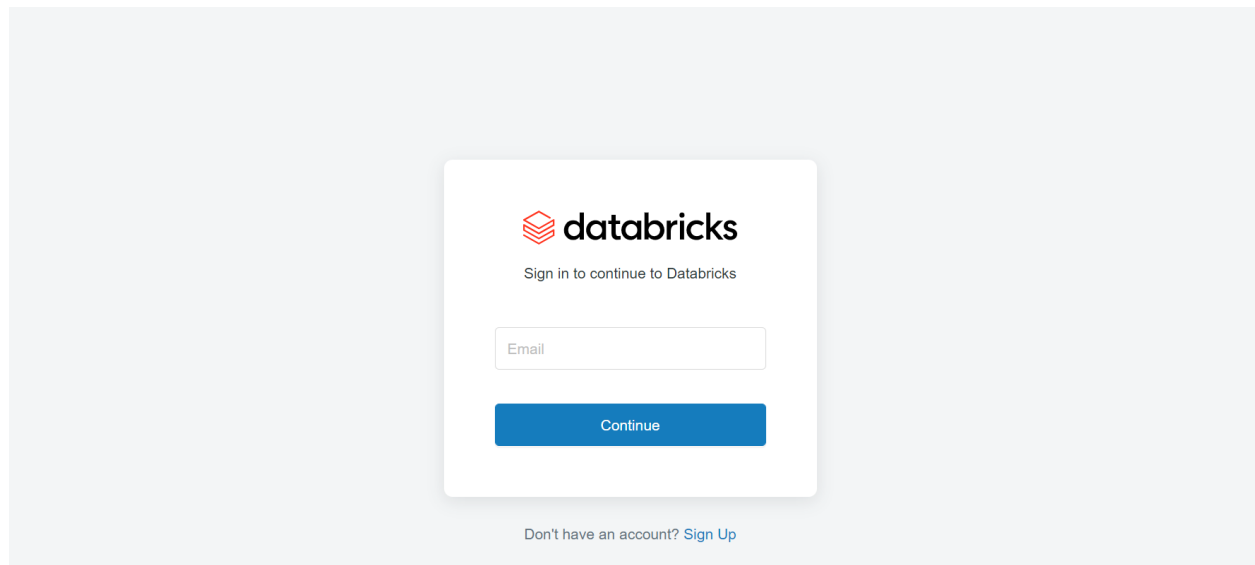


Data file mounting On Databricks

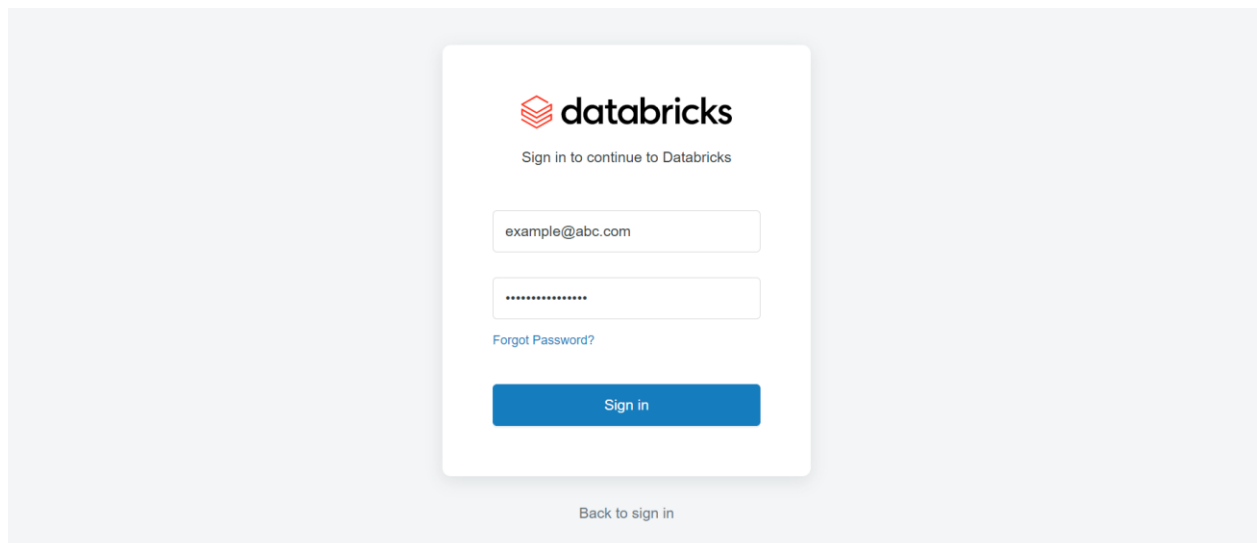
How-To Guide

Step 1: log in to databricks console.



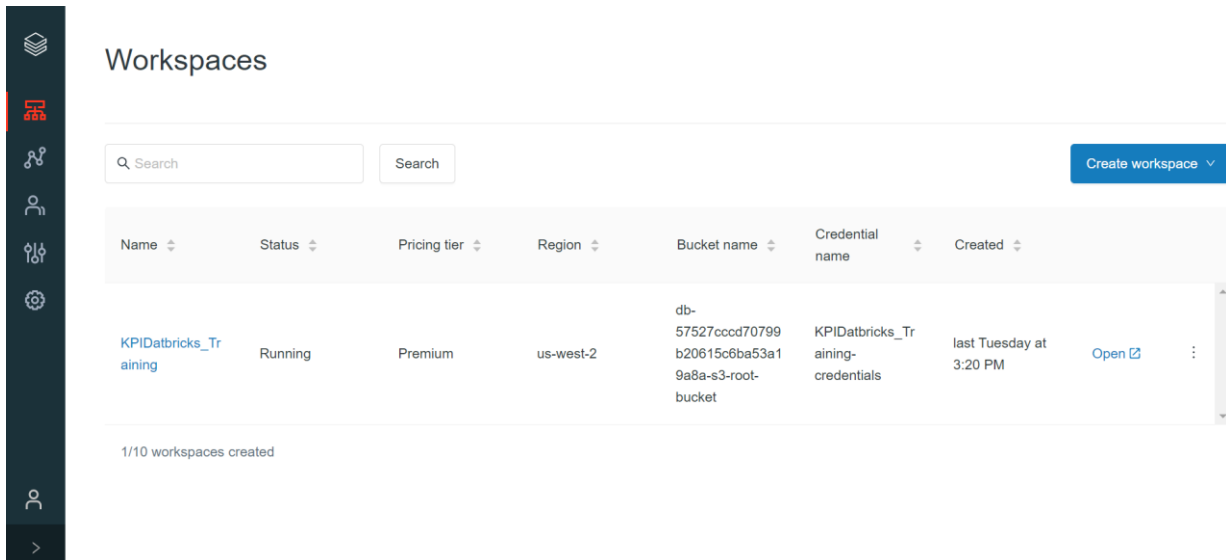
The image shows the Databricks login interface. At the top is the Databricks logo, consisting of a red cube icon and the word "databricks". Below the logo is the text "Sign in to continue to Databricks". There is a text input field labeled "Email". Below the input field is a blue button labeled "Continue". At the bottom of the form, there is a link that says "Don't have an account? [Sign Up](#)".

Enter the credentials.



The image shows the Databricks login interface with credentials entered. At the top is the Databricks logo, consisting of a red cube icon and the word "databricks". Below the logo is the text "Sign in to continue to Databricks". There is a text input field containing "example@abc.com". Below that is a password input field with masked characters "*****". Below the password field is a link that says "Forgot Password?". At the bottom of the form is a blue button labeled "Sign in". Below the button, there is a link that says "Back to sign in".

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

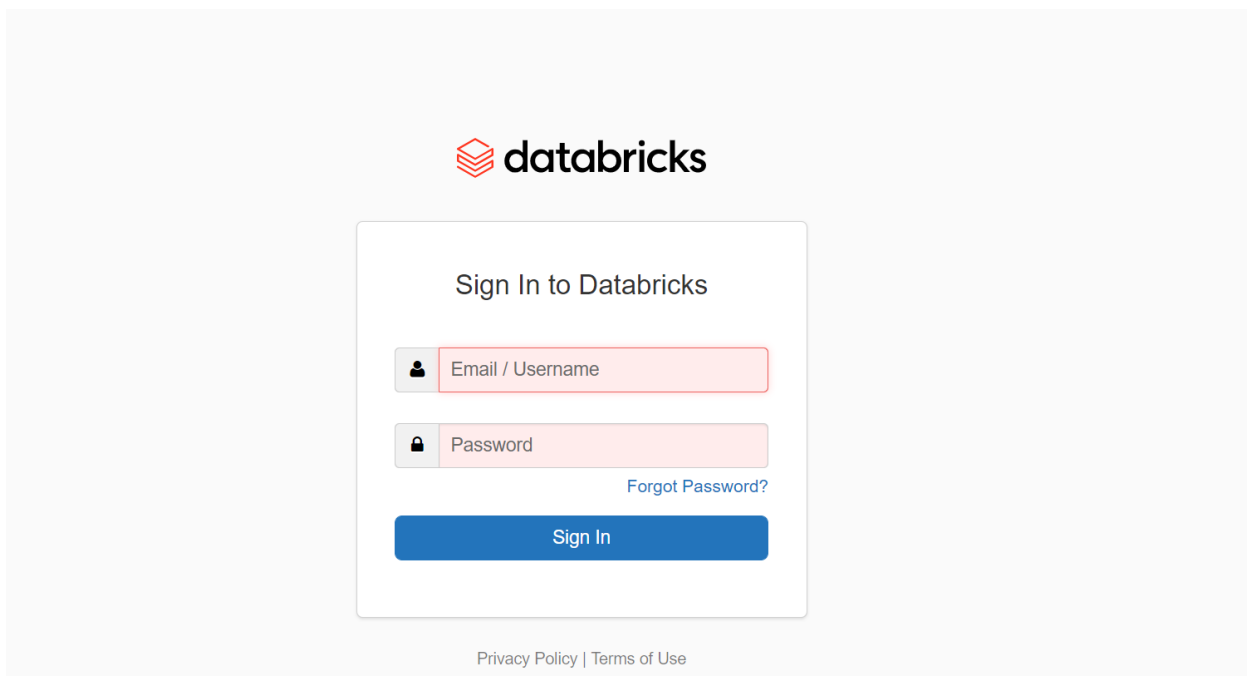


The screenshot shows the Databricks Workspaces management page. On the left is a dark sidebar with navigation icons. The main area is titled 'Workspaces' and contains a search bar, a 'Create workspace' button, and a table of workspace details.

Name	Status	Pricing tier	Region	Bucket name	Credential name	Created	
KPIDatbricks_Training	Running	Premium	us-west-2	db-57527cccd70799b20615c6ba53a19a8a-s3-root-bucket	KPIDatbricks_Training-credentials	last Tuesday at 3:20 PM	Open

1/10 workspaces created

Again enter the credentials.



The screenshot shows the Databricks Sign In page. It features the Databricks logo at the top, followed by the heading 'Sign In to Databricks'. Below this are two input fields: 'Email / Username' and 'Password'. A 'Forgot Password?' link is located below the password field. A blue 'Sign In' button is at the bottom of the form. At the very bottom of the page, there are links for 'Privacy Policy' and 'Terms of Use'.

databricks

Sign In to Databricks

Email / Username

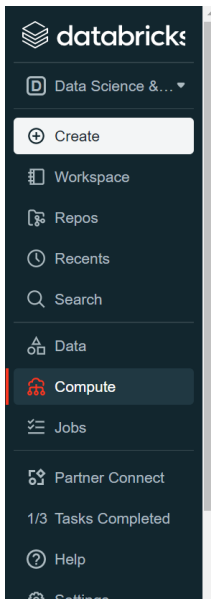
Password

[Forgot Password?](#)

Sign In

[Privacy Policy](#) | [Terms of Use](#)

Step 3: Click on compute.



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

Compute

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

All-purpose clusters Job clusters Pools Cluster policies

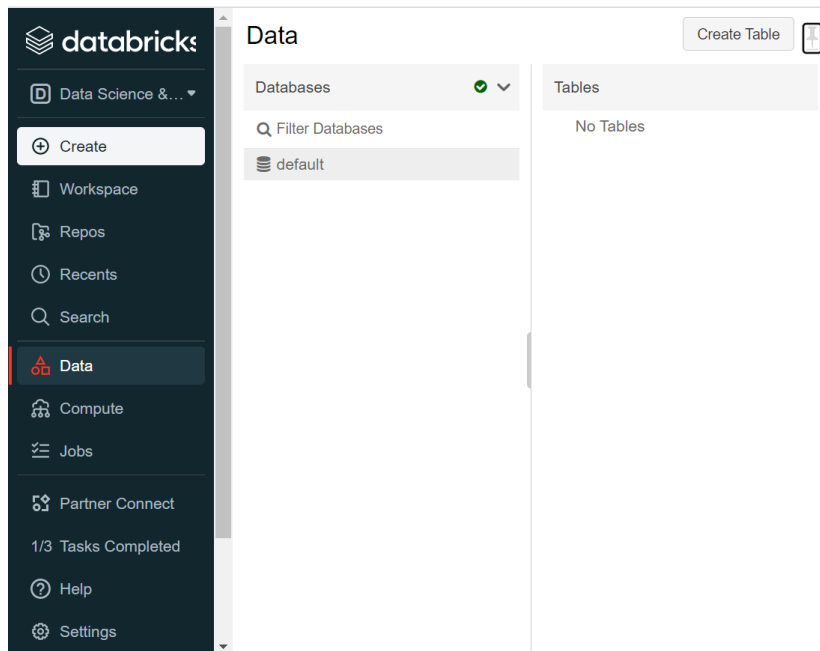
Create Cluster

Created by me Accessible by me Filter...

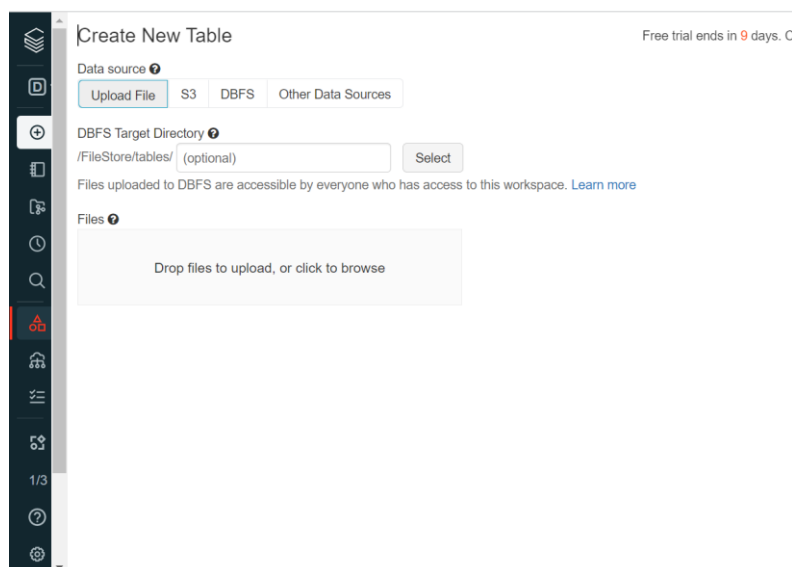
	Name	State	Nodes	Runtime	Driver	Worker	Creator		Actions
	[default]basic-starter-cluster	Running	3 (3 spot)	9.1 LTS (includes Apache Spark 3.1.2, Sc...	i3.xlarge	i3.xlarge	training@kp...	6	
	saicl1	Terminated	-	9.1 LTS (includes Apache Spark 3.1.2, Sc...	i3.xlarge	i3.xlarge	training@kp...	0	

1 - 2 of 2 < > 20 / Page Go to

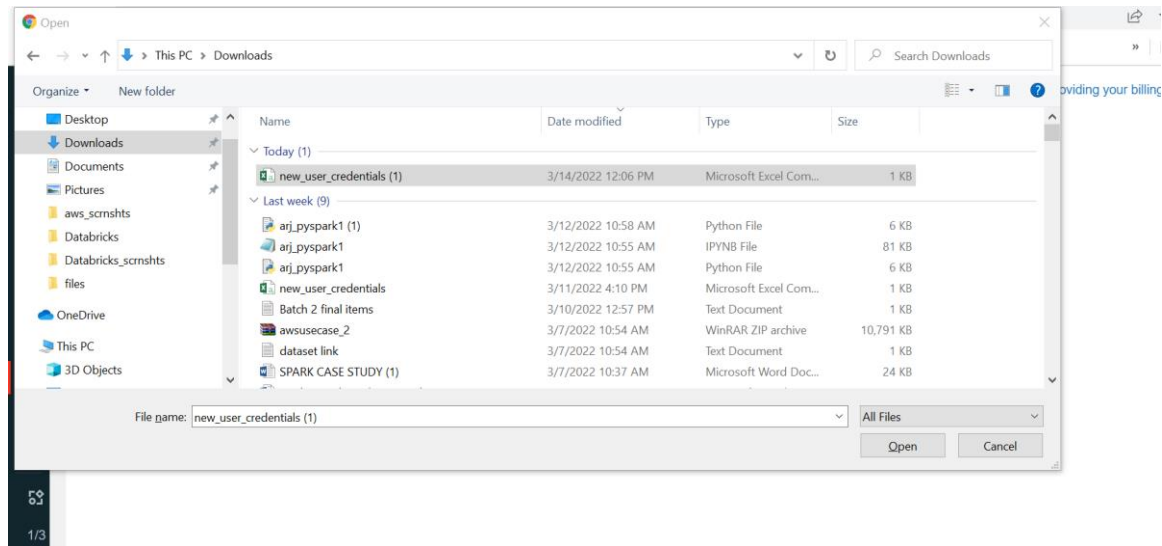
Step 5: Click on data, then on default after that click on create table.



Step 6: Click on drop files to upload tab.



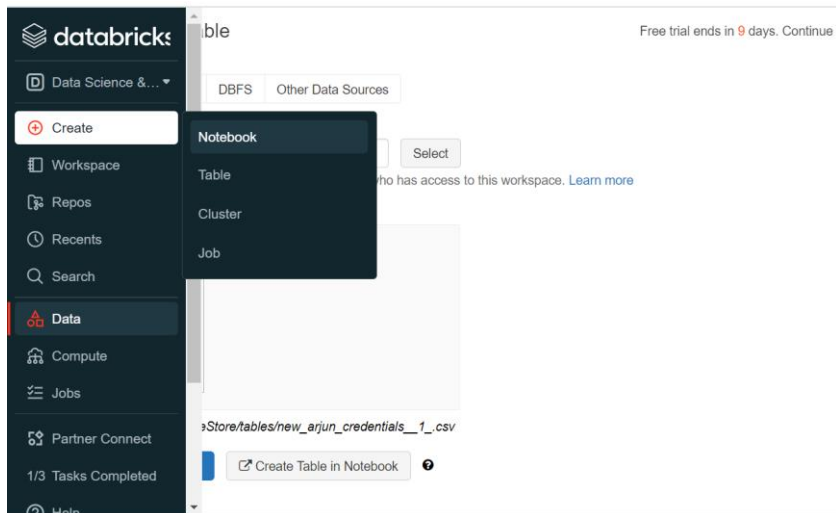
Step 7: Select the credential file, downloaded from AWS console.



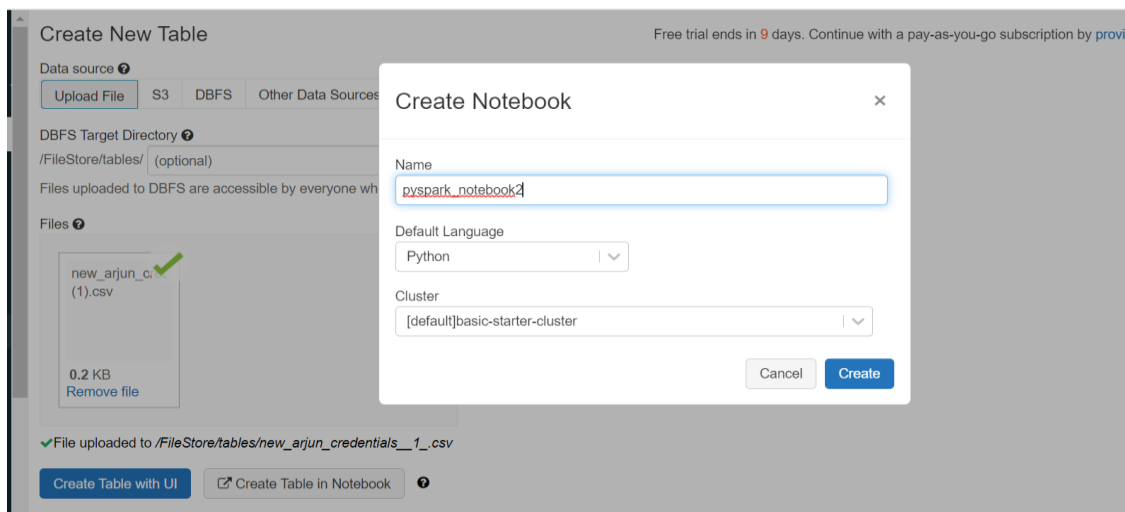
Step 8: After uploading leave it as it ease.



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 y](#)

[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, 12:38:45 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](#)

  [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