



Untitled Notebook 2024-02-16 16:49:30 Python ☆

File Edit View Run Help

Last edit was 5 minutes ago

New cell UI: OFF

Run all

example clustetr

Share

Publish



Cmd 1

SQL



```
1 %sql
2 CREATE TABLE delta.`/tmp/delta-table` USING DELTA AS SELECT col1 as id FROM VALUES 0,1,2,3,4;
```

▶ (6) Spark Jobs

```
▼ _sqldf: pyspark.sql.dataframe.DataFrame
  num_affected_rows: long
  num_inserted_rows: long
```

Query returned no results

SQL cell result stored as PySpark data frame `_sqldf`. [Learn more](#)

Command took 19.07 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:53:45 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text



Untitled Notebook 2024-02-16 16:49:30

Python



File Edit View Run Help

Last edit was now

New cell UI: OFF

Run all

example clustetr

Share

Publish



Query returned no results

Command took 19.07 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:53:45 PM on example clustetr



Cmd 2

```
1 %sql
2 SELECT * FROM delta.`/tmp/delta-table`;
```

▶ (2) Spark Jobs

▶ _sqldf: pyspark.sql.dataframe.DataFrame = [id: integer]

Table +

	id
1	0
2	1
3	2
4	3
5	4

5 rows | 11.23 seconds runtime

Refreshed now

SQL cell result stored as PySpark data frame _sqldf. [Learn more](#)

Command took 11.23 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:56:39 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text

Untitled Notebook 2024-02-16 16:49:30 Python ☆

File Edit View Run Help

Last edit was 1 minute ago

New cell UI: OFF

Run all

example clustetr

Share

Publish

Command took 11.23 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:56:39 PM on example clustetr

Cmd 3

```
1 %sql
2 INSERT OVERWRITE delta.`/tmp/delta-table` SELECT col1 as id FROM VALUES 5,6,7,8,9;
```

▶ (6) Spark Jobs

▶ _sqldf: pyspark.sql.dataframe.DataFrame = [num_affected_rows: long, num_inserted_rows: long]

Table +

	num_affected_rows	num_inserted_rows
1	5	5

1 row | 5.72 seconds runtime

Refreshed now

SQL cell result stored as PySpark data frame _sqldf. [Learn more](#)

Command took 5.72 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:57:38 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text



Untitled Notebook 2024-02-16 16:49:30

Python



File Edit View Run Help

Last edit was 2 minutes ago

New cell UI: OFF

Run all

example clustetr

Share

Publish



1 row | 5.72 seconds runtime

Refreshed now

Command took 5.72 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:57:38 PM on example clustetr

Cmd 4

```
1 %sql
2 UPDATE delta.`/tmp/delta-table` SET id = id + 100 WHERE id % 2 == 0;
```

▶ (9) Spark Jobs

▶ _sqldf: pyspark.sql.dataframe.DataFrame = [num_affected_rows: long]

Table +

	num_affected_rows
1	2

1 row | 9.28 seconds runtime

Refreshed now

SQL cell result stored as PySpark data frame _sqldf. Learn more

Command took 9.28 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:58:31 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text

Untitled Notebook 2024-02-16 16:49:30 Python ☆

File Edit View Run Help

Last edit was 2 minutes ago

New cell UI: OFF

Run all

example clustetr

Share

Publish

1 row | 9.28 seconds runtime

Refreshed now

Command took 9.28 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:58:31 PM on example clustetr

Cmd 5

```
1 %sql
2 -- Delete very even value
3 DELETE FROM delta.`/tmp/delta-table` WHERE id % 2 == 0
```

▶ (9) Spark Jobs

▶ _sqldf: pyspark.sql.dataframe.DataFrame = [num_affected_rows: long]

Table +

	num_affected_rows ▲
1	2

1 row | 6.33 seconds runtime

Refreshed now

SQL cell result stored as PySpark data frame `_sqldf`. [Learn more](#)

Command took 6.33 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:59:08 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text



Untitled Notebook 2024-02-16 16:49:30 Python ☆

File Edit View Run Help

Last edit was 1 minute ago

New cell UI: OFF

Run all

example clustetr

Share

Publish



1 2

1 row | 6.33 seconds runtime

Refreshed 3 minutes ago

SQL cell result stored as PySpark data frame `_sqldf`. [Learn more](#)

Command took 6.33 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 4:59:00 PM on example clustetr

Cmd 6

Python

```
1 data = spark.range(0, 5)
2 data.write.format("delta").save("/tmp/deltain-table")
```

▶ (6) Spark Jobs

▶ data: pyspark.sql.dataframe.DataFrame = [id: long]

Command took 6.93 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 5:02:29 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text

Command took 6.93 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 5:02:29 PM on example clustetr

Cmd 7

```
1 df = spark.read.format("delta").load("/tmp/delta-table")
2 df.show()
```

- ▶ (2) Spark Jobs

```
df: pyspark.sql.dataframe.DataFrame = [id: integer]
```

```
+---+
| id|
+---+
|  5|
|  7|
|  9|
+---+
```

Command took 1.70 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 5:03:13 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text

Untitled Notebook 2024-02-16 16:49:30 Python ☆

File Edit View Run Help

Last edit was 2 minutes ago

New cell UI: OFF

Run all

example clustetr

Share

Publish

Command took 6.93 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 5:02:29 PM on example clustetr

Cmd 7

```
1 df = spark.read.format("delta").load("/tmp/deltain-table")
2 df.show()
```

▶ (3) Spark Jobs

▶ df: pyspark.sql.dataframe.DataFrame = [id: long]

```
+---+
| id|
+---+
|  0|
|  1|
|  2|
|  3|
|  4|
+---+
```

Command took 1.86 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 5:03:44 PM on example clustetr

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text


```
1 data = spark.range(5, 10)
2 data.write.format("delta").mode("overwrite").save("/tmp/deltain-table")
```

- ▶ (6) Spark Jobs

```
data: pyspark.sql.dataframe.DataFrame = [id: long]
```

Command took 5.16 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 5:04:22 PM on example clustetr

Cmd 9

```
1 from delta.tables import *
2 from pyspark.sql.functions import *
3
4 deltaTable = DeltaTable.forPath(spark, "/tmp/delta-table")
```

Command took 0.35 seconds -- by akashpeesapati@gmail.com at 2/16/2024, 5:05:40 PM on example clustetr

Cmd 10

1

[Shift+Enter] to run

[Shift+Ctrl+Enter] to run selected text



Untitled Notebook 2024-02-16 16:49:30 Python ☆

File Edit View Run Help

Last edit was 1 minute ago

New cell UI: OFF

Run all

example clustetr

Share

Publish



```
1 # Update every even value by adding 100 to it
2 deltaTable.update(
3   condition = expr("id % 2 == 0"),
4   set = { "id": expr("id + 100") })
5
6 # Delete every even value
7 deltaTable.delete(condition = expr("id % 2 == 0"))
8
9 # Upsert (merge) new data
10 newData = spark.range(0, 20)
11
12 deltaTable.alias("oldData") \
13   .merge(
14     newData.alias("newData"),
15     "oldData.id = newData.id" ) \
16   .whenMatchedUpdate(set = { "id": col("newData.id") }) \
17   .whenNotMatchedInsert(values = { "id": col("newData.id") }) \
18   .execute()
19
20 deltaTable.toDF().show()
```

▶ (25) Spark Jobs

▶ `newData: pyspark.sql.dataframe.DataFrame = [id: long]`

```
3|
4|
7|
8|
9|
17|
18|
19|
12|
13|
14|
5|
6|
15|
16|
0|
1|
10|
11|
++
```