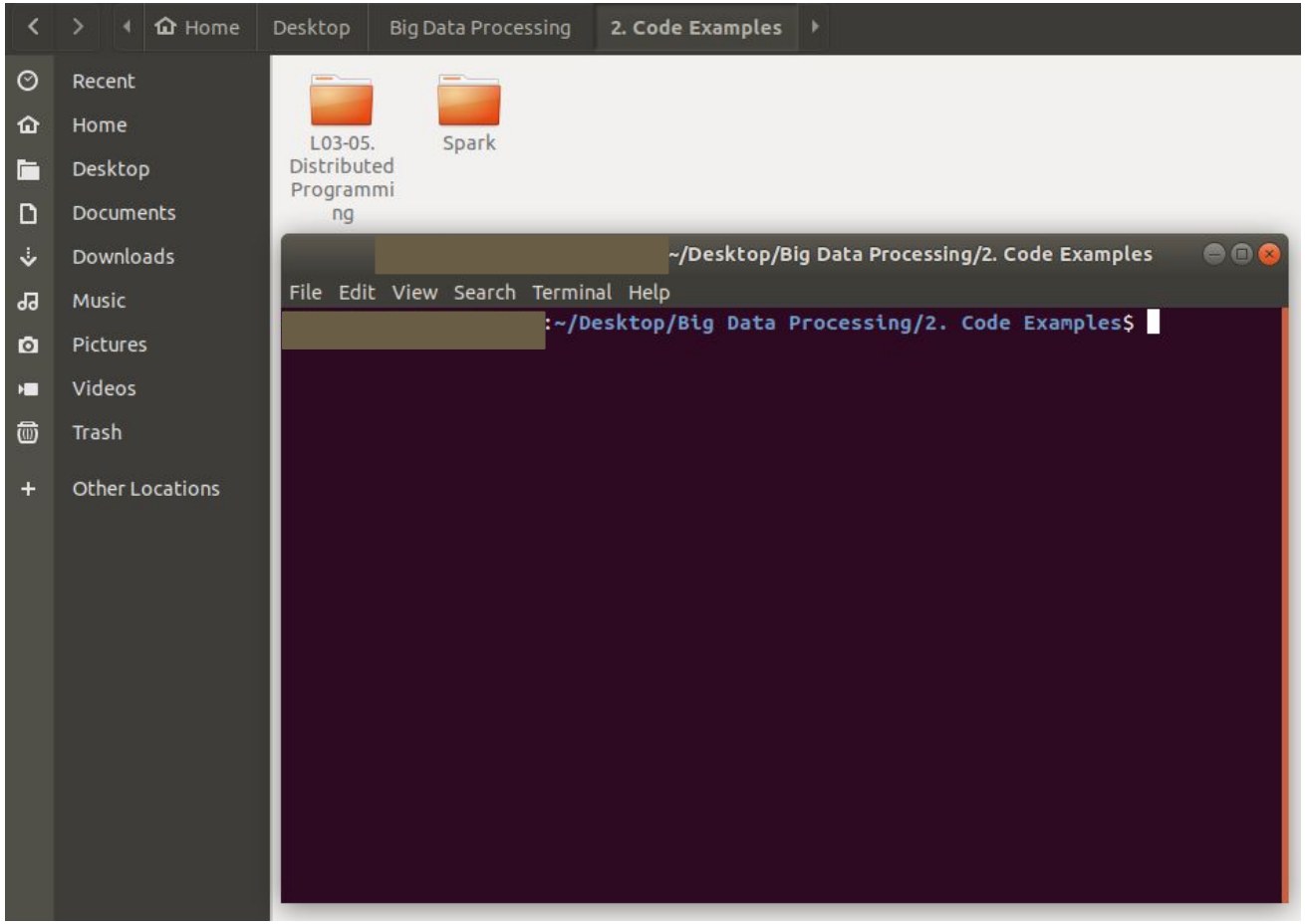
Databricks: An Online Platform for Data Engineers

How to...

Install the Databricks Command Line Application (CLI)

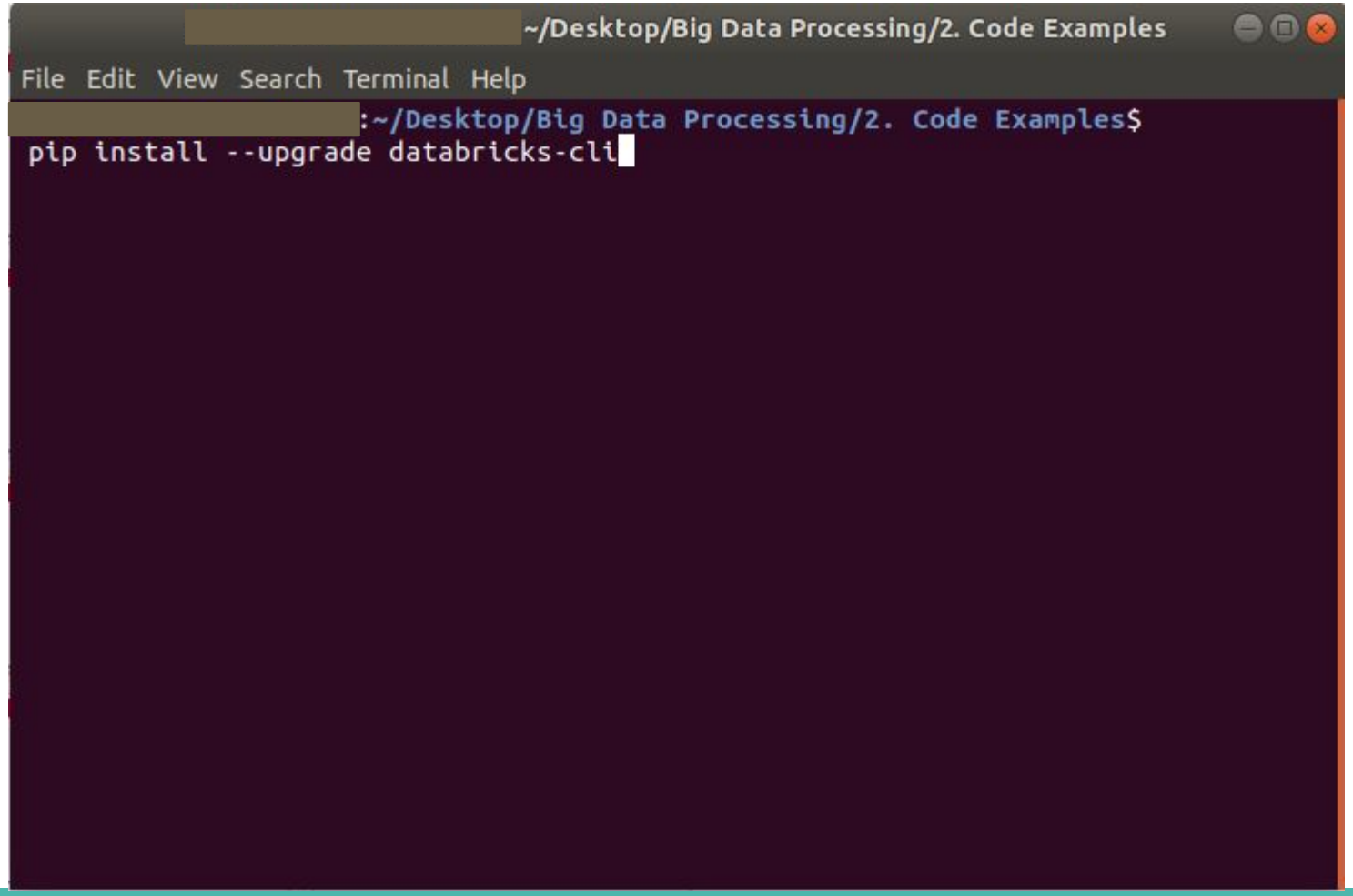
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   The CLI allows us to perform some operations on Databricks from a terminal in our local machine.



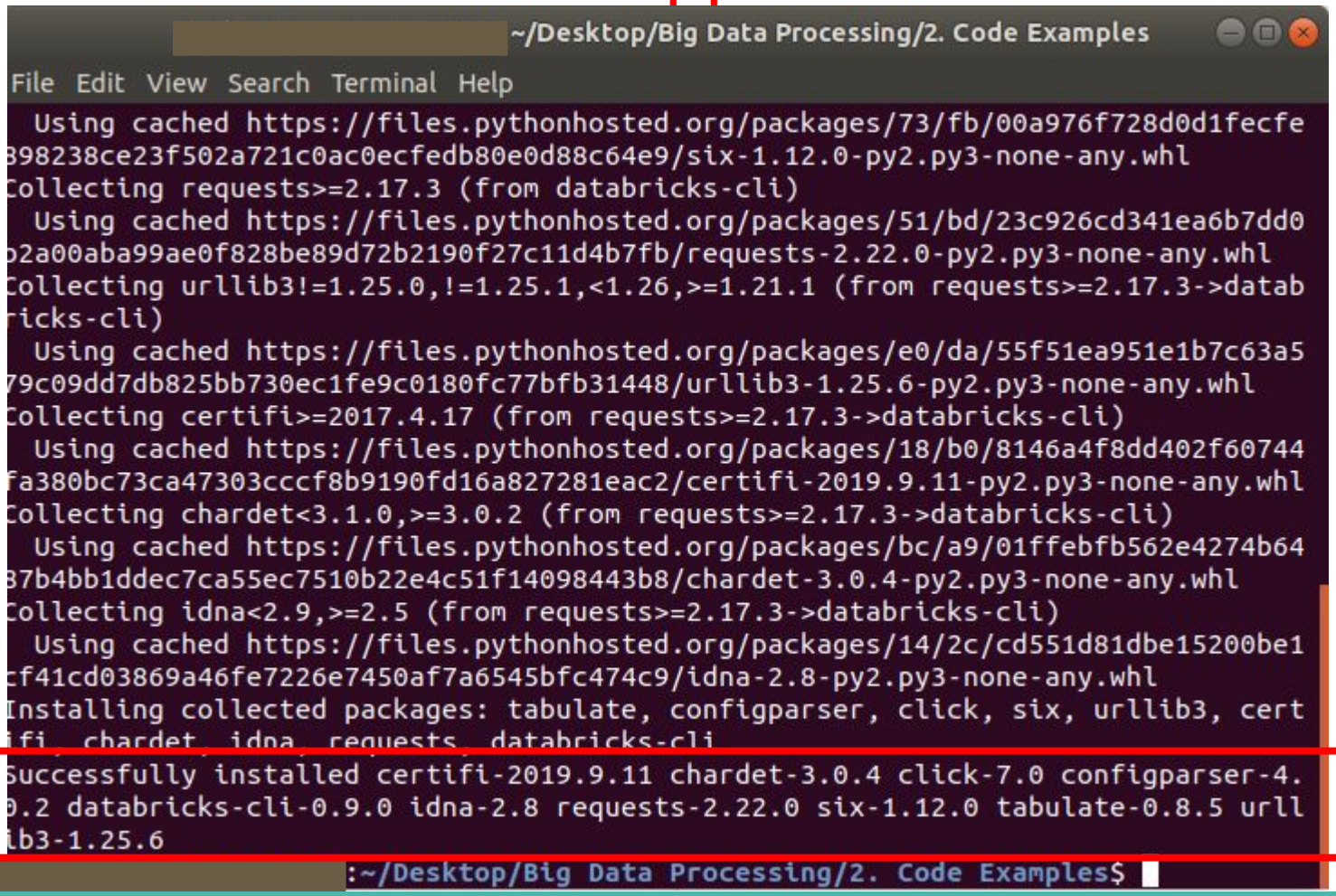
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   We can install the Databricks CLI via **pip**.



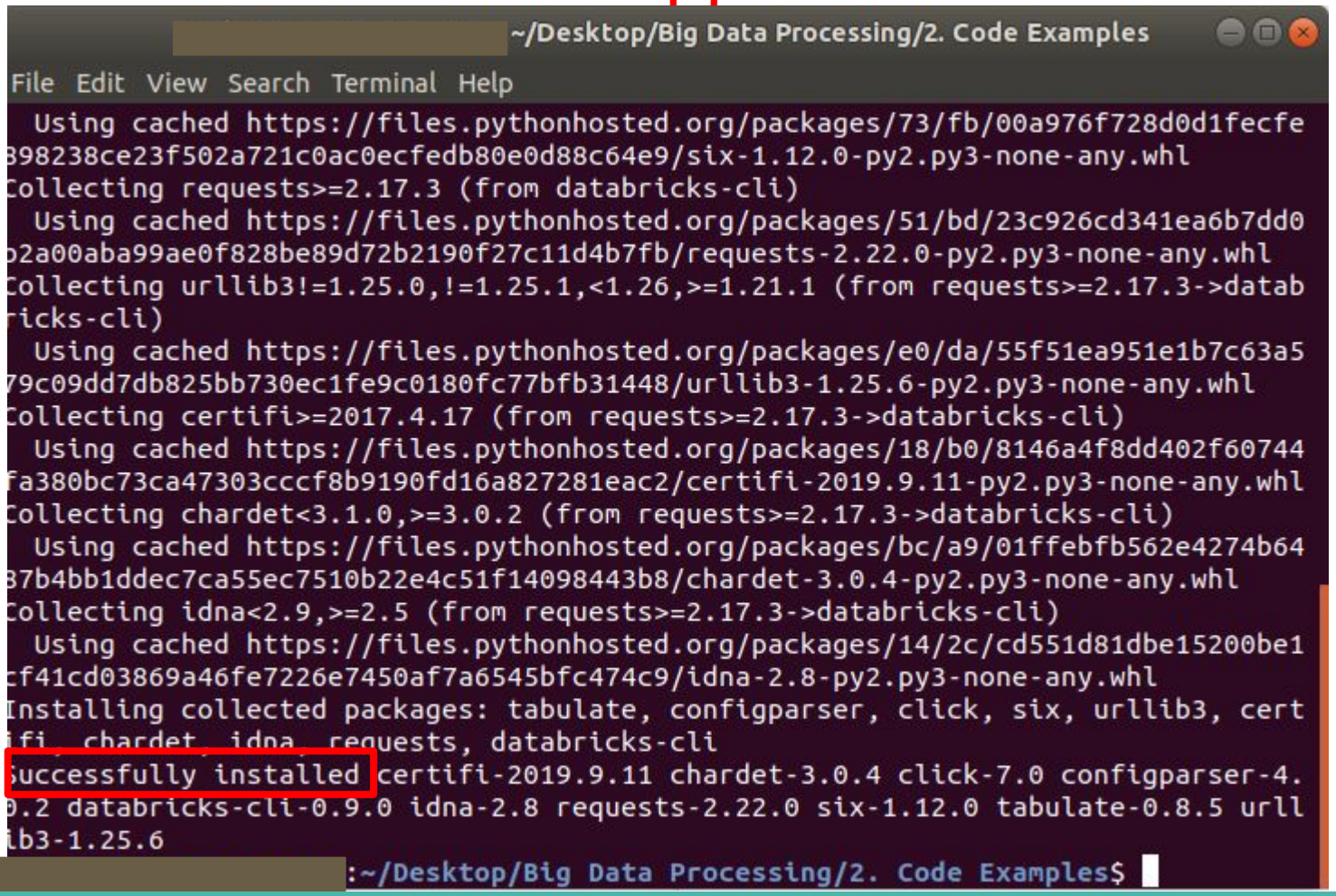
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   We can install the Databricks CLI via **pip**.



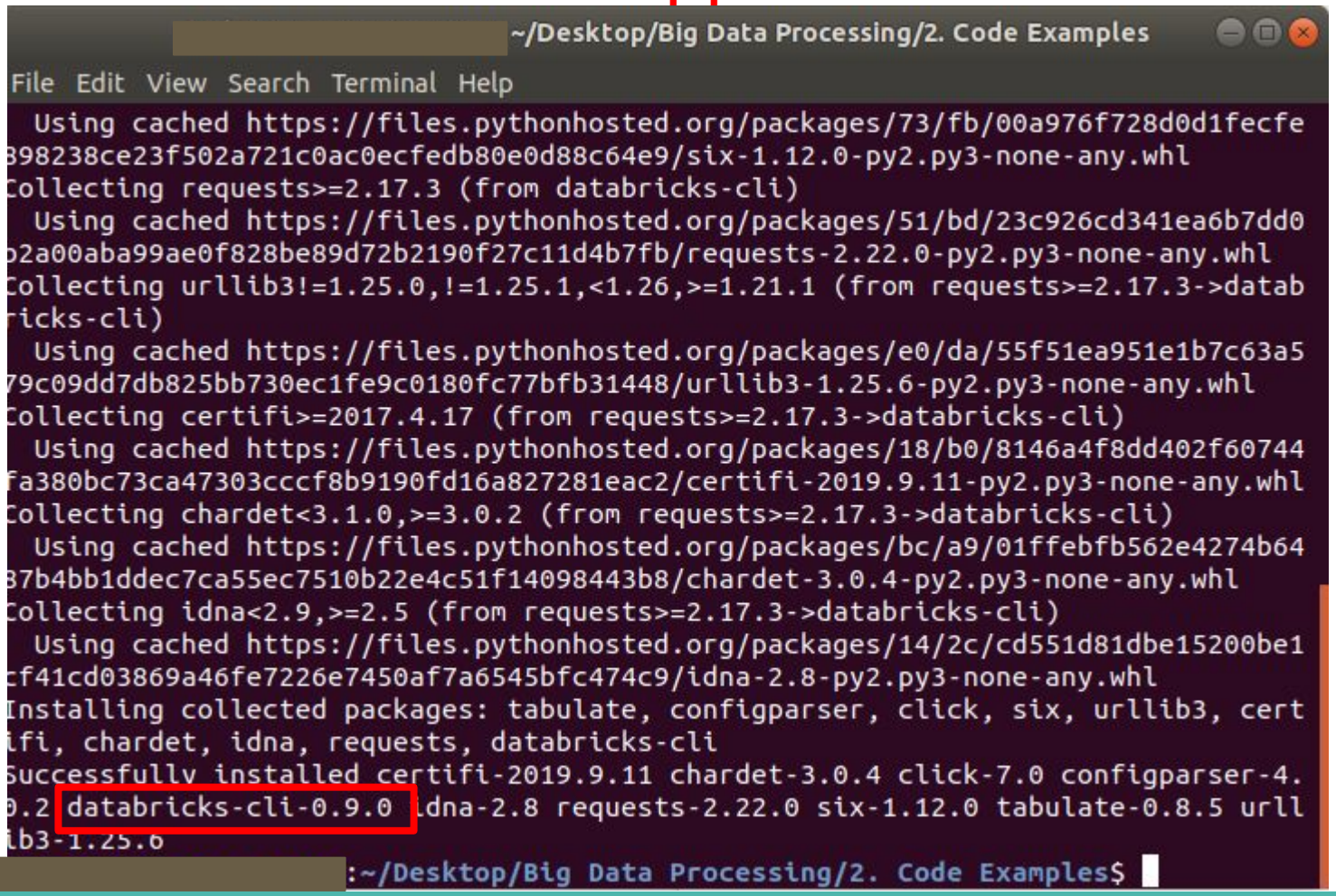
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   We can install the Databricks CLI via **pip**.



Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   We can install the Databricks CLI via **pip**.

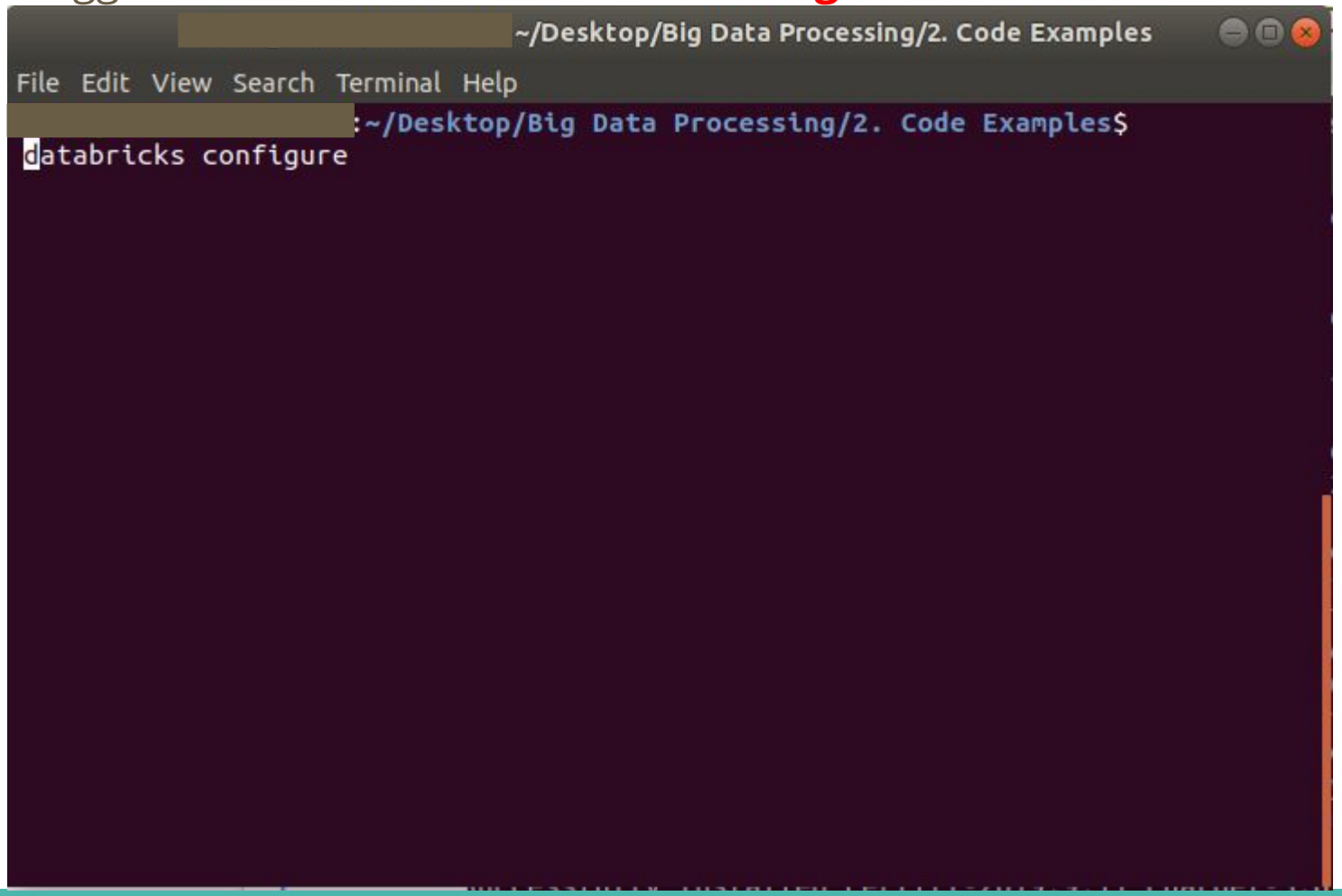


Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   Once installed, we must configure the Databricks CLI.

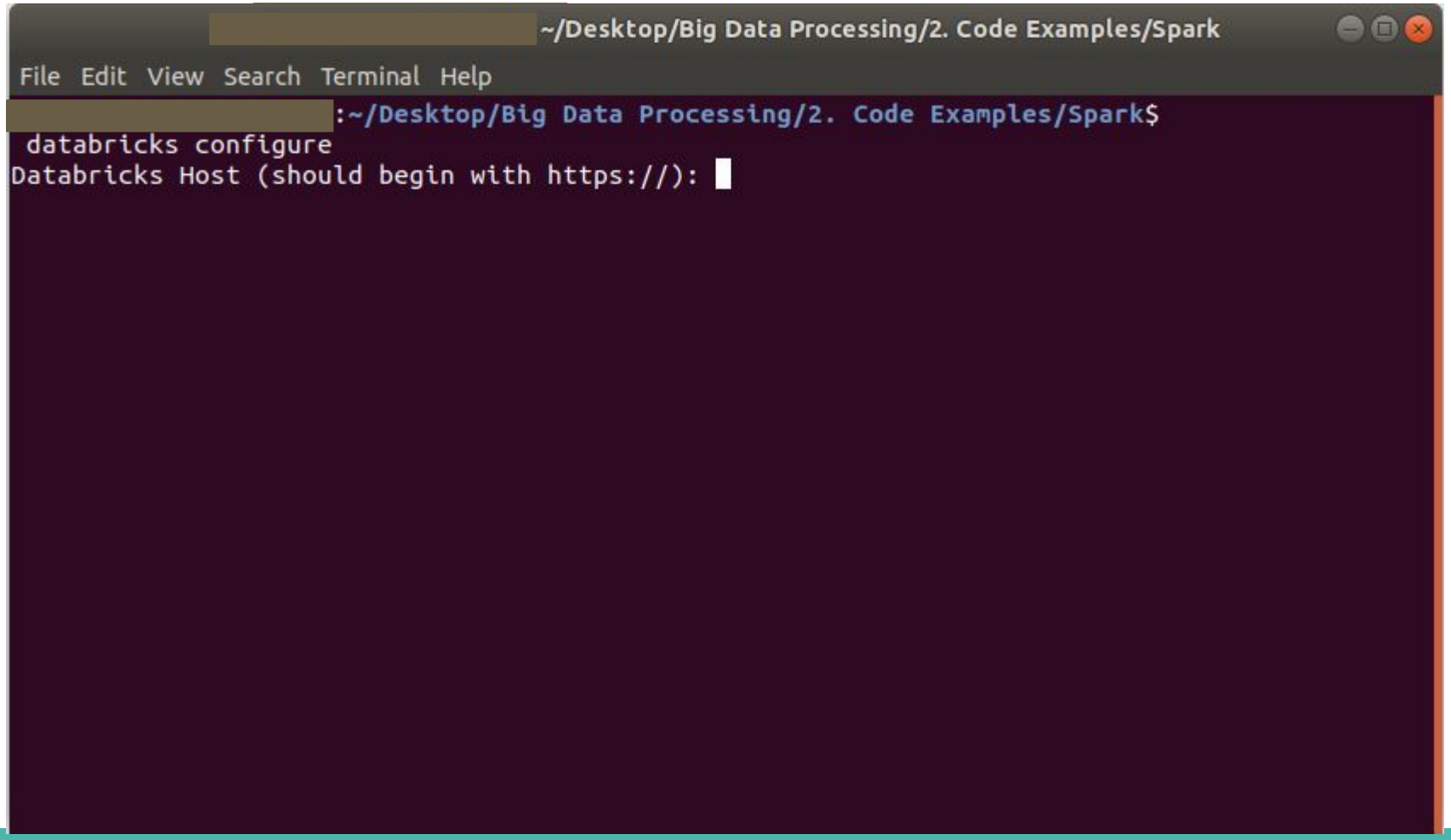
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   We trigger the command **databricks configure**



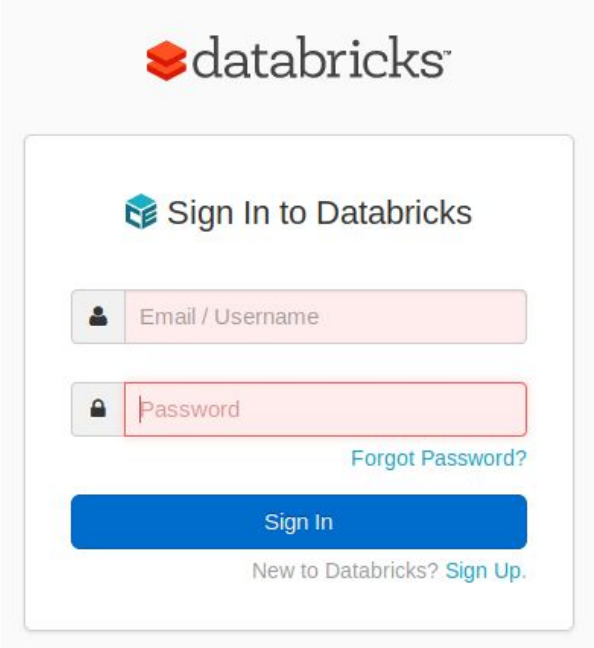
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   a. We must indicate our databricks host.



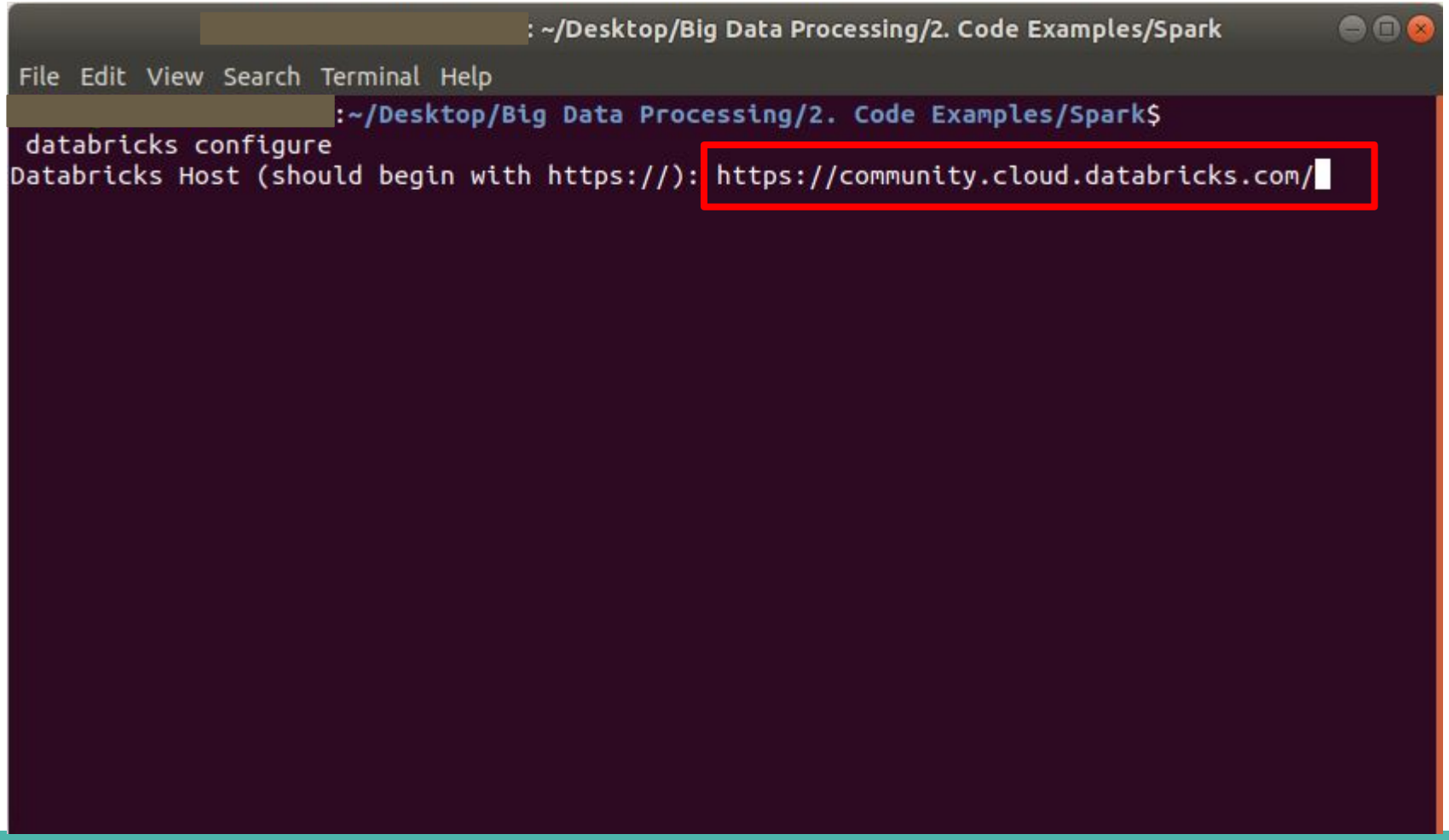
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   a. We must indicate our databricks host.
      In our case, our host is the website for Databricks Community Edition:
      https://community.cloud.databricks.com



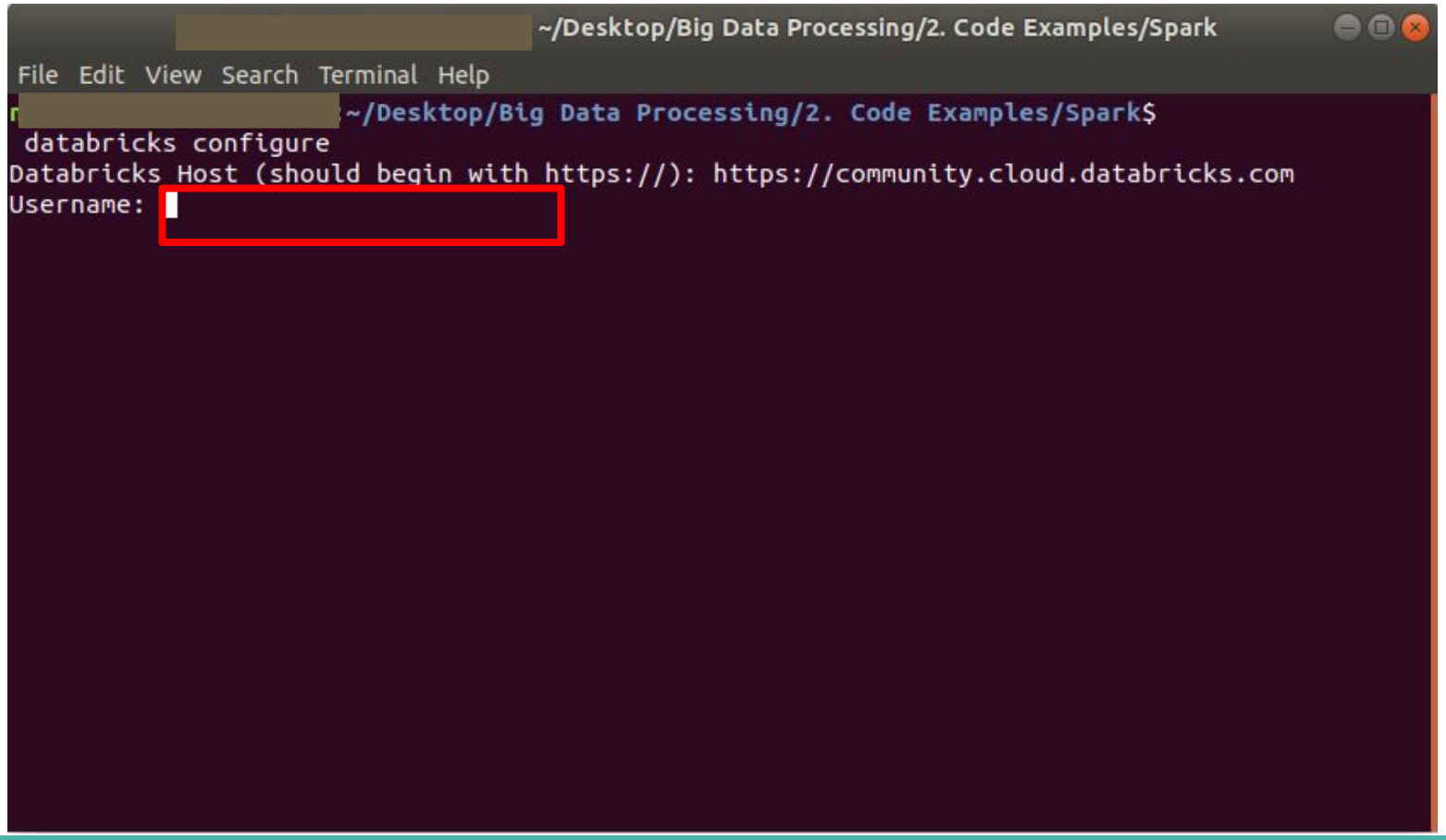
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   a. We must indicate our databricks host.



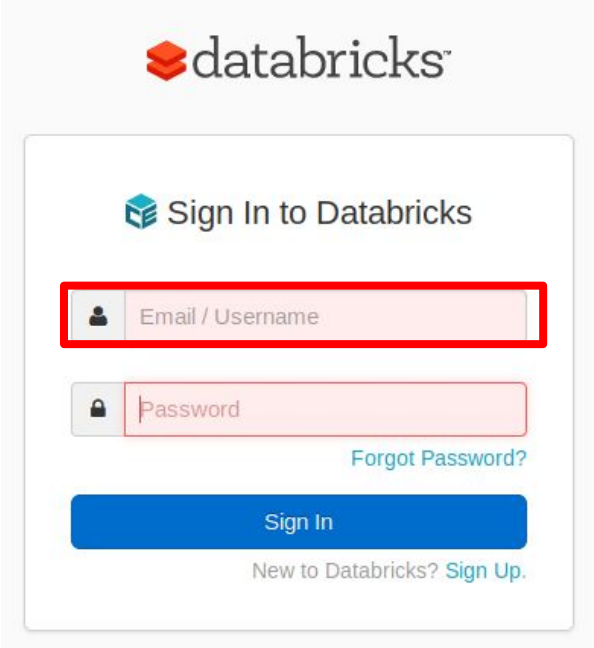
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   b. We must indicate our username (same as when we sign in at https://community.cloud.databricks.com).



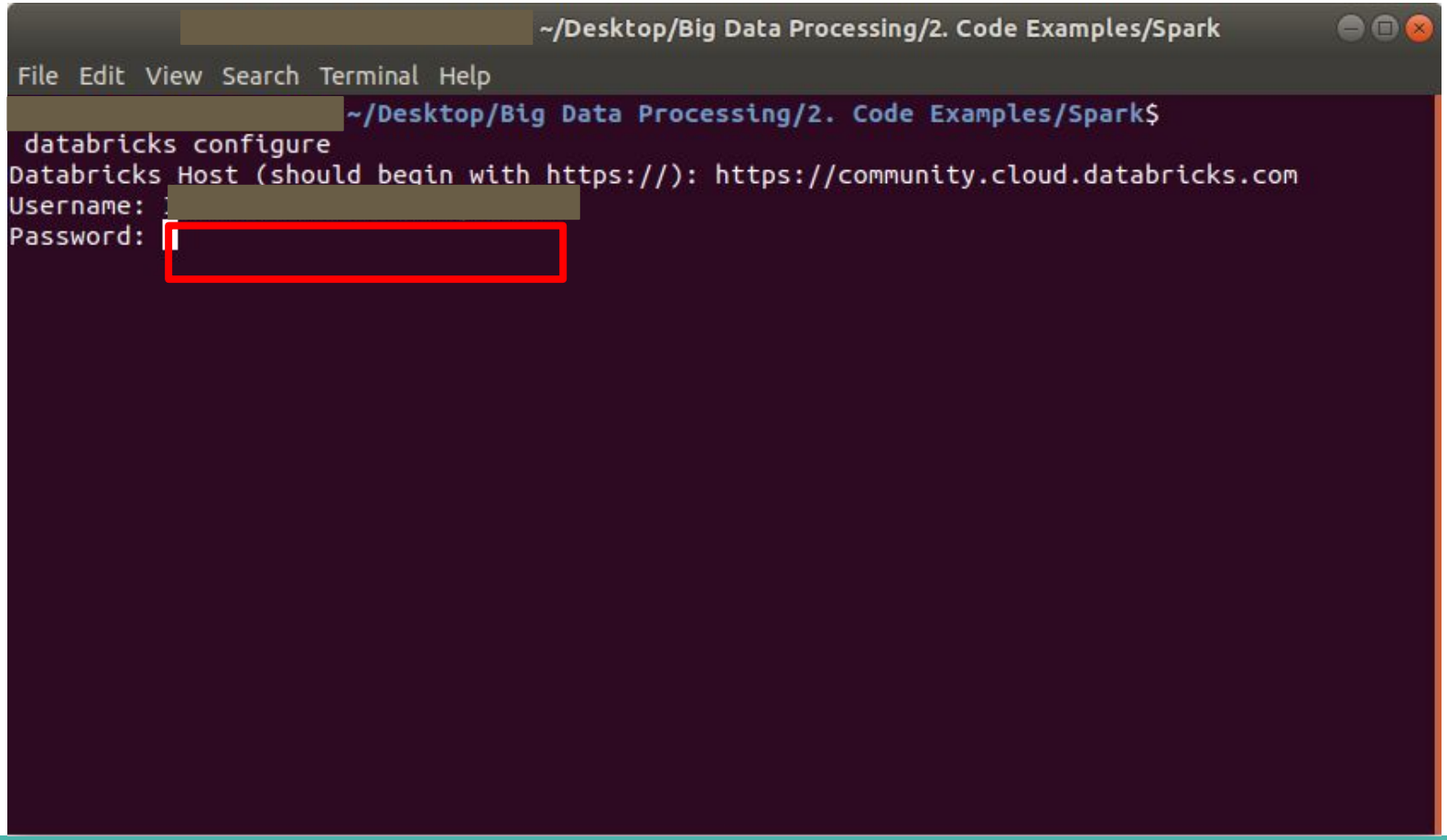
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   b. We must indicate our username (same as when we sign in at https://community.cloud.databricks.com).



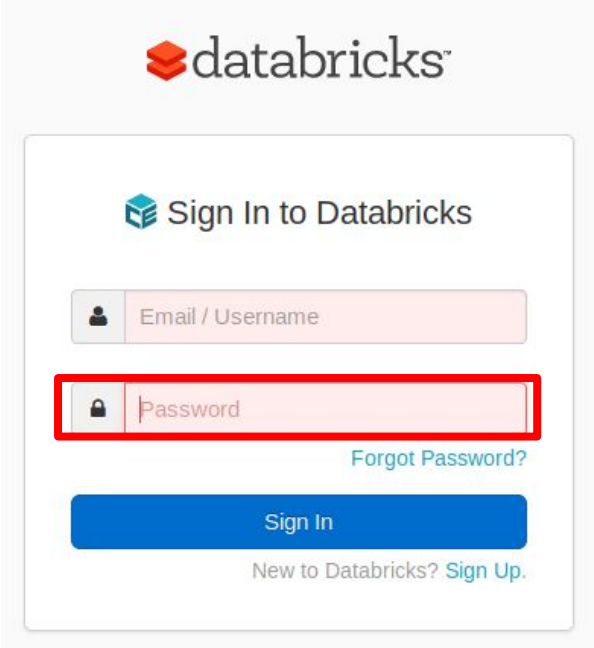
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   c. We must indicate our password (same as when we sign in at https://community.cloud.databricks.com).



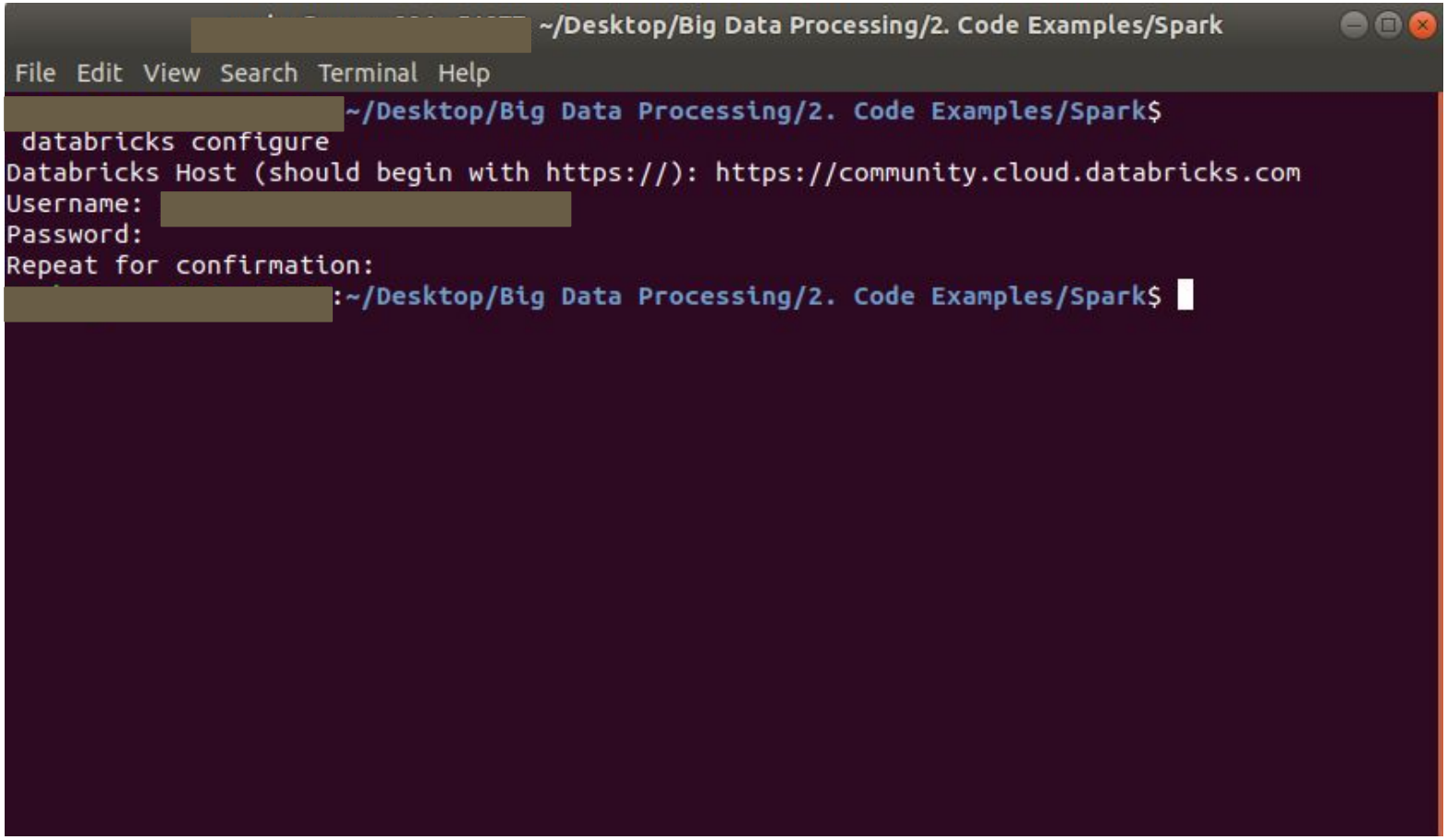
Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   c. We must indicate our password (same as when we sign in at https://community.cloud.databricks.com).



Databricks: An Online Platform for Data Engineers

6. Databricks provides a Command Line Interface (CLI).
   c. We must indicate our password (same as when we sign in at https://community.cloud.databricks.com).



Databricks: An Online Platform for Data Engineers

How to...

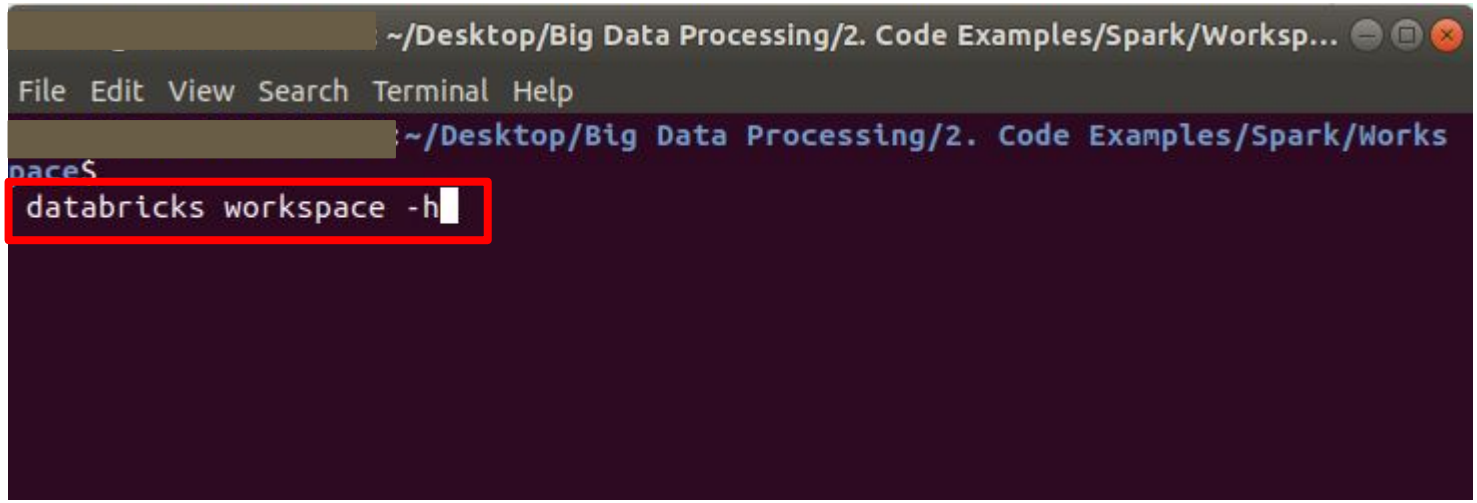
Upload an entire directory
from our local machine
to our Databricks Workspace

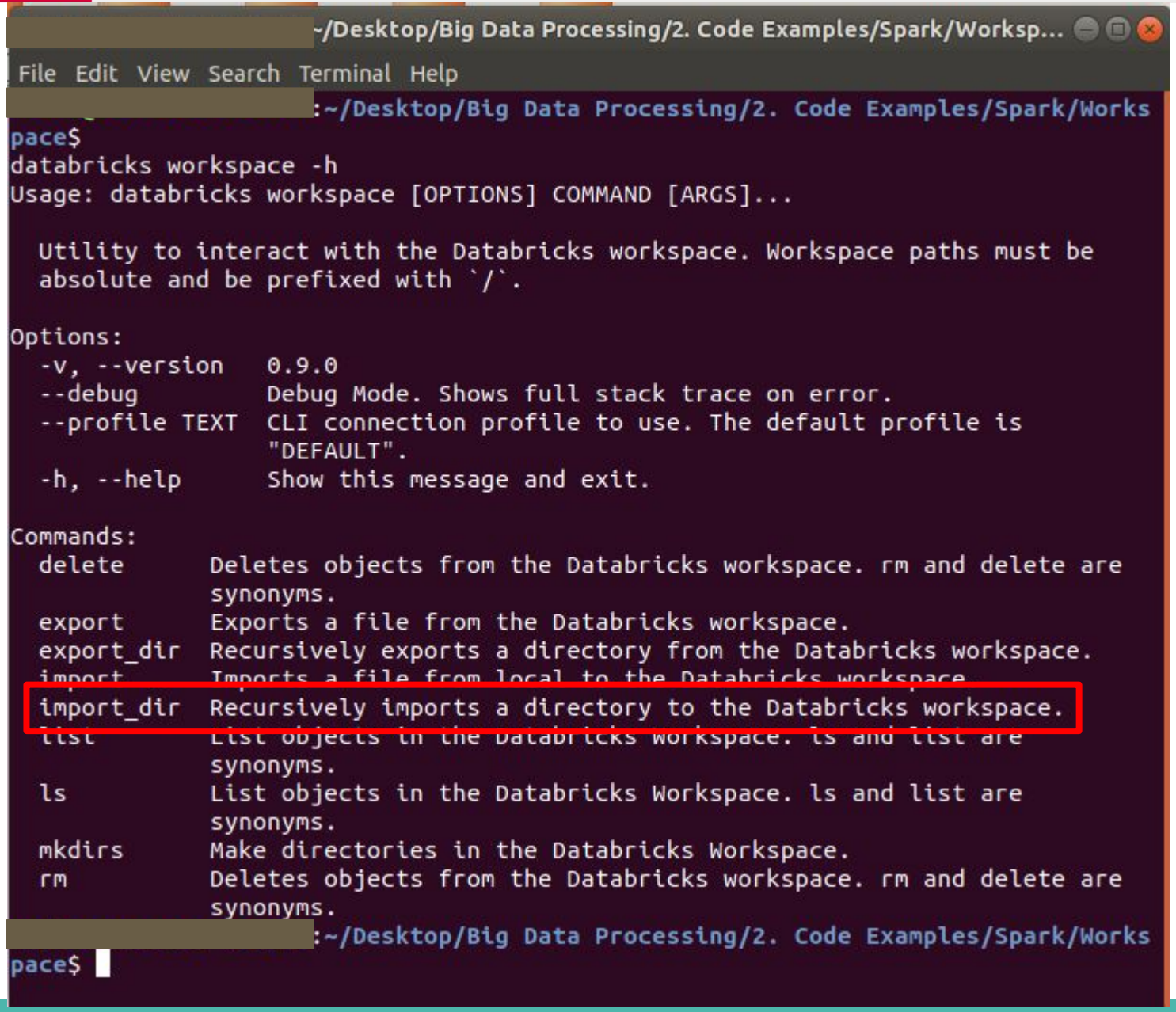
Databricks: An Online Platform for Data Engineers

7. The Databricks CLI has some **workspace** specific commands:

Databricks: An Online Platform for Data

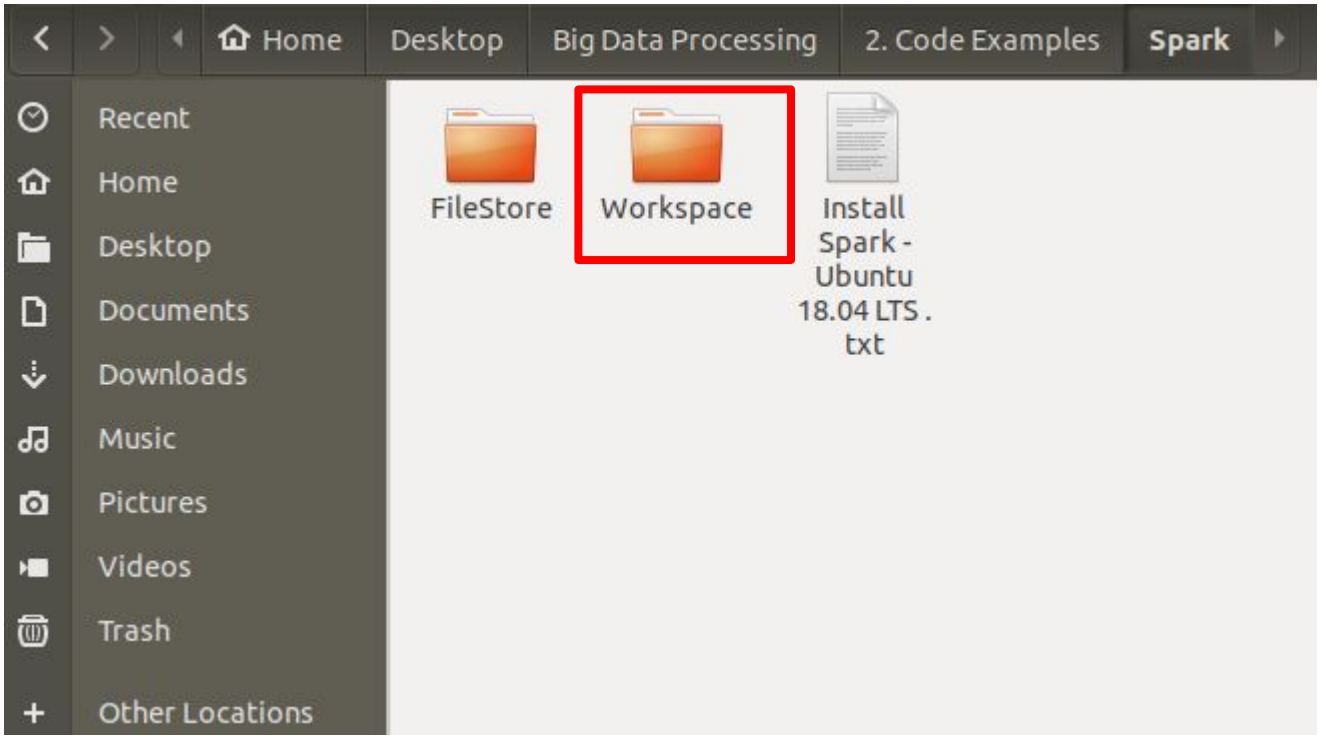
7. The Databricks CLI has some **workspace** specific commands:




~/Desktop/Big Data Processing/2. Code Examples/Spark/Worksp...
File Edit View Search Terminal Help
:~/Desktop/Big Data Processing/2. Code Examples/Spark/Works
pace$
databricks workspace -h
Usage: databricks workspace [OPTIONS] COMMAND [ARGS]...
Utility to interact with the Databricks workspace. Workspace paths must be absolute and be prefixed with `/`.
Options:
-v, --version 0.9.0
--debug Debug Mode. Shows full stack trace on error.
--profile TEXT CLI connection profile to use. The default profile is "DEFAULT".
-h, --help Show this message and exit.
Commands:
delete Deletes objects from the Databricks workspace. rm and delete are synonyms.
export Exports a file from the Databricks workspace.
export_dir Recursively exports a directory from the Databricks workspace.
import_dir Recursively imports a directory to the Databricks workspace.
ls List objects in the Databricks Workspace. ls and list are synonyms.
mkdirs Make directories in the Databricks Workspace.
rm Deletes objects from the Databricks workspace. rm and delete are synonyms.
:~/Desktop/Big Data Processing/2. Code Examples/Spark/Works
pace$

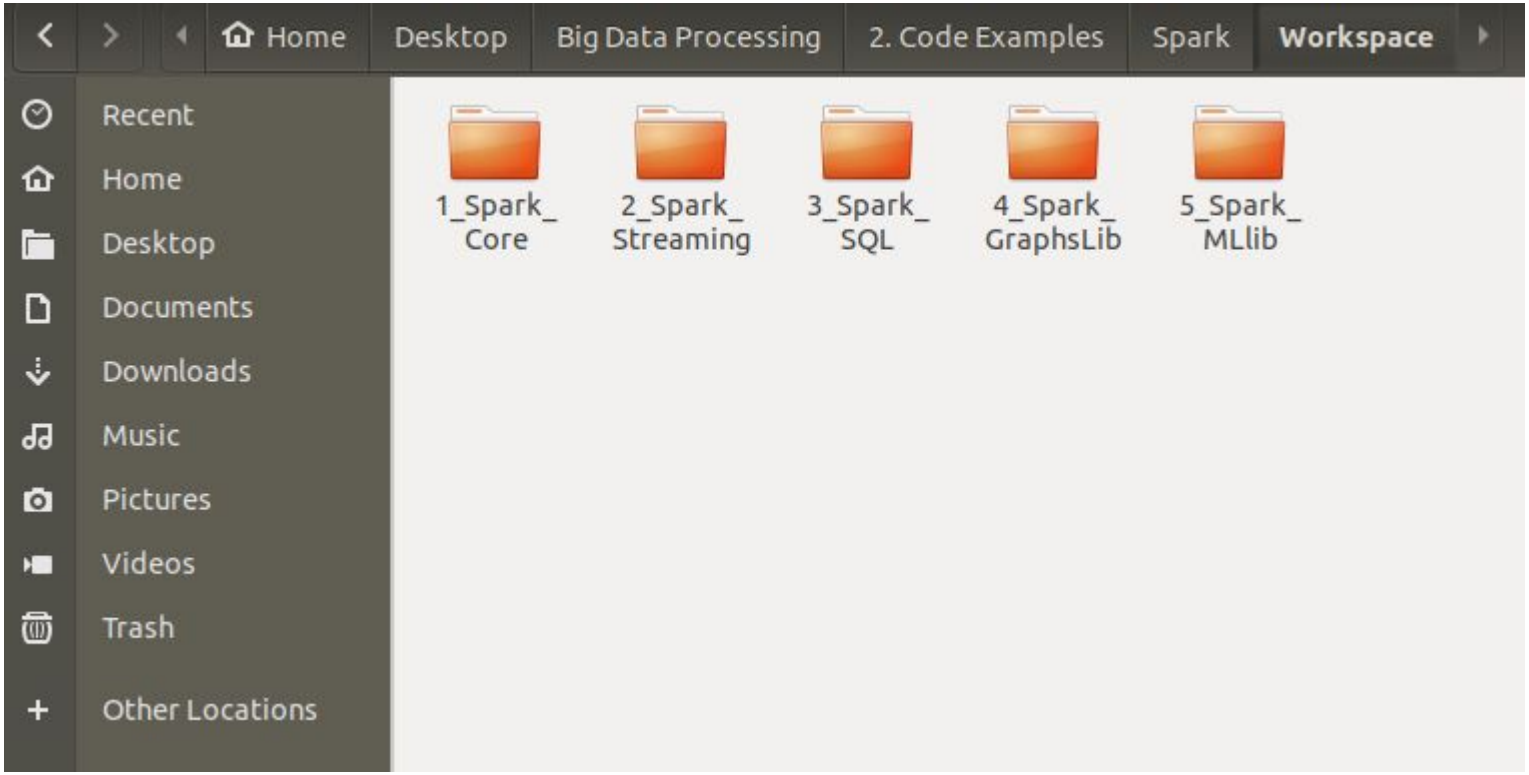
Databricks: An Online Platform for Data Engineers

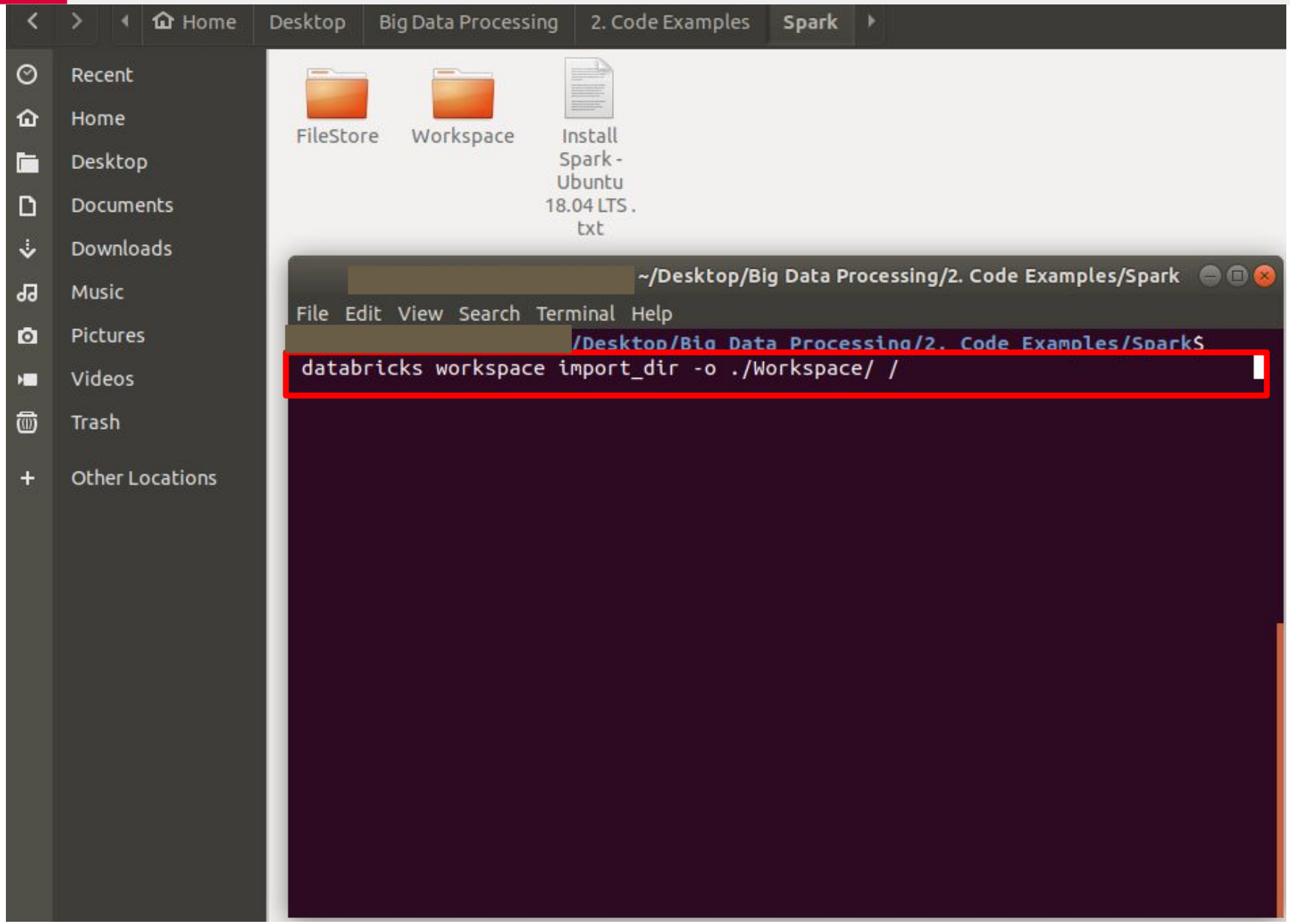
7. We can use the command **import_dir** command to upload our set of Spark code examples (available in Canvas) to our Databricks Workspace.



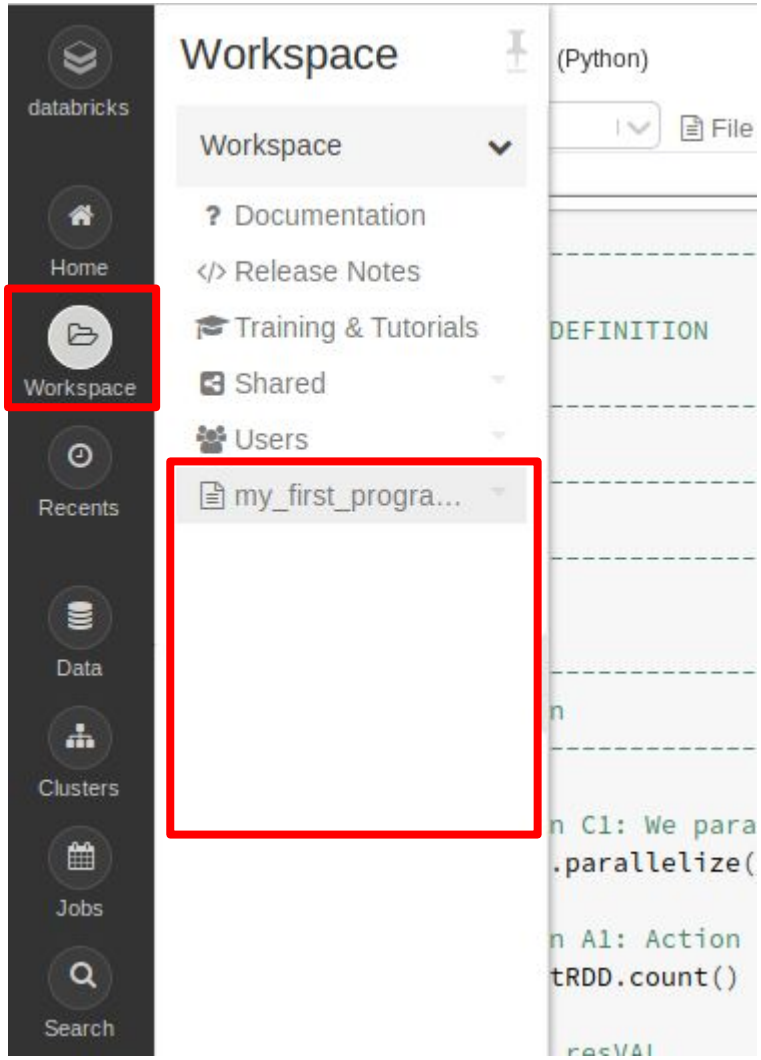
Databricks: An Online Platform for Data Engineers

7. We can use the command **import_dir** command to upload our set of Spark code examples (available in Canvas) to our Databricks Workspace.




Home
Desktop
Big Data Processing
2. Code Examples
Spark
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
FileStore
Workspace
Install Spark - Ubuntu 18.04 LTS . txt
~/Desktop/Big Data Processing/2. Code Examples/Spark
File Edit View Search Terminal Help
databricks workspace import_dir -o ./Workspace/ /

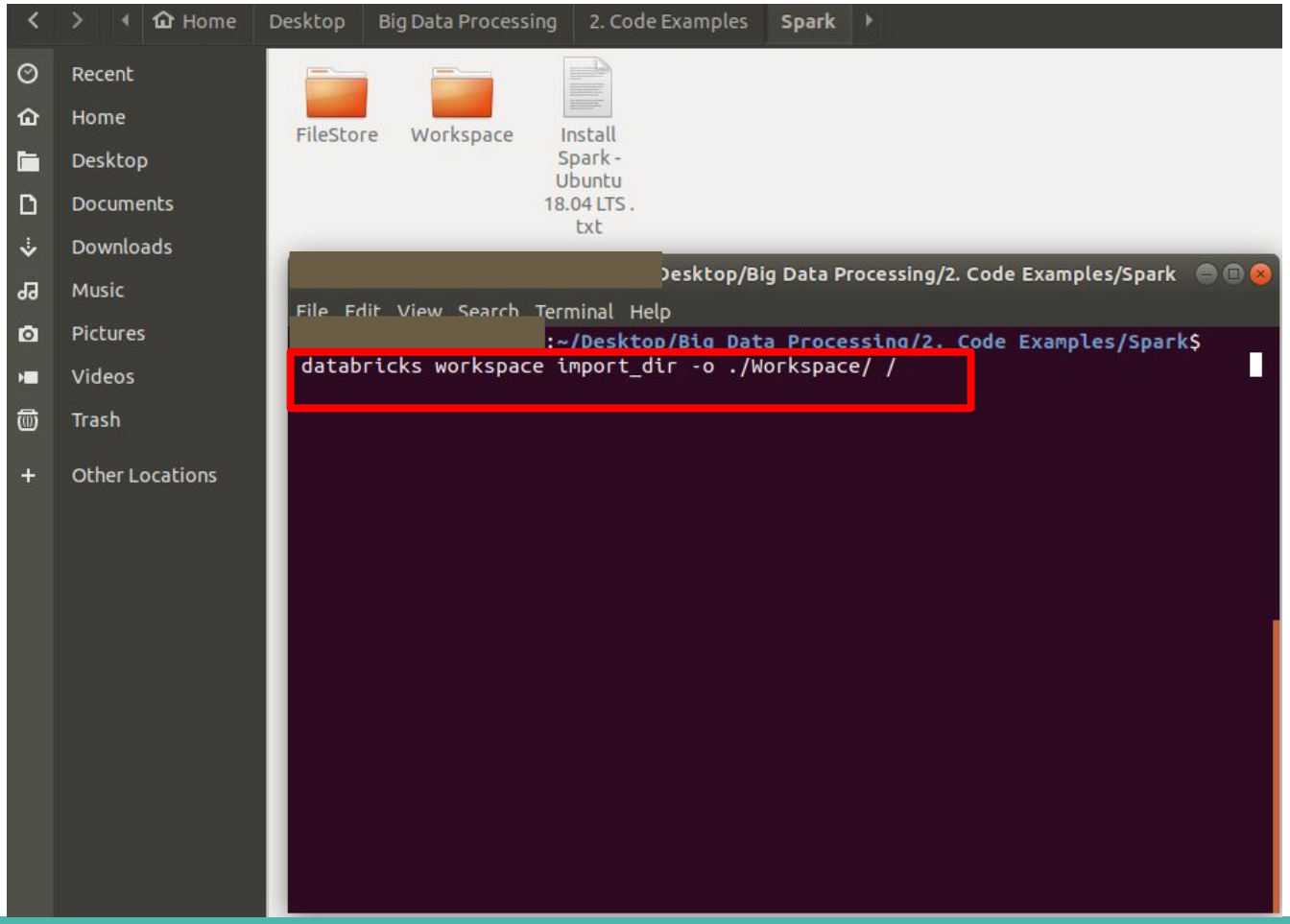
Databricks: An Online Platform for Data Engin

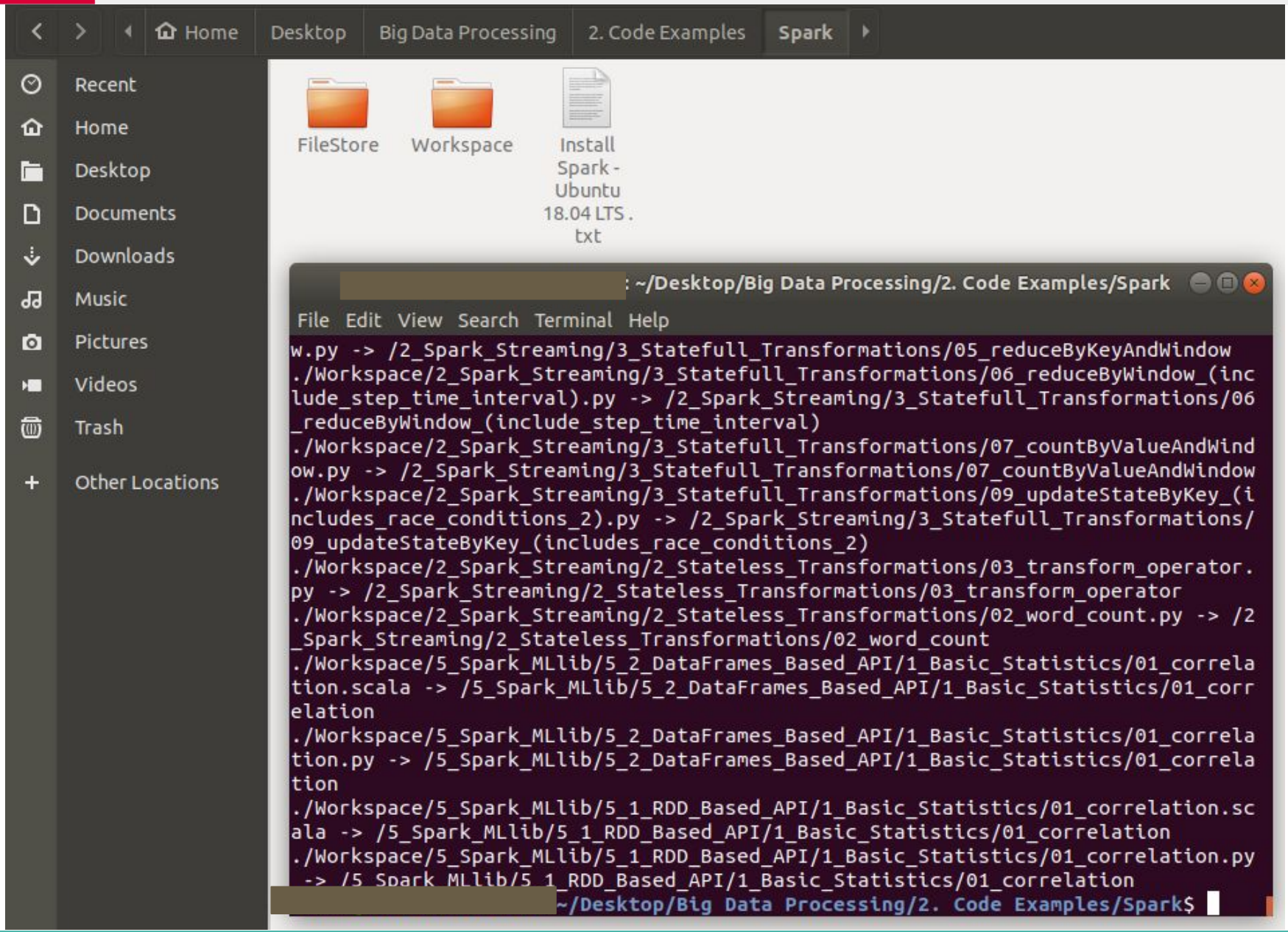


As we can see, before running the command, our **Databricks Workspace** only has the Python program we created before.

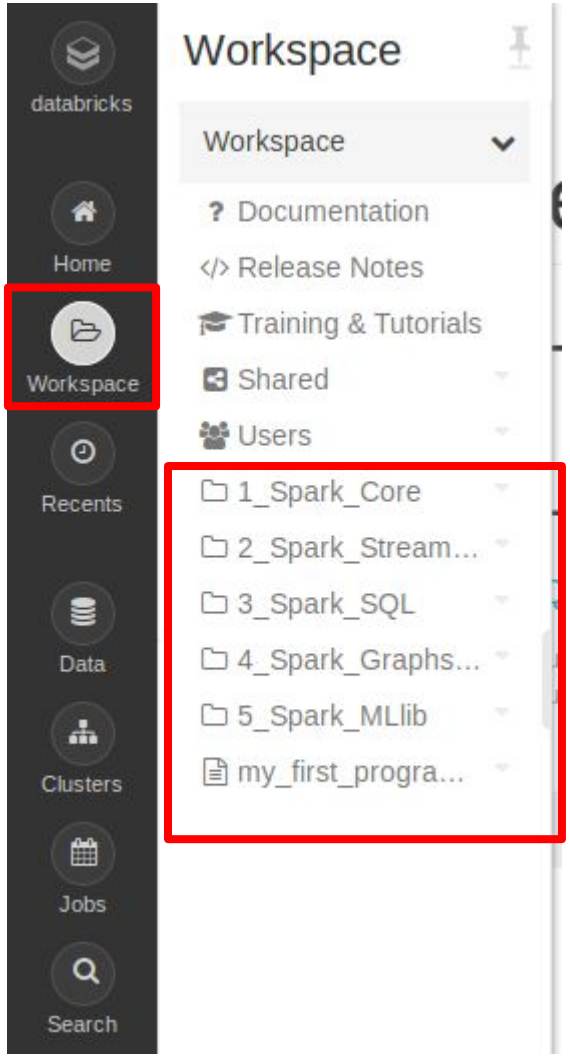
Databricks: An Online Platform for Data Engin

7. Running the command will take a couple of minutes.
   > databricks workspace import_dir -o ./Workspace/ /




Home
Desktop
Big Data Processing
2. Code Examples
Spark
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Other Locations
FileStore
Workspace
Install Spark - Ubuntu 18.04 LTS .txt
~/Desktop/Big Data Processing/2. Code Examples/Spark
File Edit View Search Terminal Help
w.py -> /2_Spark_Streaming/3_Statefull_Transformations/05_reduceByKeyAndWindow
./Workspace/2_Spark_Streaming/3_Statefull_Transformations/06_reduceByWindow_(include_step_time_interval).py -> /2_Spark_Streaming/3_Statefull_Transformations/06_reduceByWindow_(include_step_time_interval)
./Workspace/2_Spark_Streaming/3_Statefull_Transformations/07_countByValueAndWindow.py -> /2_Spark_Streaming/3_Statefull_Transformations/07_countByValueAndWindow
./Workspace/2_Spark_Streaming/3_Statefull_Transformations/09_updateStateByKey_(includes_race_conditions_2).py -> /2_Spark_Streaming/3_Statefull_Transformations/09_updateStateByKey_(includes_race_conditions_2)
./Workspace/2_Spark_Streaming/2_Stateless_Transformations/03_transform_operator.py -> /2_Spark_Streaming/2_Stateless_Transformations/03_transform_operator
./Workspace/2_Spark_Streaming/2_Stateless_Transformations/02_word_count.py -> /2_Spark_Streaming/2_Stateless_Transformations/02_word_count
./Workspace/5_Spark_MLlib/5_2_DataFrames_Based_API/1_Basic_Statistics/01_correlation.scala -> /5_Spark_MLlib/5_2_DataFrames_Based_API/1_Basic_Statistics/01_correlation
./Workspace/5_Spark_MLlib/5_2_DataFrames_Based_API/1_Basic_Statistics/01_correlation.py -> /5_Spark_MLlib/5_2_DataFrames_Based_API/1_Basic_Statistics/01_correlation
./Workspace/5_Spark_MLlib/5_1_RDD_Based_API/1_Basic_Statistics/01_correlation.scala -> /5_Spark_MLlib/5_1_RDD_Based_API/1_Basic_Statistics/01_correlation
./Workspace/5_Spark_MLlib/5_1_RDD_Based_API/1_Basic_Statistics/01_correlation.py -> /5_Spark_MLlib/5_1_RDD_Based_API/1_Basic_Statistics/01_correlation
~/Desktop/Big Data Processing/2. Code Examples/Spark$

Databricks: An Online Platform for Data Engin



As we can see, after running the command, our **Databricks Workspace** contains our set of Spark code examples that we will use during the semester.

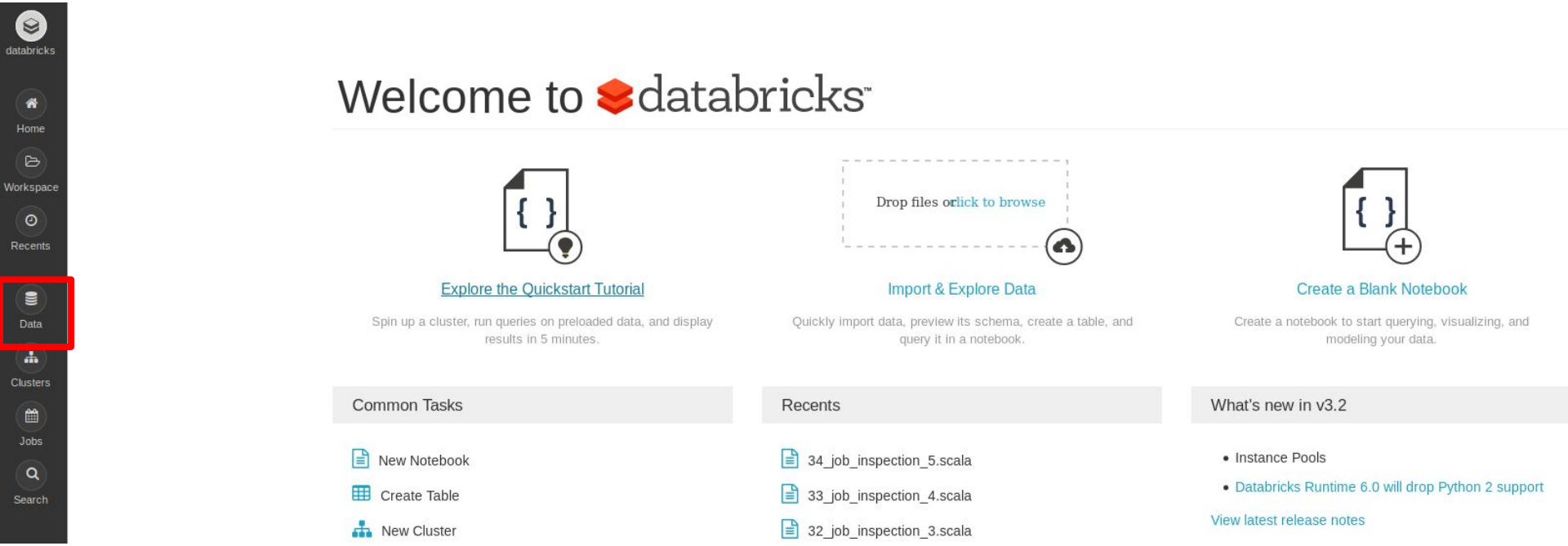
Databricks: An Online Platform for Data Engineers

How to...

Upload a Dataset
from our local machine
to the Databricks File System (DBFS).

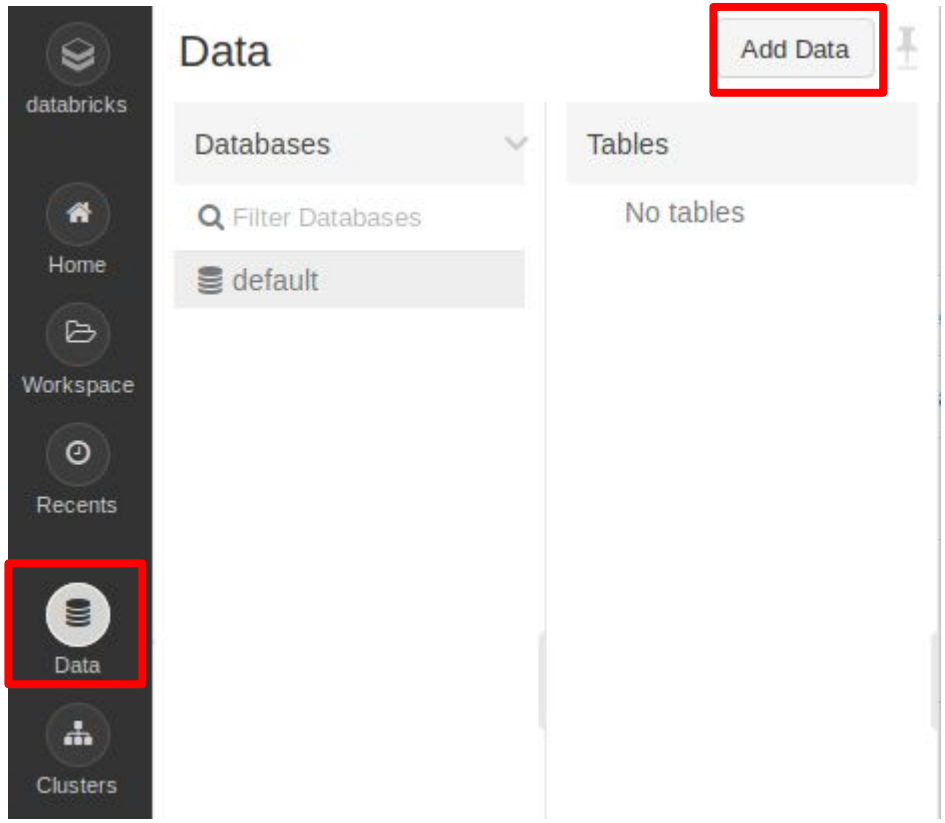
Databricks: An Online Platform for Data Engineers

8. We can see the files stored in the online Databricks File System (DBFS) by clicking in **Data**.



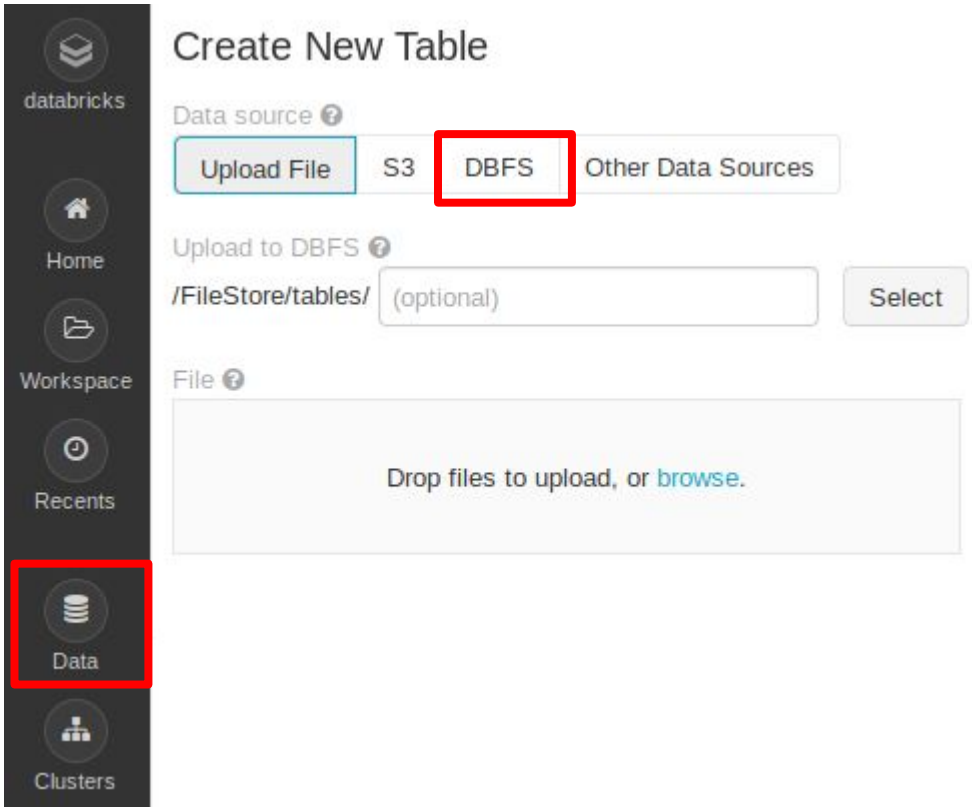
Databricks: An Online Platform for Data Engineers

8. We can see the files stored in the online Databricks File System (DBFS) by clicking in **Data**.



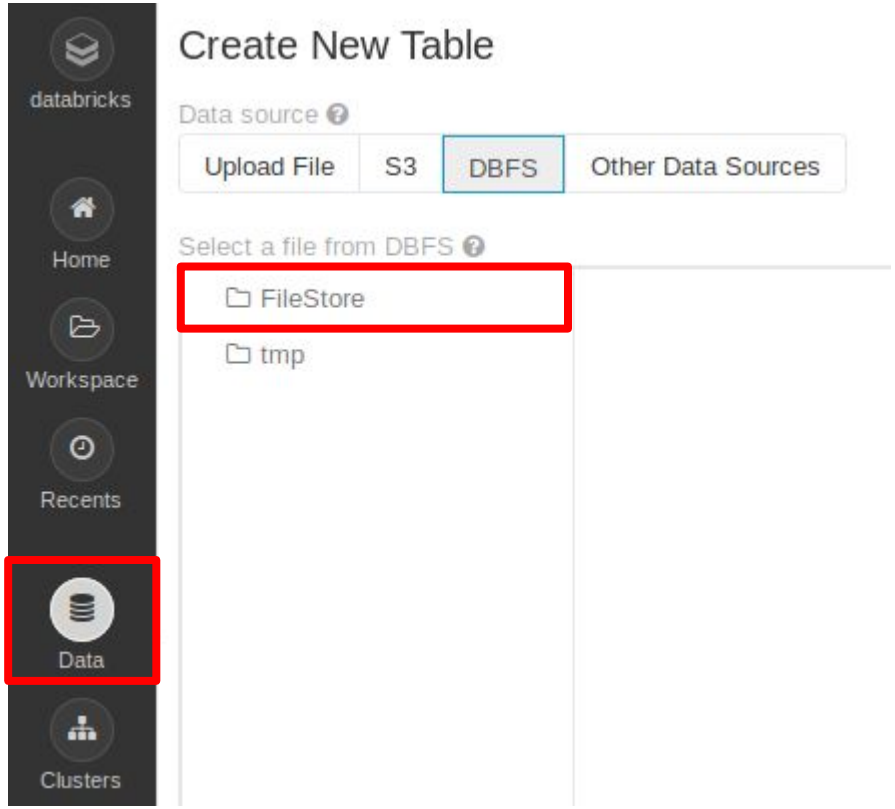
Databricks: An Online Platform for Data Engineers

8. We can see the files stored in the online Databricks File System (DBFS) by clicking in **Data**.



Databricks: An Online Platform for Data Engineers

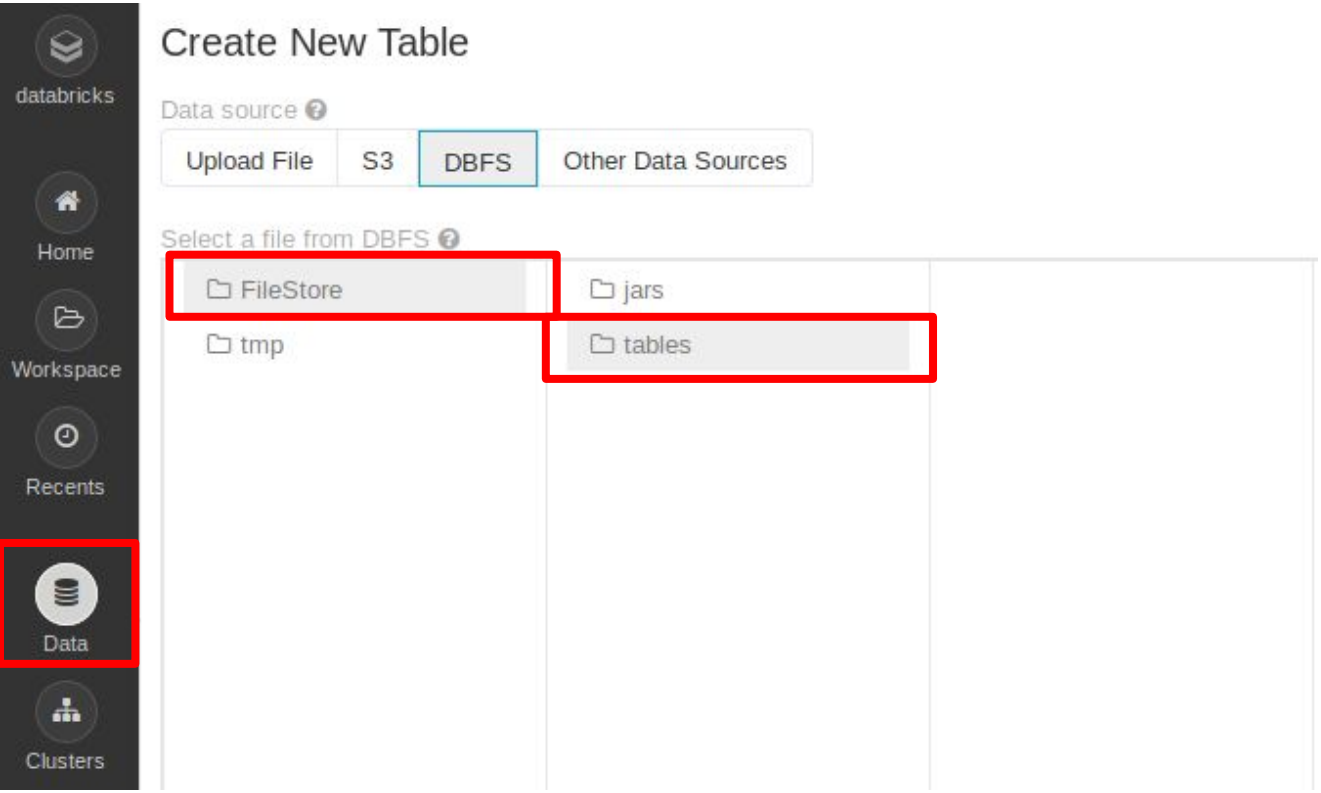
8. We can see the files stored in the online Databricks File System (DBFS) by clicking in **Data**.



Any **Dataset** being used during the semester is to be placed at **dbfs:/FileStore/tables/**

Databricks: An Online Platform for Data Engineers

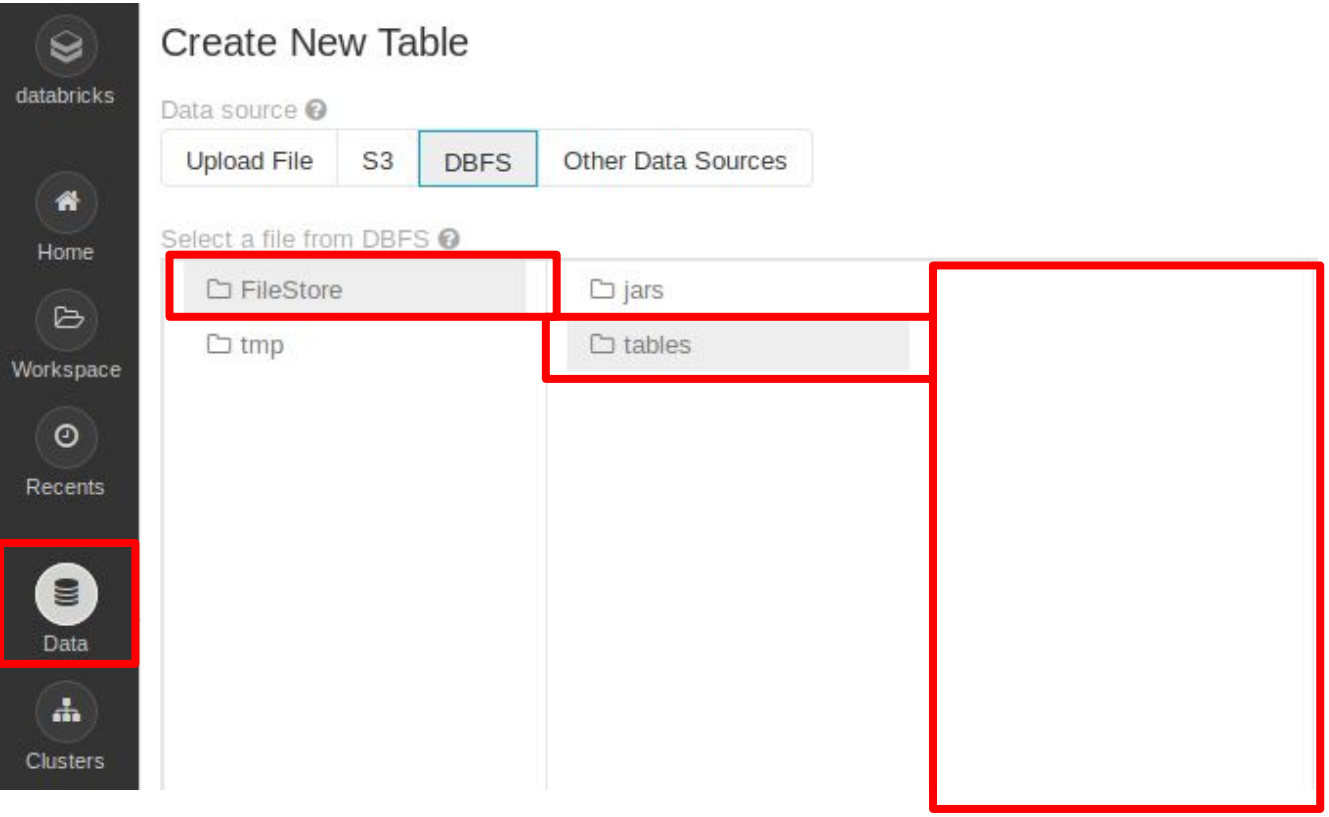
8. We can see the files stored in the online Databricks File System (DBFS) by clicking in **Data**.



Any **Dataset** being used during the semester is to be placed at **dbfs:/FileStore/tables/**

Databricks: An Online Platform for Data Engineers

8. We can see the files stored in the online Databricks File System (DBFS) by clicking in **Data**.



Any **Dataset** being used during the semester is to be placed at **dbfs:/FileStore/tables/**

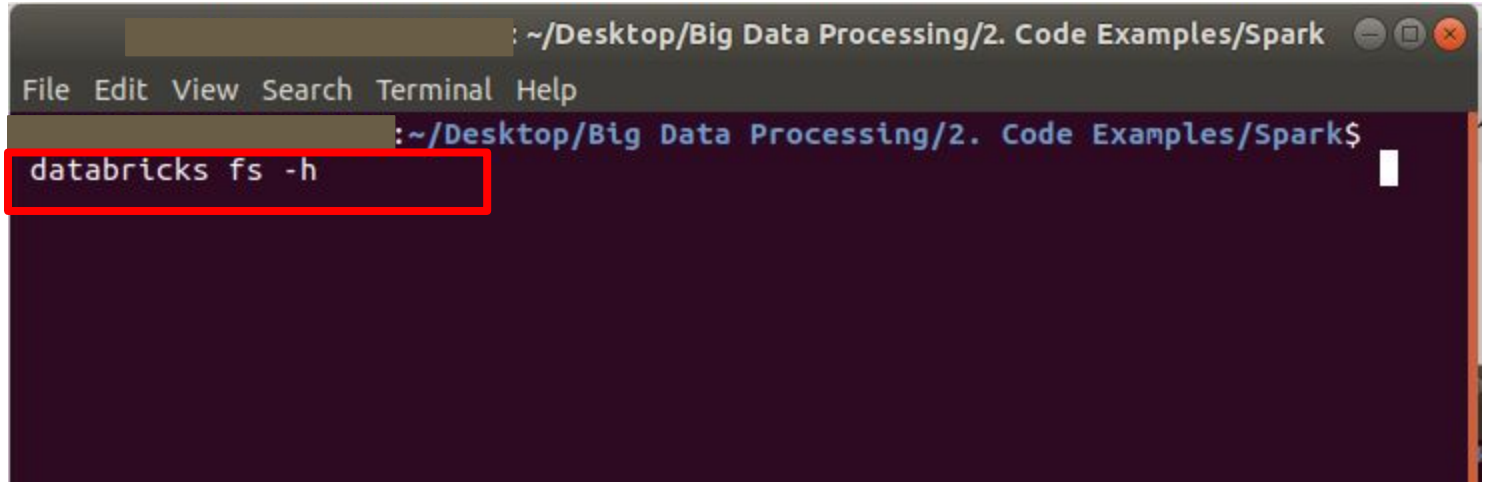
As we can see by the moment there is no Dataset store at DBFS.

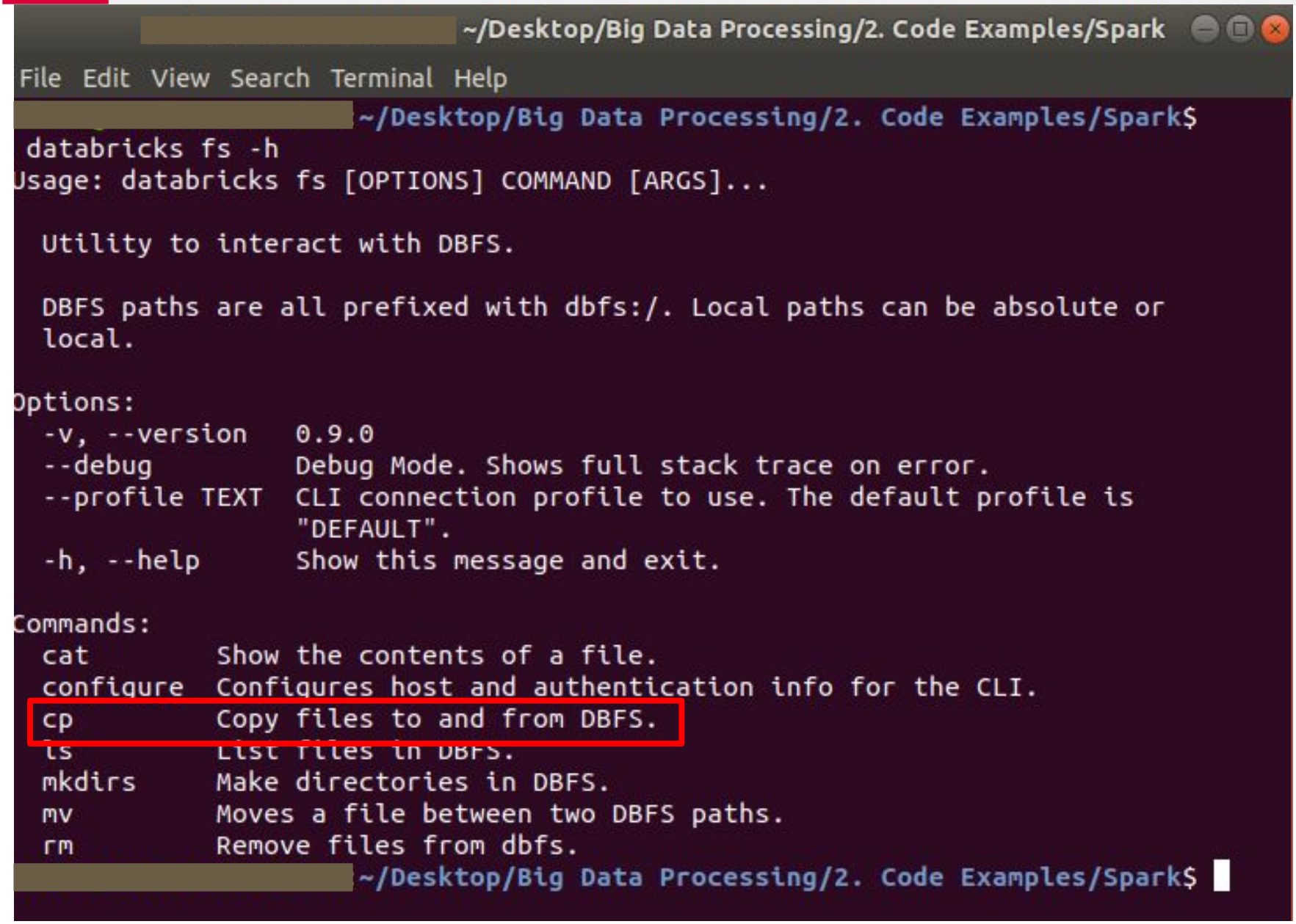
Databricks: An Online Platform for Data Engineers

8. The CLI has some **Databricks File System (DBFS)** specific commands:

Databricks: An Online Platform for Data Engineers

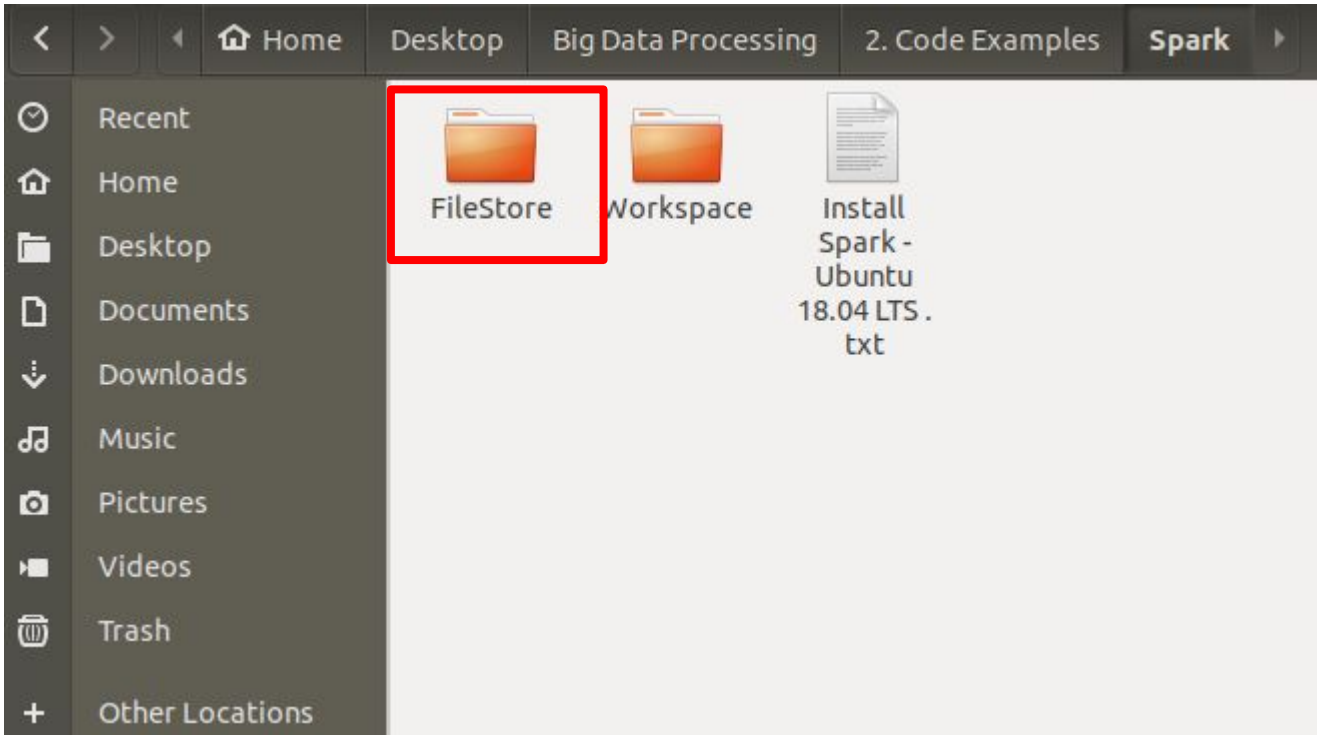
8. The CLI has some **Databricks File System (DBFS)** specific commands:




~/Desktop/Big Data Processing/2. Code Examples/Spark
File Edit View Search Terminal Help
~/Desktop/Big Data Processing/2. Code Examples/Spark$ databricks fs -h
Usage: databricks fs [OPTIONS] COMMAND [ARGS]...
Utility to interact with DBFS.
DBFS paths are all prefixed with dbfs:/. Local paths can be absolute or local.
Options:
-v, --version 0.9.0
--debug Debug Mode. Shows full stack trace on error.
--profile TEXT CLI connection profile to use. The default profile is "DEFAULT".
-h, --help Show this message and exit.
Commands:
cat Show the contents of a file.
configure Configures host and authentication info for the CLI.
cp Copy files to and from DBFS.
ls List files in DBFS.
mkdirs Make directories in DBFS.
mv Moves a file between two DBFS paths.
rm Remove files from dbfs.
~/Desktop/Big Data Processing/2. Code Examples/Spark$

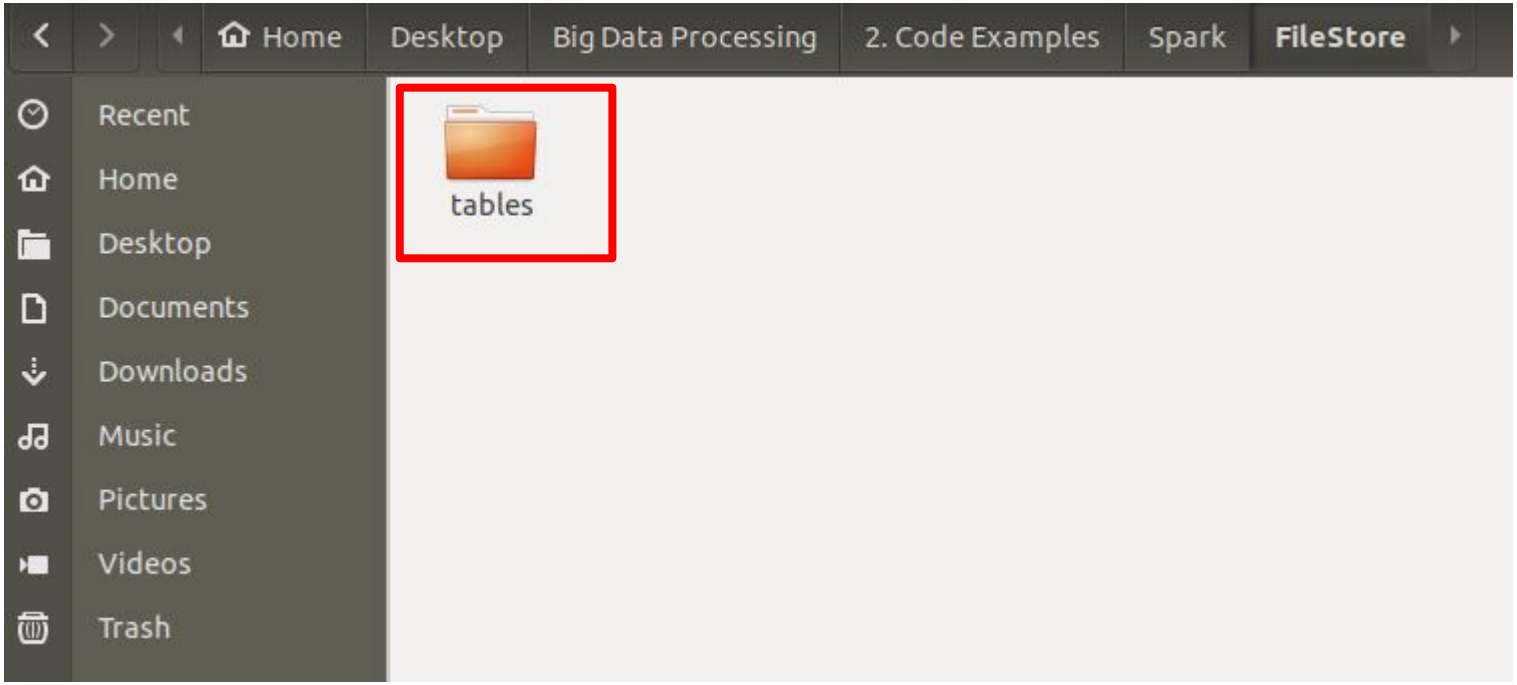
Databricks: An Online Platform for Data Engineers

8. We can use the command **cp** to upload our set of Datasets to DBFS.



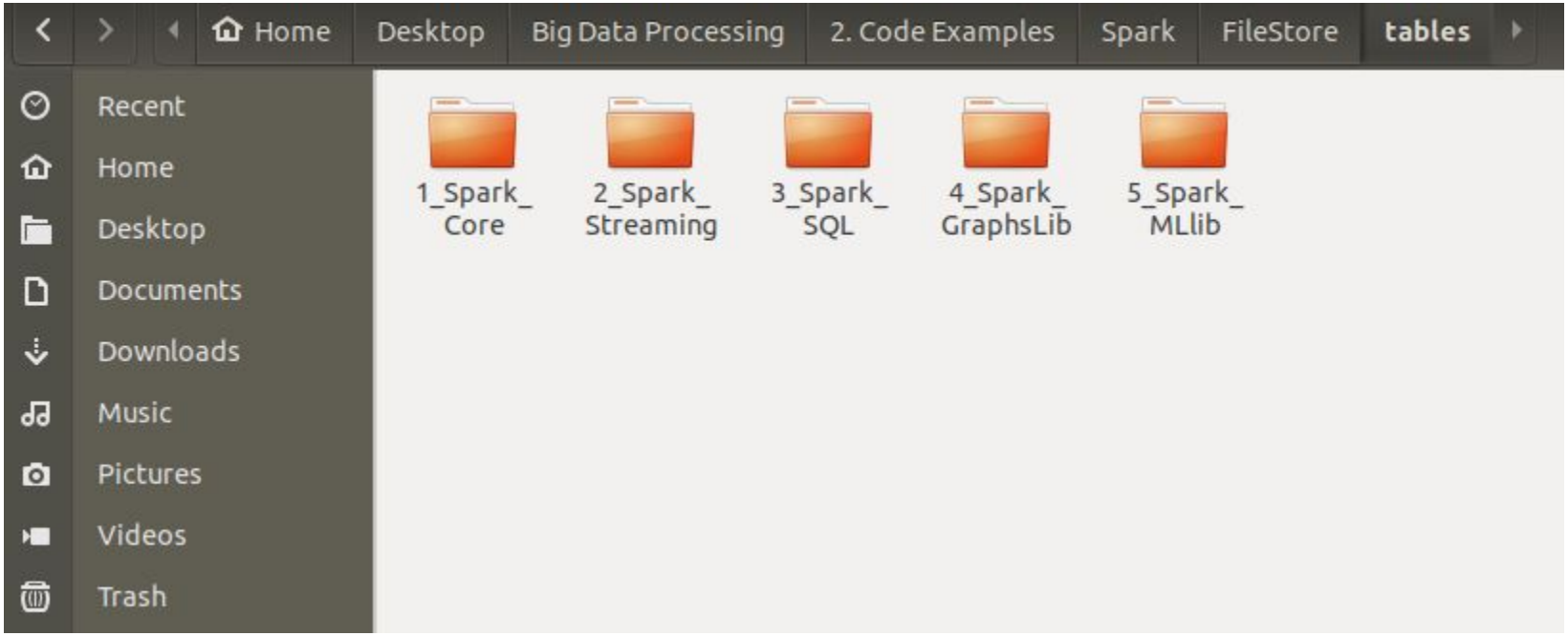
Databricks: An Online Platform for Data Engineers

8. We can use the command **cp** to upload our set of Datasets to DBFS.



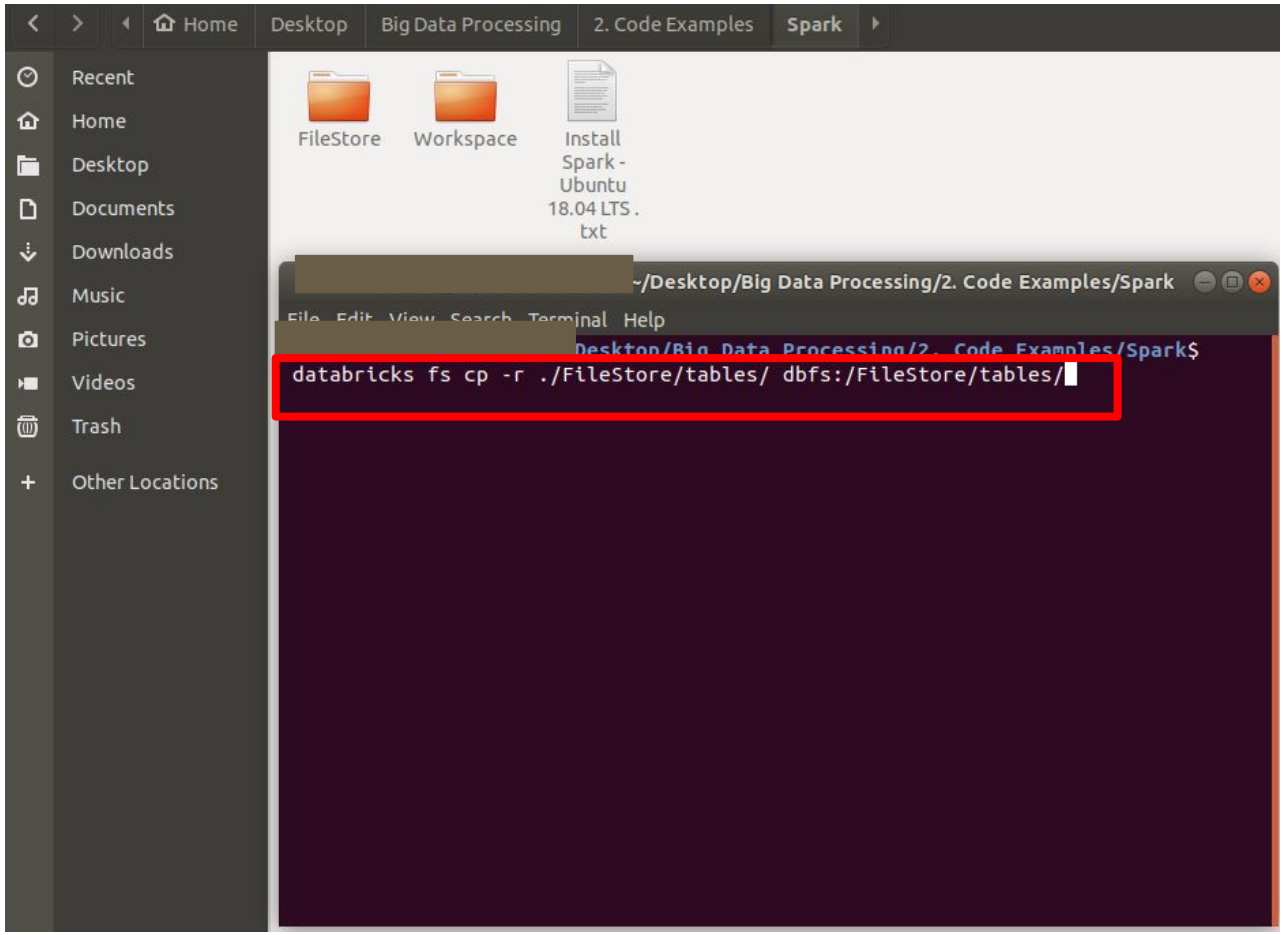
Databricks: An Online Platform for Data Engineers

8. We can use the command **cp** to upload our set of Datasets to DBFS.



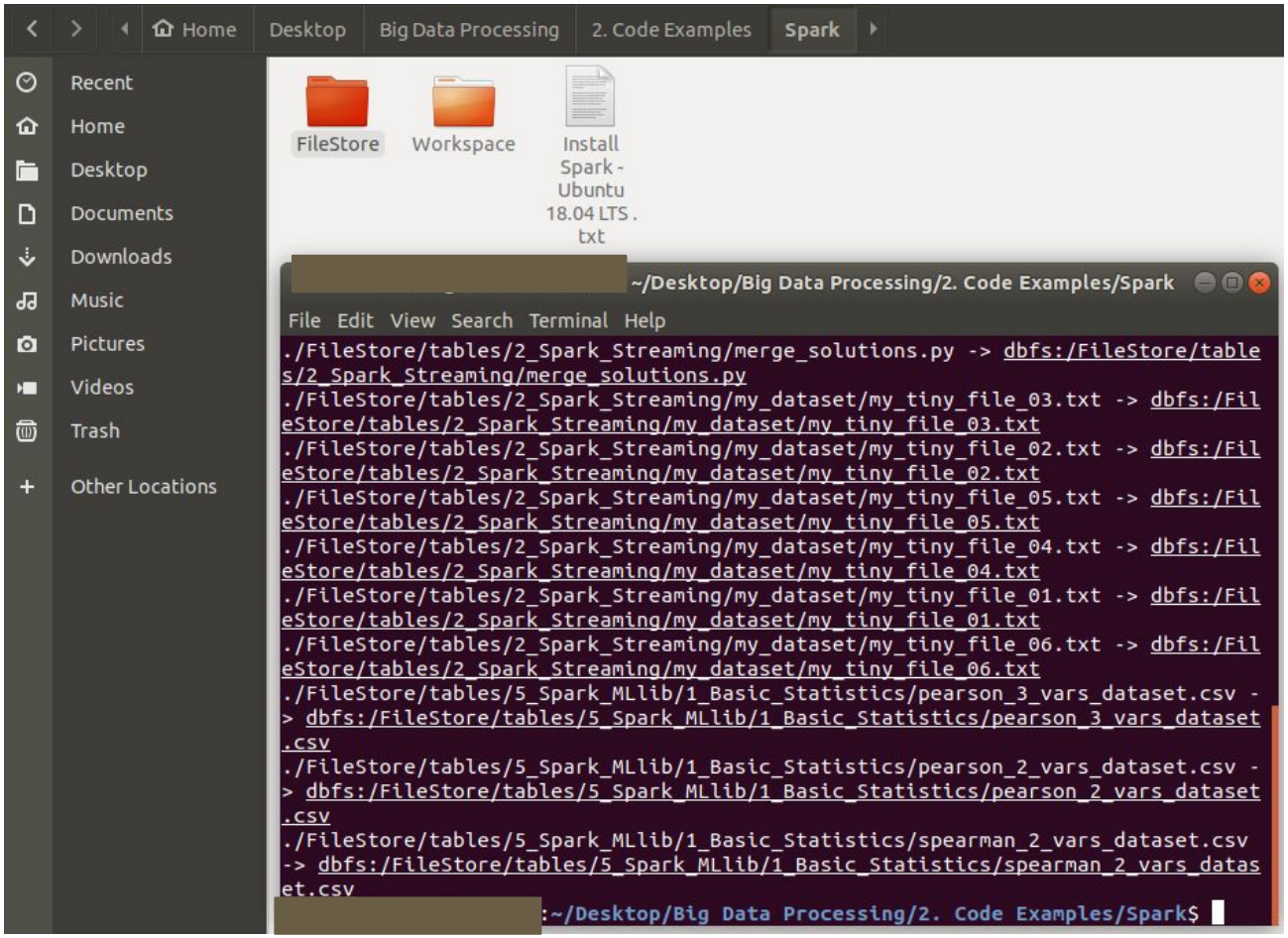
Databricks: An Online Platform for Data Engin

8. Running the command will take a couple of minutes.
   > databricks fs cp -r ./FileStore/tables dbfs:/FileStore/tables/

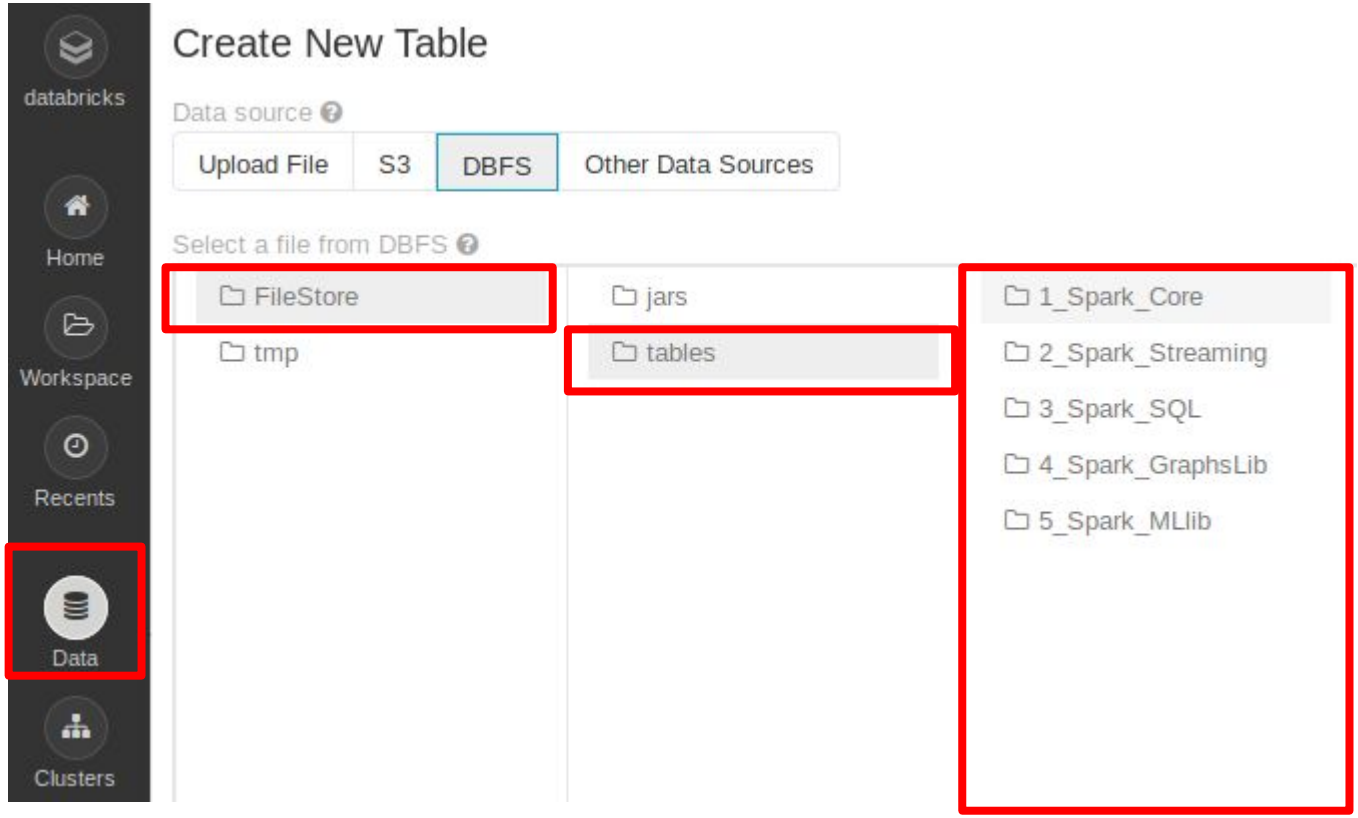


Databricks: An Online Platform for Data Engin

8. Running the command will take a couple of minutes.
   > databricks fs cp -r ./FileStore/tables dbfs:/FileStore/tables/



Databricks: An Online Platform for Data Engin



As we can see, after running the command, our **Datasets** are now stored at dbfs:/FileStore/tables/

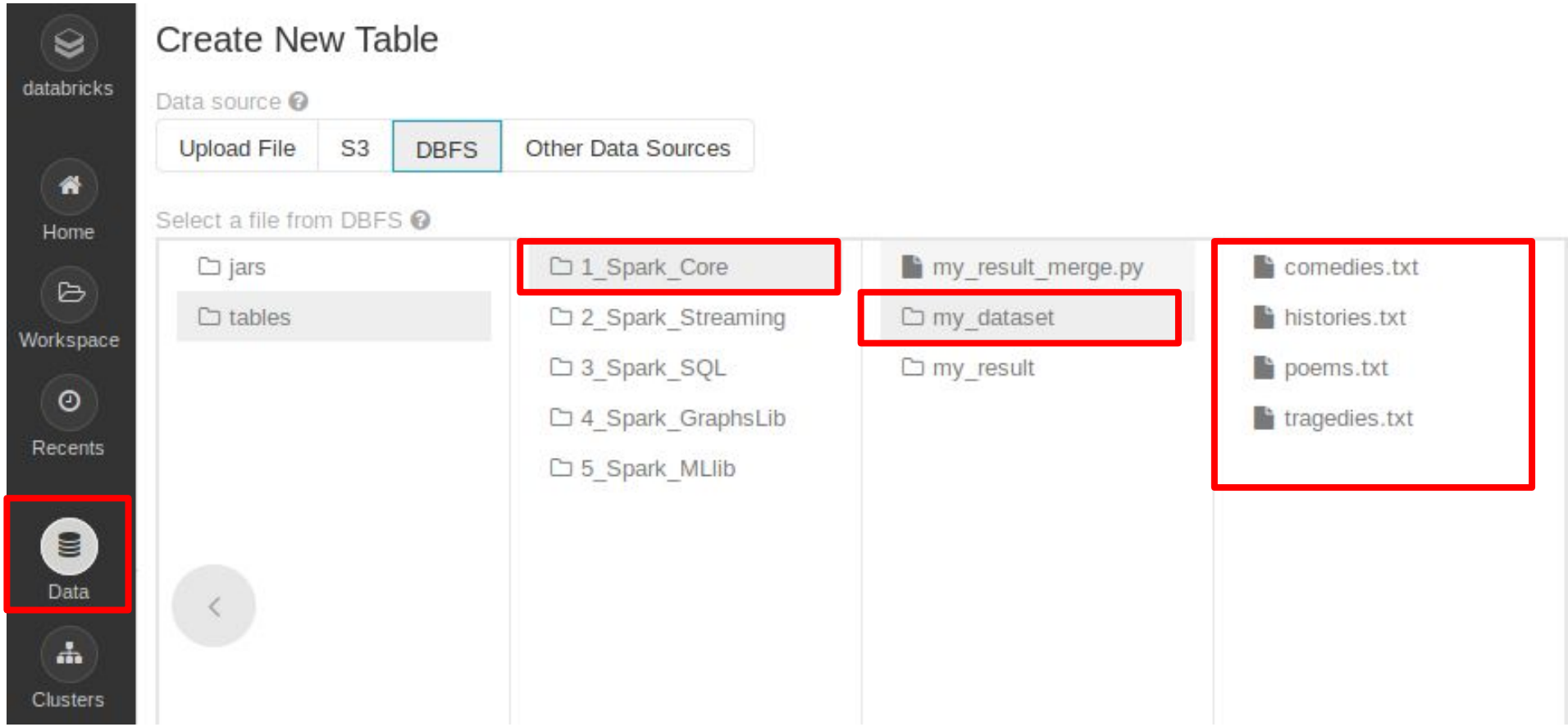
Databricks: An Online Platform for Data Engineers

How to...

Run a program / Spark Application
reading in a Dataset from DBFS
and possibly storing the results at DBFS.

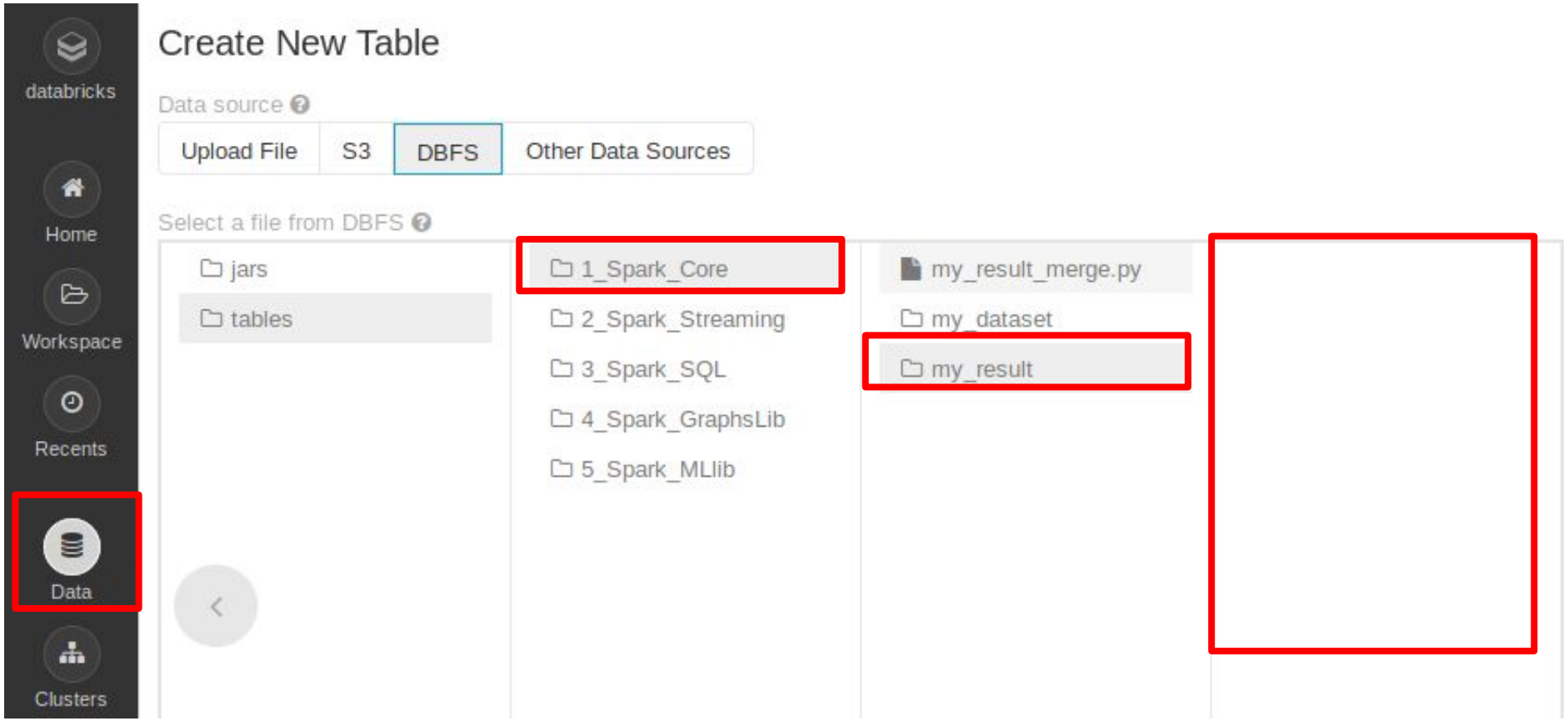
Databricks: An Online Platform for Data Engin

9. As we can see, Spark Core contains **my_dataset** of 4 text files:
   comedies.txt histories.txt poems.txt tragedies.txt



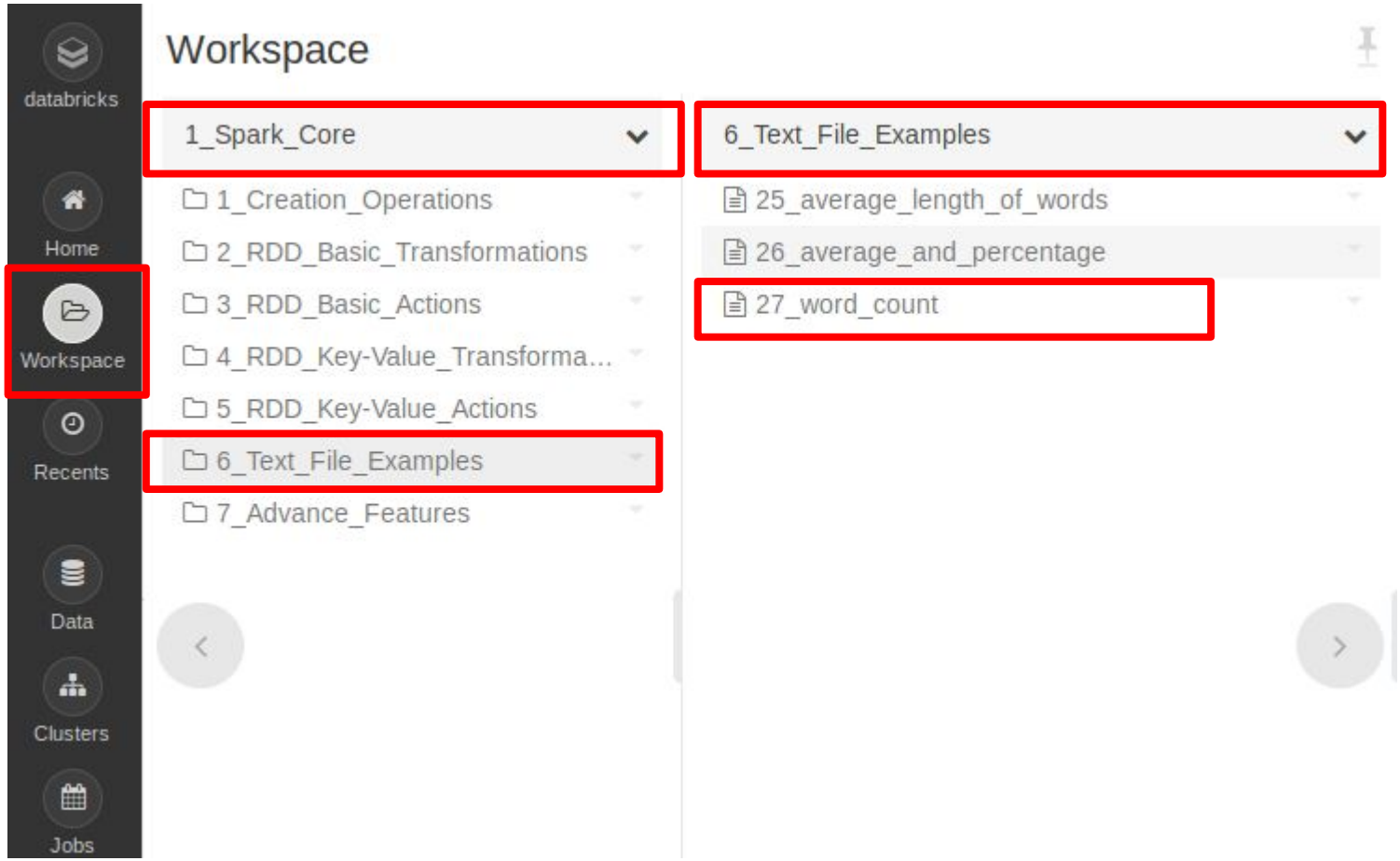
Databricks: An Online Platform for Data Engin

9. As we can see, Spark Core contains the folder **my_result**, being empty.



Databricks: An Online Platform for Data Engin

9. Without explaining the program itself, our code example:
   Workspace/1_Spark_Core/6_Text_File_Examples/27_word_count.py



Databricks: An Online Platform for Data Engin

9. Without explaining the program itself, our code example Workspace/1_Spark_Core/6_Text_File_Examples/27_word_count.py

   ➢ Reads the **Dataset** from **dbfs:/FileStore/1_Spark_Core/my_dataset**

Databricks: An Online Platform for Data Engin

9. Without explaining the program itself, our code example Workspace/1_Spark_Core/6_Text_File_Examples/27_word_count.py

   - ➢ Reads the **Dataset** from **dbfs:/FileStore/1_Spark_Core/my_dataset**
   - ➢ Computes the word count of the words appearing in the files of my_dataset: comedies.txt, histories.txt, poems.txt, tragedies.txt

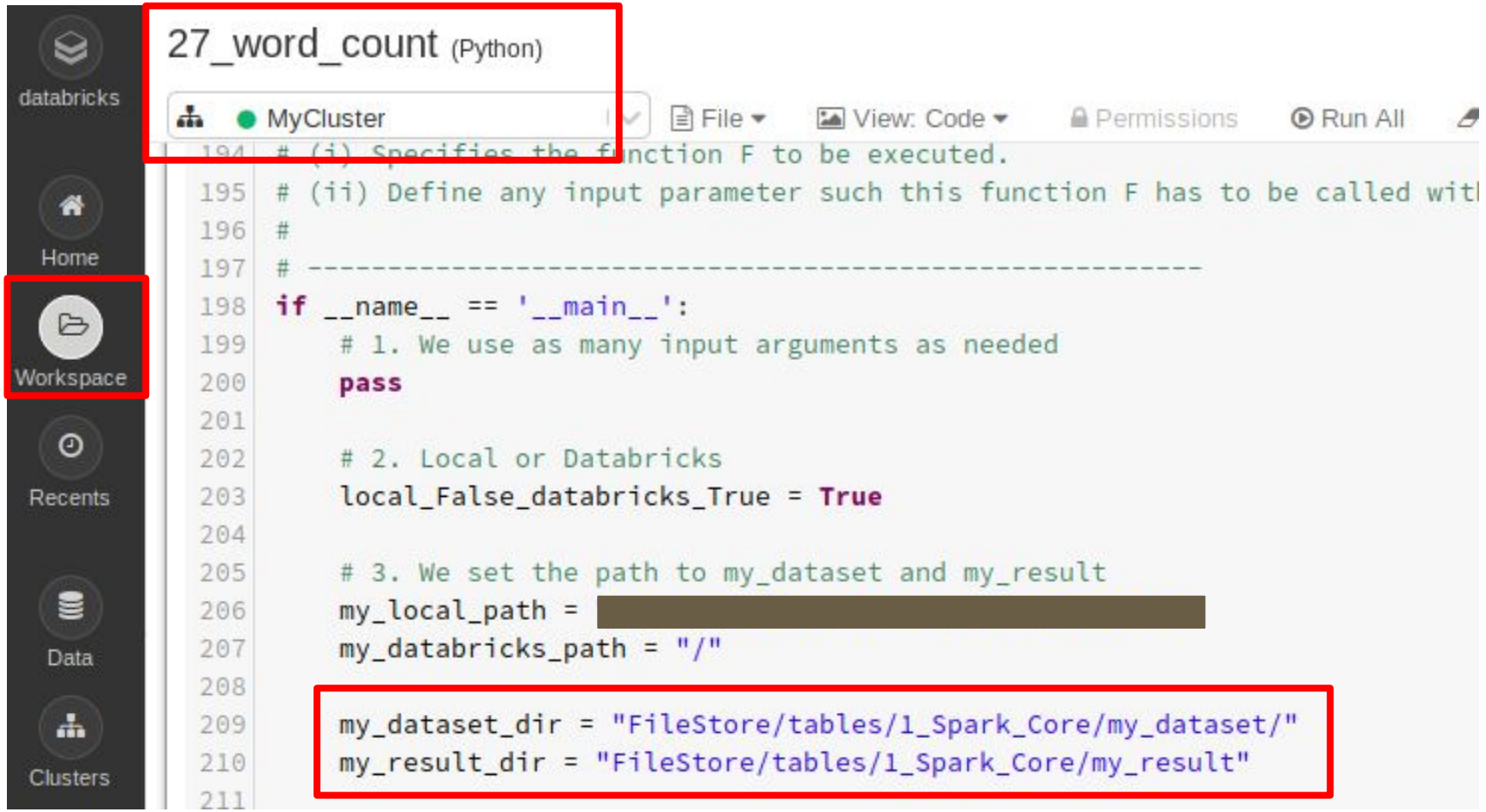
Databricks: An Online Platform for Data Engin

9. Without explaining the program itself, our code example Workspace/1_Spark_Core/6_Text_File_Examples/27_word_count.py

   - ➢ Reads the **Dataset** from **dbfs:/FileStore/1_Spark_Core/my_dataset**
   - ➢ Computes the word count of the words appearing in the files of my_dataset: comedies.txt, histories.txt, poems.txt, tragedies.txt
   - ➢ Writes the results to new files placed in the directory **dbfs:/FileStore/1_Spark_Core/my_result**

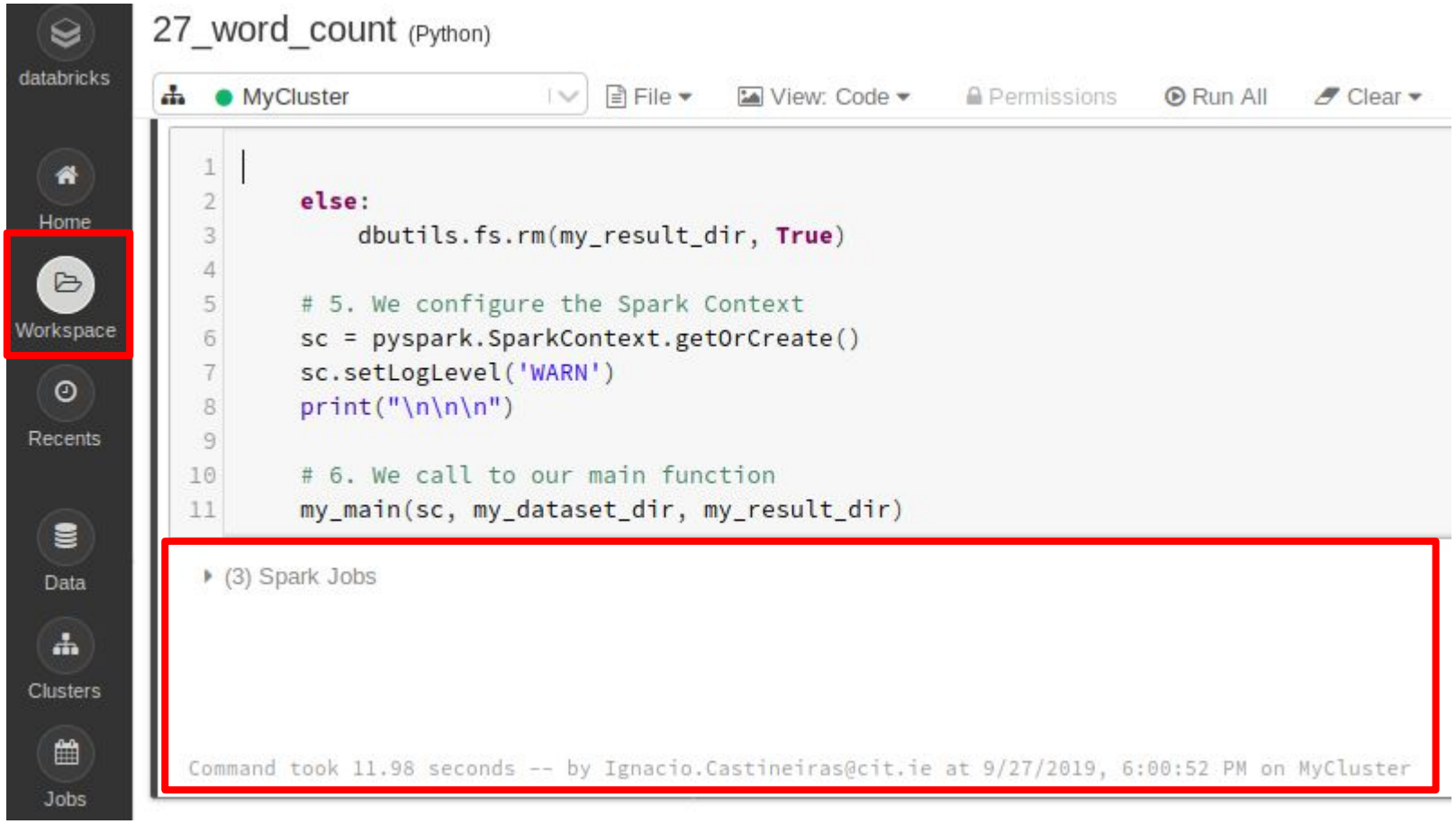
Databricks: An Online Platform for Data Engin

9. Without explaining the program itself, our code example Workspace/1_Spark_Core/6_Text_File_Examples/27_word_count.py



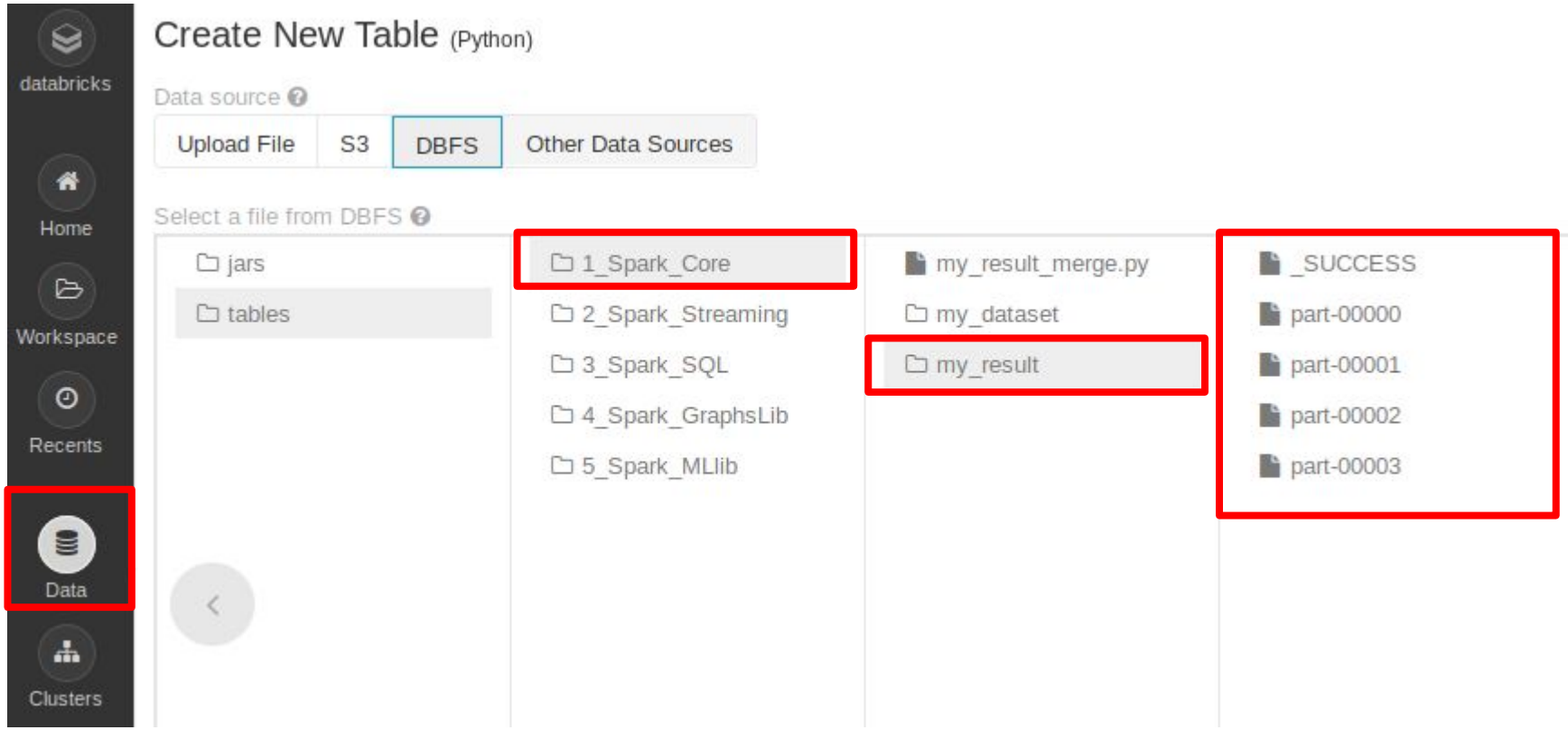
Databricks: An Online Platform for Data Engin

9. Running the program / Spark application leads to no visual results:
   Workspace/1_Spark_Core/6_Text_File_Examples/27_word_count.py



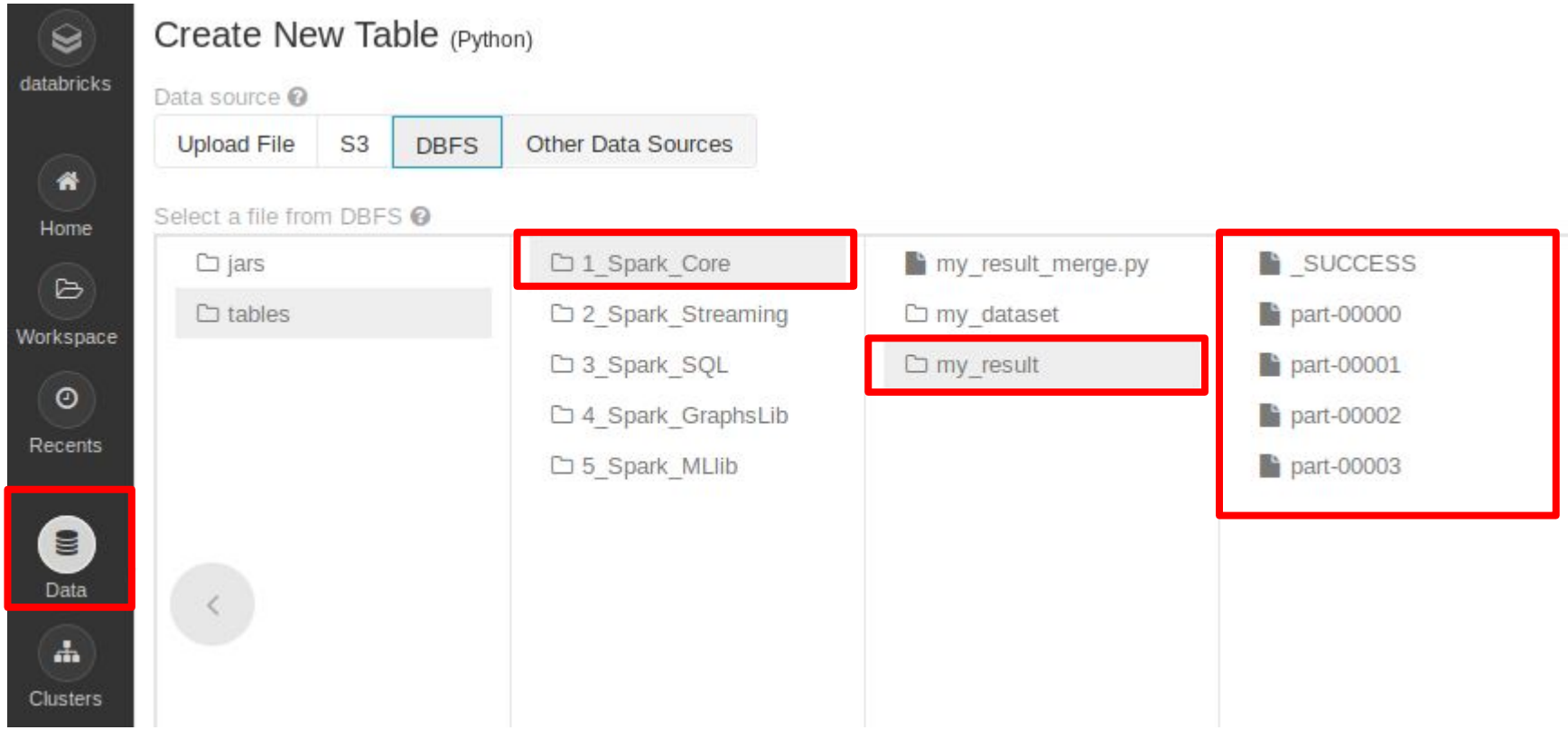
Databricks: An Online Platform for Data Engin

9. However, the folder **my_result** in the DBFS has been populated:



Databricks: An Online Platform for Data Engin

9. Unfortunately we cannot visualise the content of these DBFS files.

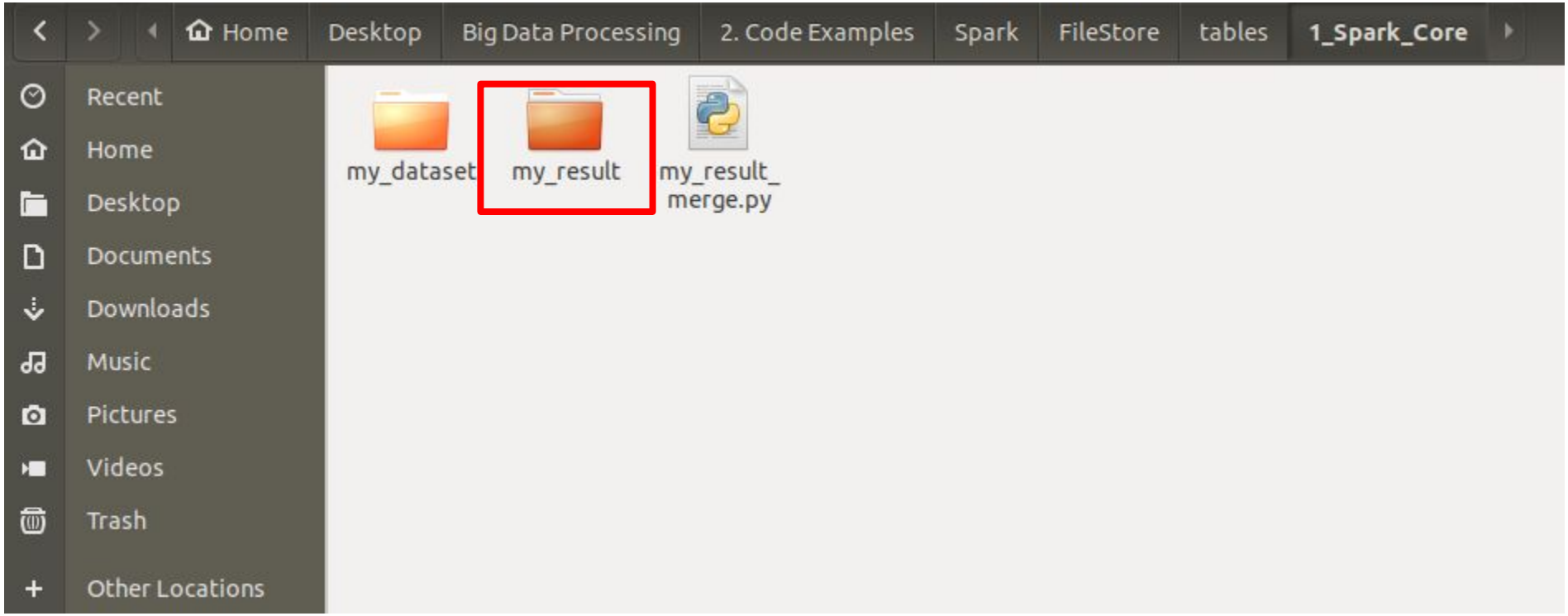


Databricks: An Online Platform for Data Engin

9. But we can use the DBFS command cp to copy them to our local machine!

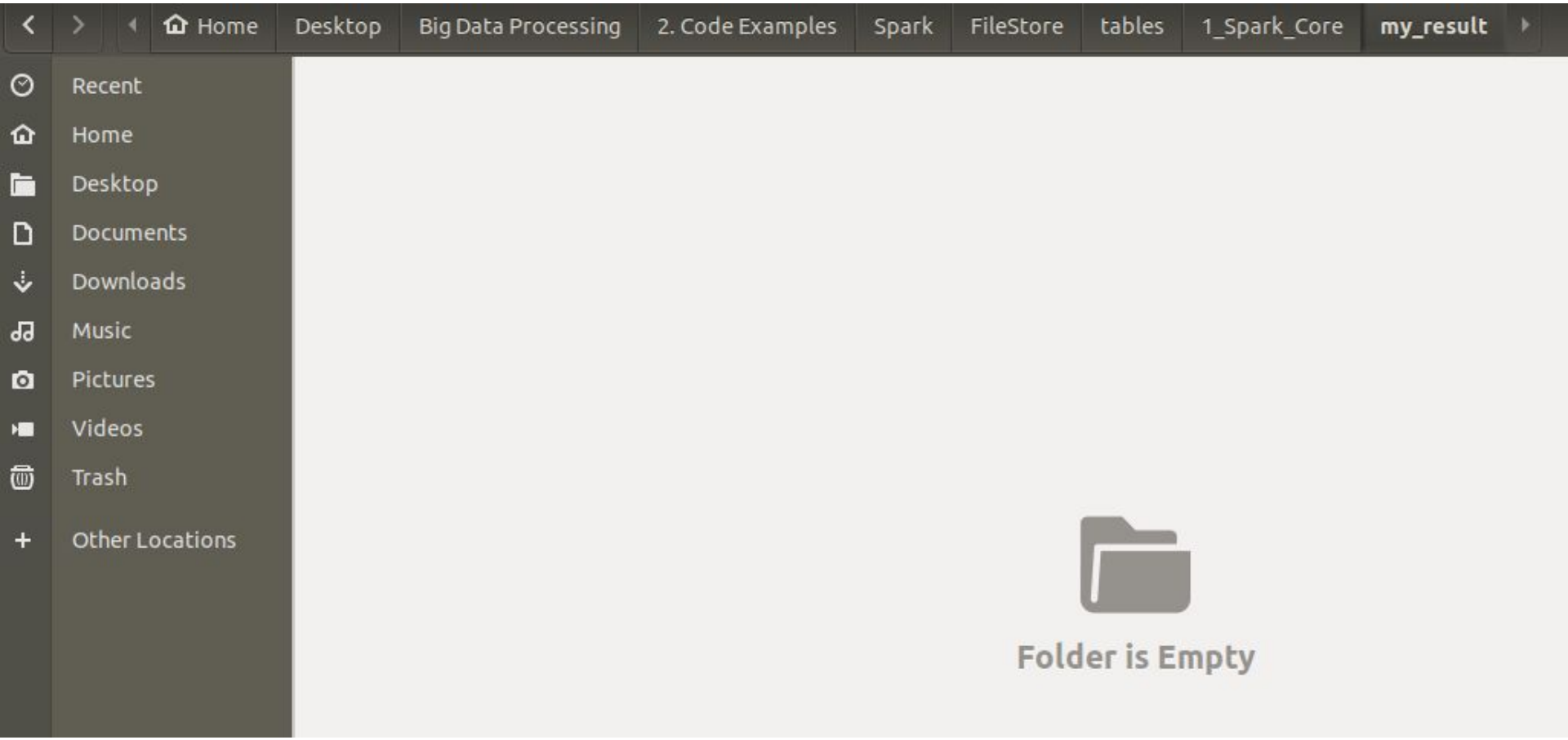
Databricks: An Online Platform for Data Engin

9. As we can see, before running the command **cp** the local folder **my_result** is empty.



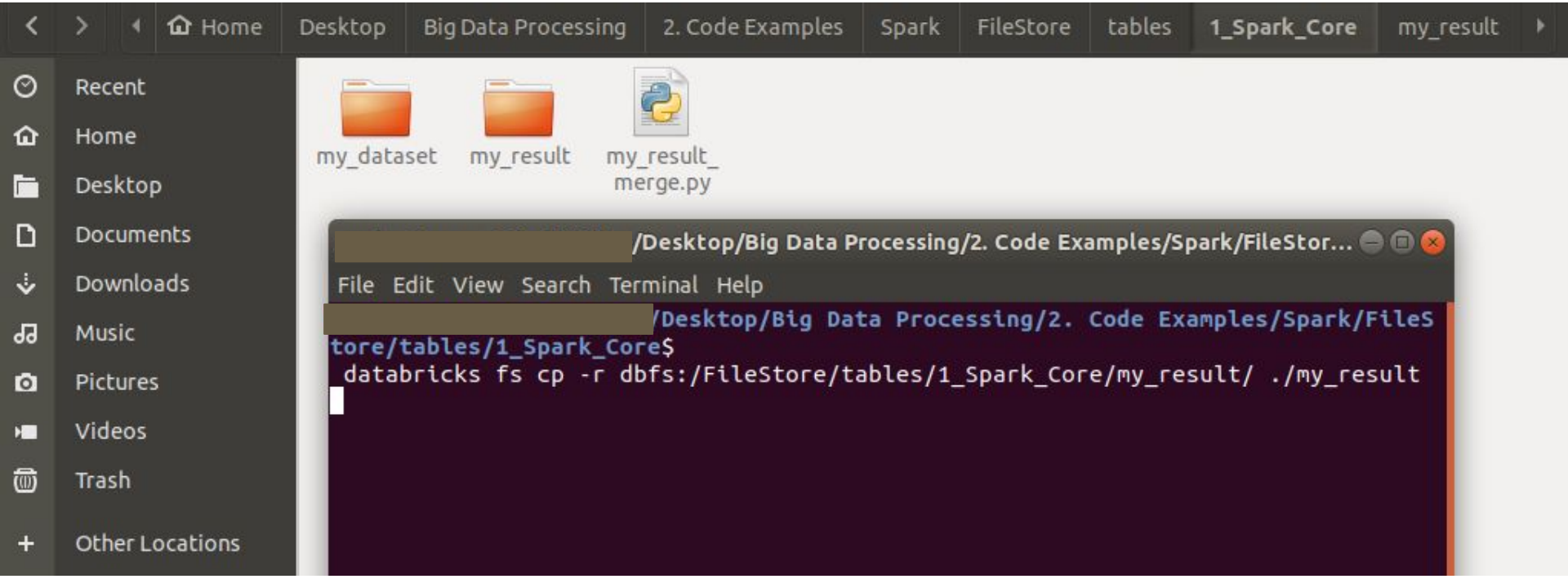
Databricks: An Online Platform for Data Engin

9. As we can see, before running the command **cp** the local folder **my_result** is empty.



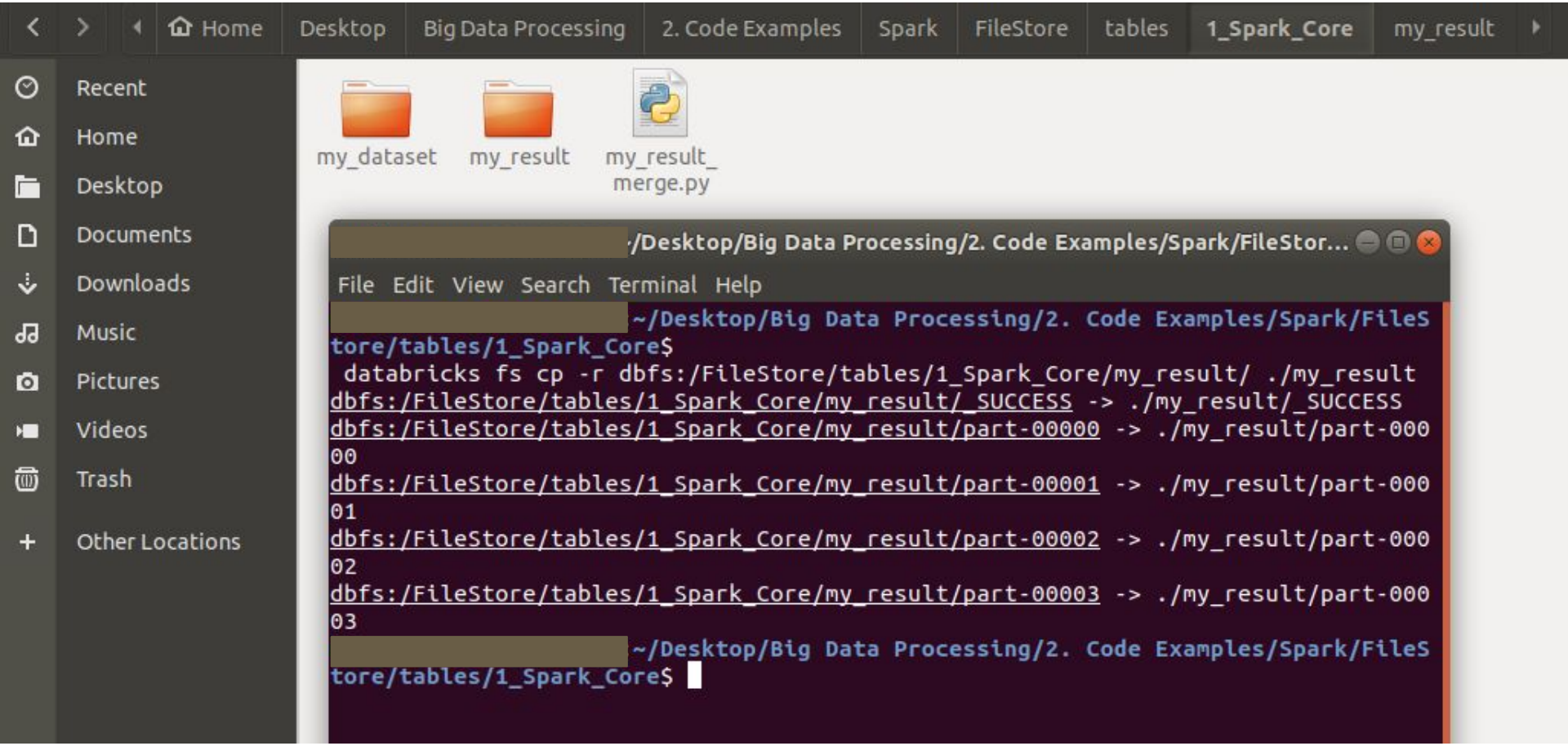
Databricks: An Online Platform for Data Engin

9. We run the command **cp** to copy the files.



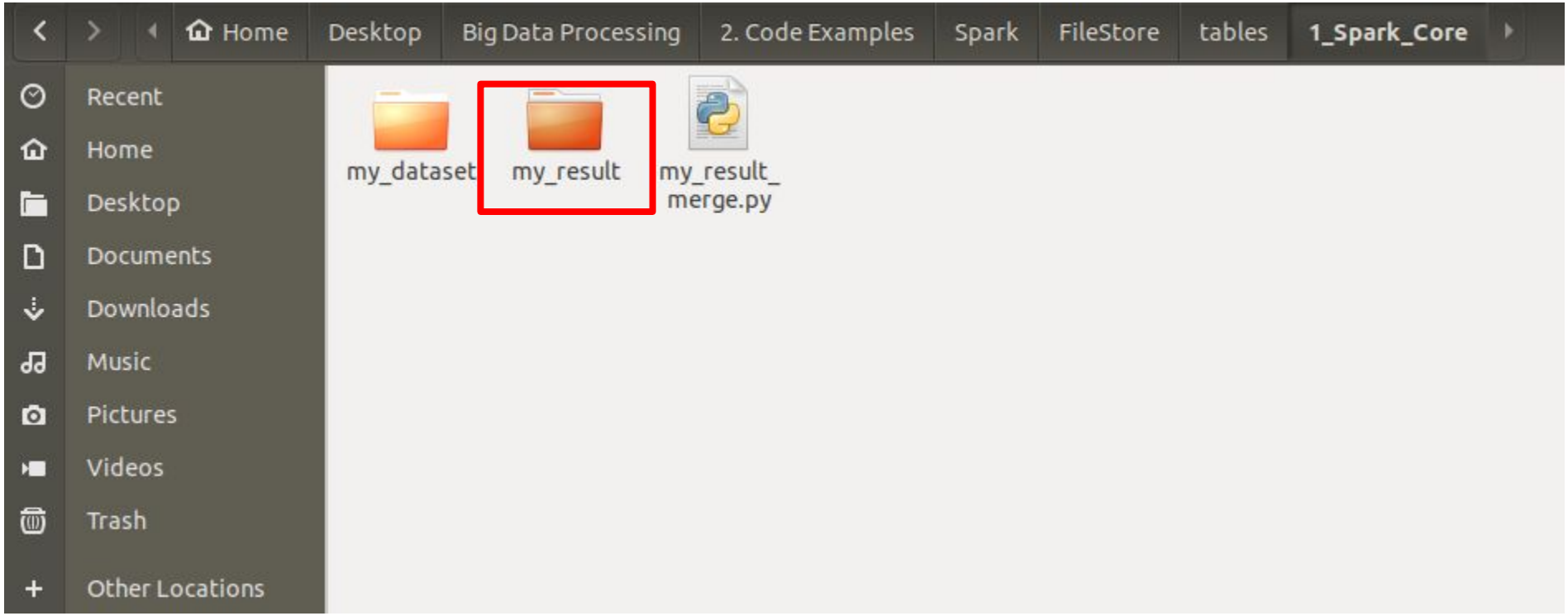
Databricks: An Online Platform for Data Engin

9. We run the command **cp** to copy the files.



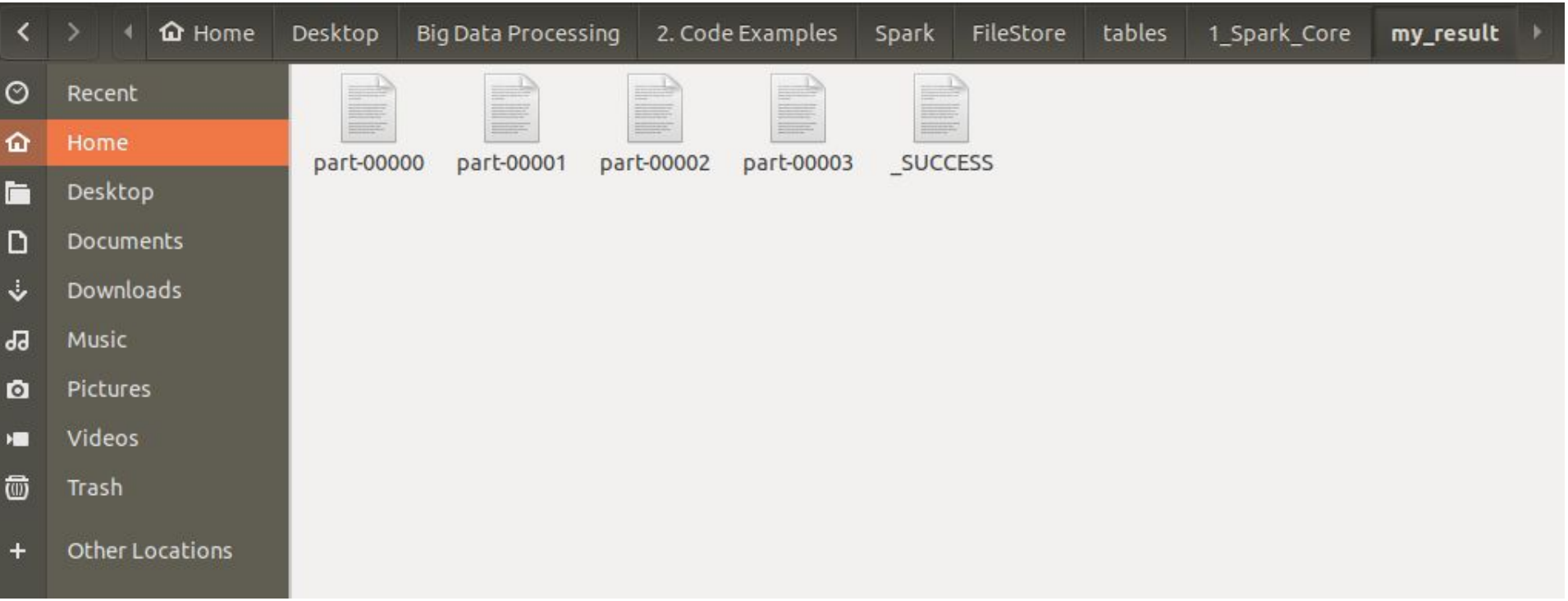
Databricks: An Online Platform for Data Engin

9. We run the command **cp** to copy the files.



Databricks: An Online Platform for Data Engin

9. And now that the solution files of running the Spark Application are back to our computer we can open them to see their content.



Databricks: An Online Platform for Data Engin

9. And now that the solution files of running the Spark Application are back to our computer we can open them to see their content.

Databricks: An Online Platform for Data Engineers

Databricks provides much more functionality, and I encourage you to explore it.

That being said, the one covered in these slides represents all the functionality we need to use for this module.

Thank you for your attention!