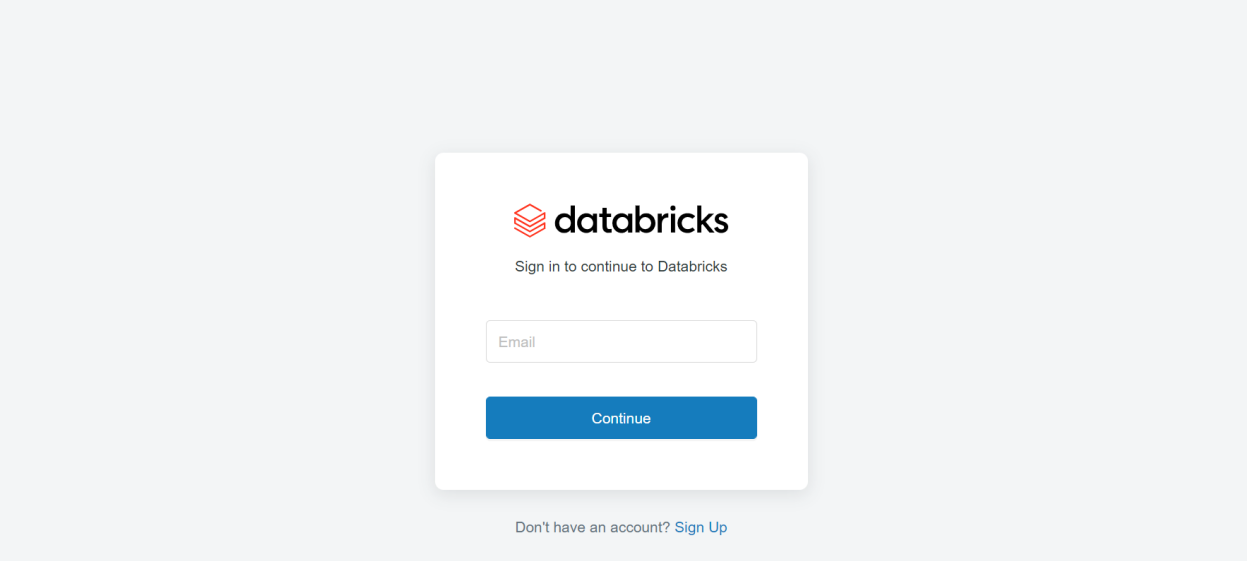
Data file mounting On Databricks

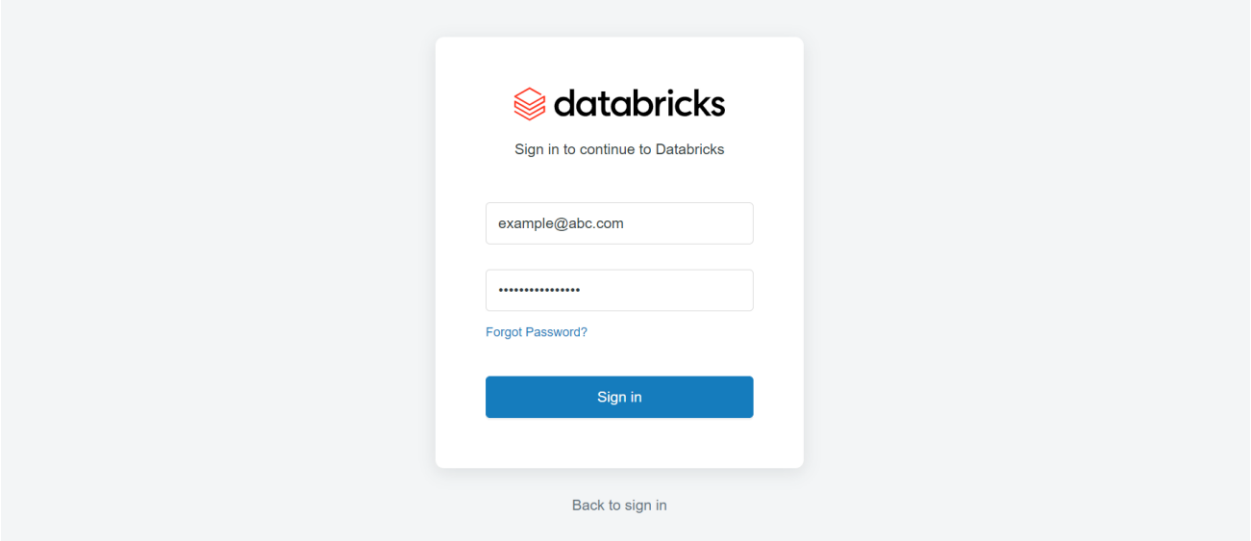
databricks™

How-To Guide

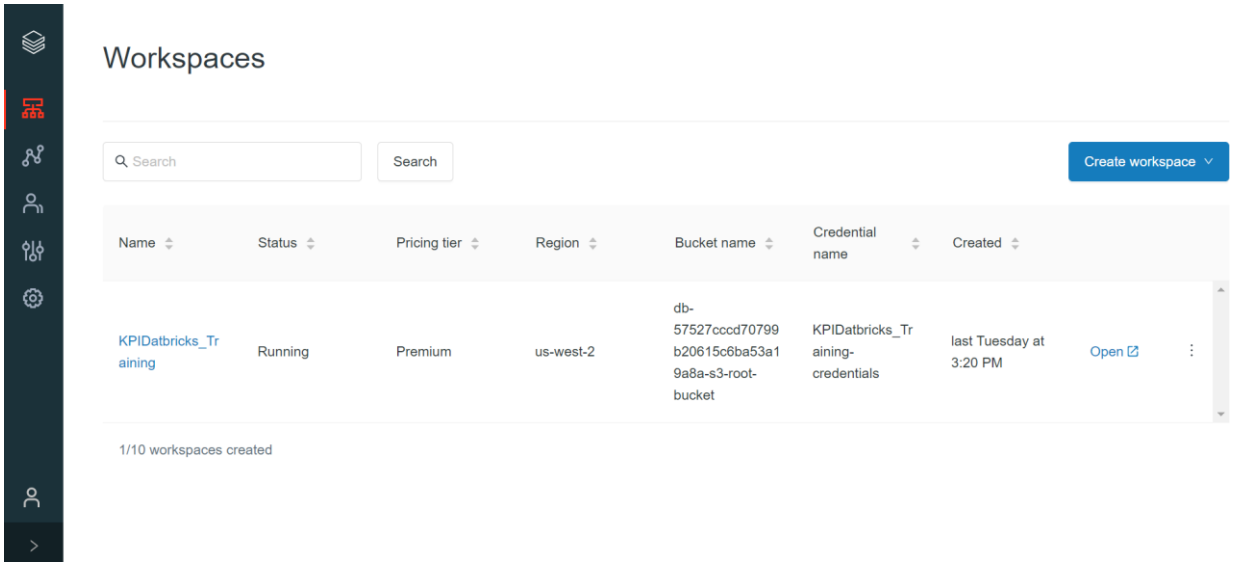
Step 1: log in to databricks console.



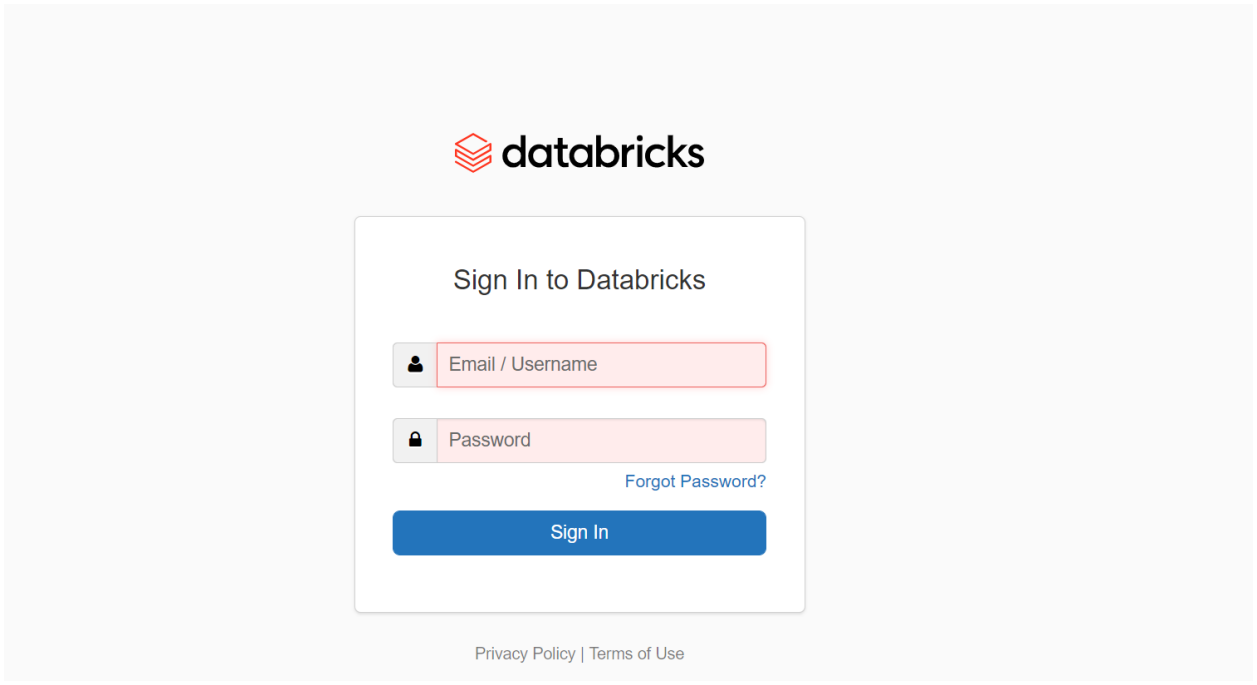
Enter the credentials.



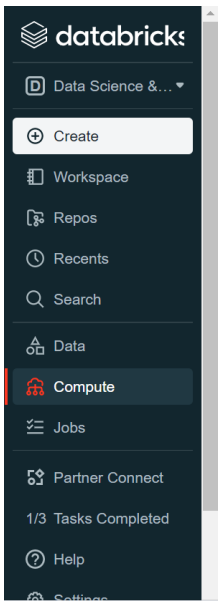
Step 2: As cluster is already created, click on open hyper link.



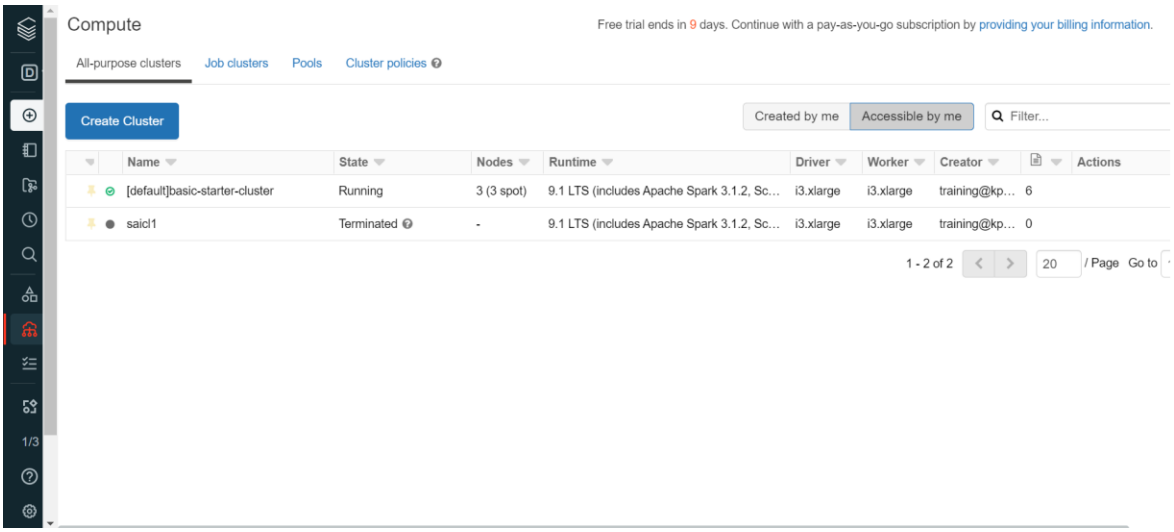
Again enter the credentials.



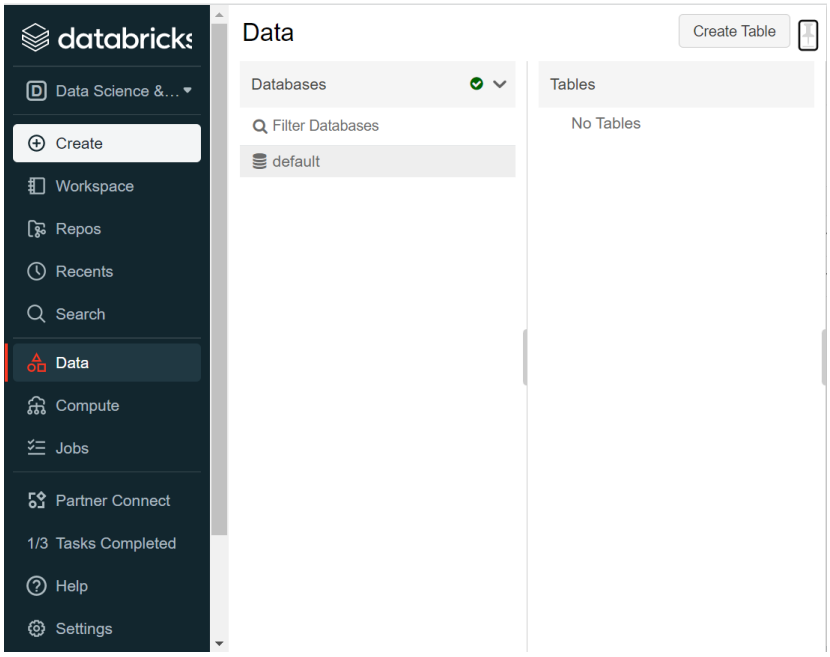
Step 3: Click on compute.



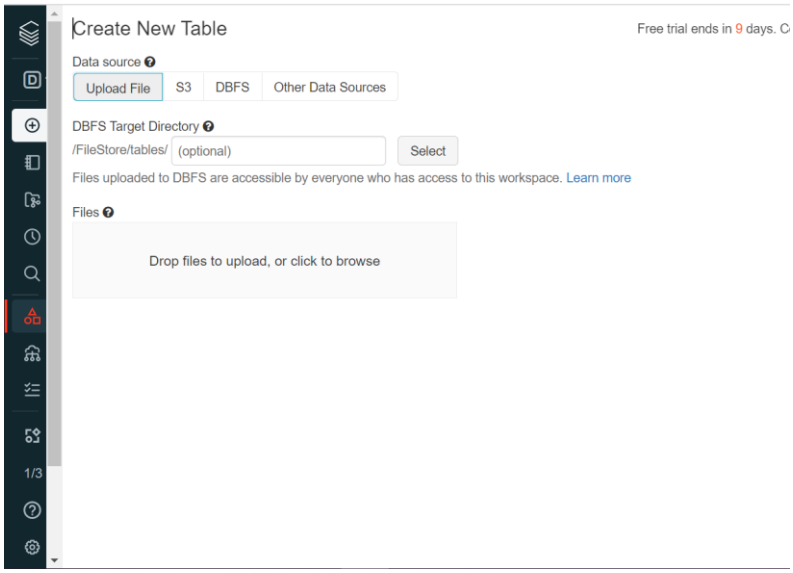
Step 4: Run the cluster, check out for the confirmation with the green tick mark.



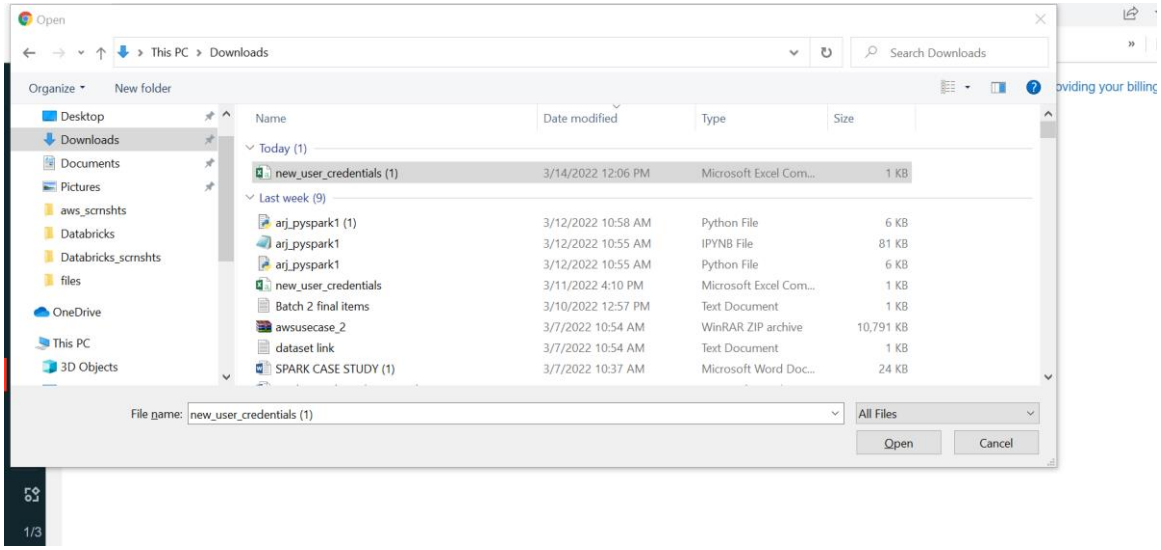
Step 5: Click on data, then on default after that click on create table.



Step 6: Click on drop files to upload tab.

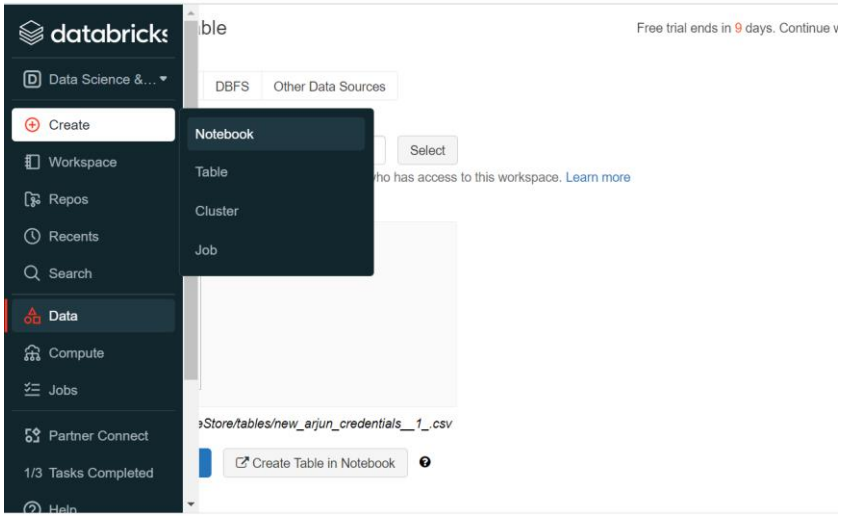


Step 7: Select the credential file, downloaded from AWS console.

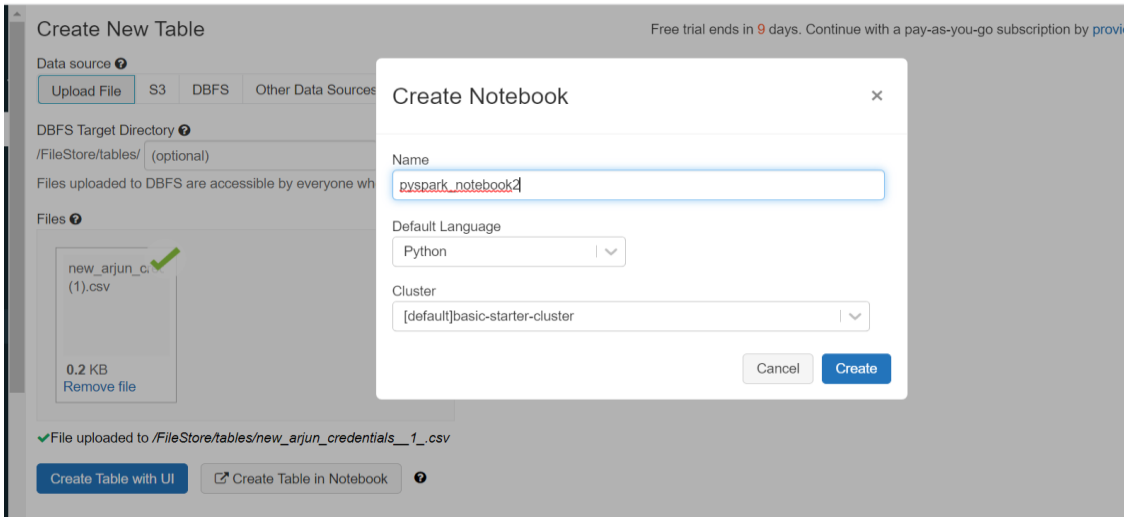


Step 8: After uploading leave it as it ease.

Step 9: Create notebook as shown below.



Give a notebook name and click on create button.



Run the scripts as below.

pyspark_notebook2 Python

Free trial ends in 9 days. Continue with a pay-as-you-go subscription by providing

[default]basic-start...

Cmd 1

```
1  display(dbutils.fs.ls("/FileStore/tables/"))
```

(2) Spark Jobs

Table | Data Profile

	path	name	size
1	dbfs:/FileStore/tables/Covid_Filtered_Data.csv	Covid_Filtered_Data.csv	44987720
2	dbfs:/FileStore/tables/Crops_price-1.csv	Crops_price-1.csv	1813
3	dbfs:/FileStore/tables/Crops_price-2.csv	Crops_price-2.csv	1813
4	dbfs:/FileStore/tables/Crops_price.csv	Crops_price.csv	1835
5	dbfs:/FileStore/tables/bamboo_plantation.txt	bamboo_plantation.txt	666
6	dbfs:/FileStore/tables/bamboo_plantation_15_16.txt	bamboo_plantation_15_16.txt	666
7	dbfs:/FileStore/tables/new_arjun_credentials__1_.csv	new_arjun_credentials__1_.csv	209

pyspark_notebook2 Python

Free trial ends in 9 days. Continue with a pay-as-you-go subscription by providing y

[default]basic-start...

Cmd 3

```
1  file_type = "csv"
2  first_row_is_header = "true"
3  delimiter = ","
4  aws_key_df = spark.read.format(file_type).option("header", first_row_is_header).option("separator",
   delimiter).load("dbfs:/FileStore/tables/new_arjun_credentials__1_.csv")
```

(1) Spark Jobs

aws_key_df: pyspark.sql.dataframe.DataFrame = [User name: string, Password: string ... 3 more fields]

Command took 0.93 seconds -- by training@kpipartners.com at 3/14/2022, 12:38:45 PM on [default]basic-starter-cluster

Cmd 4

```
1  aws_key_df.show()
```

(1) Spark Jobs

```
+--------------------+--------+--------------------+--------------------+--------------------+
|           User name|Password|       Access key ID|   Secret access key|  Console login link|
+--------------------+--------+--------------------+--------------------+--------------------+
|arjun_databricks_...|    null|AKIA54P67JZBUMKGT3VG|TmPL6n84A+HIwaVBN...|https://954554338...|
+--------------------+--------+--------------------+--------------------+--------------------+
```

Command took 0.18 seconds -- by training@kpipartners.com at 3/14/2022, 12:39:22 PM on [default]basic-starter-cluster

```
1  access_key = aws_key_df.select('Access key ID')
2  secret_key = aws_key_df.select('Secret access key')
3  access_key.show()
4  secret_key.show()
```

▸ (2) Spark Jobs

▸ access_key: pyspark.sql.dataframe.DataFrame = [Access key ID: string]

▸ secret_key: pyspark.sql.dataframe.DataFrame = [Secret access key: string]

```
+--------------------+
|       Access key ID|
+--------------------+
|AKIA54P67JZBUMKGT3VG|
+--------------------+

+--------------------+
|   Secret access key|
+--------------------+
|TmPL6n84A+HIwaVBN...|
+--------------------+
```

Command took 0.23 seconds -- by training@kpipartners.com at 3/14/2022, 12:49:00 PM on

pyspark_notebook2 Python

Free trial ends in 8 days. Continue with a pay-as-you-go subscription by providing y

[default]basic-start...

```
1  access_key = aws_key_df.select('Access key ID').collect()[0]['Access key ID']
```

▸ (1) Spark Jobs

Command took 0.12 seconds -- by training@kpipartners.com at 3/15/2022, 11:22:14 AM on [default]basic-starter-cluster

Cmd 9

```
1  secret_key = aws_key_df.select('Secret access key').collect()[0]['Secret access key']
```

▸ (1) Spark Jobs

Command took 0.17 seconds -- by training@kpipartners.com at 3/15/2022, 11:22:35 AM on [default]basic-starter-cluster

Cmd 10

```
1  encoded_secret_key = urllib.parse.quote(string=secret_key, safe="")
```

Command took 0.01 seconds -- by training@kpipartners.com at 3/15/2022, 11:22:48 AM on [default]basic-starter-cluster

Cmd 11

```
1  url1 = "s3n://{0}:{1}@{2}".format(access_key, encoded_secret_key, "mount_file")
```

pyspark_notebook2 Python Free trial ends in 8 days. Continue with a pay-as-you-go subscription by providi

[default]basic-start...

```
1 encoded_secret_key = urllib.parse.quote(string=secret_key, safe="")
```

Command took 0.01 seconds -- by training@kpipartners.com at 3/15/2022, 11:22:48 AM on [default]basic-starter-cluster

Cmd 11

```
1 url1 = "s3n://{0}:{1}@{2}".format(access_key, encoded_secret_key, "mount_file")
```

Command took 0.02 seconds -- by training@kpipartners.com at 3/15/2022, 11:23:46 AM on [default]basic-starter-cluster

Cmd 12

```
1 dbutils.fs.mount(url1, "/mnt/marksFolder")
```

Cmd 11

```
1 dbutils.fs.mount(url1, "/mnt/marksFolder")
```

Command took 0.02 seconds -- by training@kpipartners.com at 3/15/2022, 11:23:46 AM on [default]basic-starter-cluster

Cmd 12

```
1 display(dbutils.fs.ls("/mnt/marksFolder"))
```

Shift+Enter to run

```
display(dbutils.fs.ls("/mnt/"))
```

path	name	size
dbfs:/mnt/marksFolder/	marksFolder/	0
dbfs:/mnt/mnt/	mnt/	0
dbfs:/mnt/mount1/	mount1/	0
dbfs:/mnt/mymount/	mymount/	0
dbfs:/mnt/mymount2/	mymount2/	0
dbfs:/mnt/mys3bucket/	mys3bucket/	0
dbfs:/mnt/newmarksFolder/	newmarksFolder/	0

```
%fs mounts
```

mountPoint	source	encryptionType
/mnt//mnt/mount1	s3a://training-labtarget1	
/databricks-datasets	databricks-datasets	sse-s3
/mnt/mount1	s3a://training-labtarget1	
/mnt/marksFolder	s3n://marksFolder	
/databricks/mlflow-tracking	databricks/mlflow-tracking	sse-s3