



(<https://databricks.com>)



## Delta Lake

### Module 4 Assignment

This final assignment is broken up into 2 parts:

1. Completing this Delta Lake notebook
  - Submitting question answers to Coursera
  - Uploading notebook to Coursera for peer reviewing
2. Answering 3 free response questions on Coursera platform

#### ★ In this notebook you:

- Create a Delta table with partitioned columns
- Delete records
- Update existing records
- Time travel!

For each **bold** question, input its answer in Coursera.

```
%run ../Includes/Classroom-Setup
```

Data mounted to /mnt/davis ...

OK

Create a new Delta table called `fireCallsPartitioned` partitioned by `Priority`.

```
CREATE OR REPLACE TEMPORARY VIEW fireCallsParquet
USING parquet
OPTIONS (
  path "/mnt/davis/fire-calls/fire-calls-clean.parquet/"
)
```

OK

```
-- TODO
CREATE DATABASE IF NOT EXISTS Databricks;
USE Databricks;
DROP TABLE IF EXISTS fireCallsPartitioned;
```

```
CREATE TABLE fireCallsPartitioned
USING DELTA
PARTITIONED BY (Priority)
AS
  SELECT * FROM fireCallsParquet
```

Query returned no results

### Question 1

How many folders were created? Enter the number of records you see from the output below (include the `_delta_log` in your count).

```
%fs ls dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned
```

Table					
	path	name	size	modificationTime	
1	dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned/Priority=1/	Priority=1/	0	0	
2	dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned/Priority=2/	Priority=2/	0	0	
3	dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned/Priority=3/	Priority=3/	0	0	
4	dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned/Priority=A/	Priority=A/	0	0	
5	dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned/Priority=B/	Priority=B/	0	0	
6	dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned/Priority=C/	Priority=C/	0	0	
7	dbfs:/user/hive/warehouse/databricks.db/firecallspartitioned/Priority=E/	Priority=E/	0	0	
9 rows					

Question 2

Delete all the records where city is null. How many records are left in the delta table?

```
-- TODO

DELETE FROM firecallspartitioned WHERE City IS NULL
```

Table		
	num_affected_rows	
1	550	
1 row		

Question 3

After you deleted all records where the City is null, how many files were removed? Hint: Look at operationsMetrics in the transaction log using the DESCRIBE HISTORY command.

```
-- TODO

DESCRIBE HISTORY firecallspartitioned
```

Table					
	version	timestamp	userId	userName	operation
1	1	2023-08-02T13:31:00.000+0000	8174178738153308	akanchhakant@gmail.com	DELETE
2	0	2023-08-02T13:25:34.000+0000	8174178738153308	akanchhakant@gmail.com	CREATE TABLE AS SELECT
2 rows					

Question 4

There are quite a few missing Call\_Type\_Group values. Use the UPDATE command to replace any null values with Non Life-threatening .

After you replace the null values, how many Non Life-threatening call types are there?

```
-- TODO

UPDATE firecallspartitioned SET Call_Type_Group = 'Non Life-threatening' WHERE Call_Type_Group IS NULL
```

Table		
	num_affected_rows	

1	246459
---	--------

1 row

Question 5

Travel back in time to the earliest version of the Delta table (version 0). How many records were there?

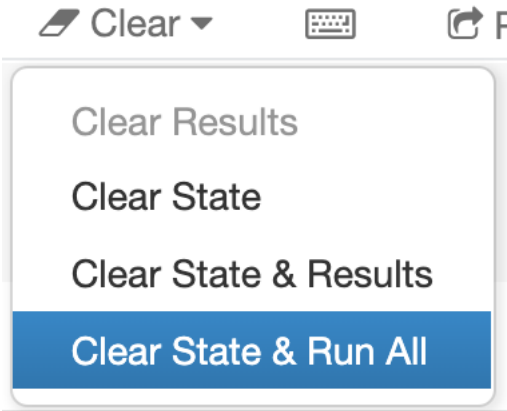
```
-- TODO

SELECT COUNT(*)
FROM firecallspartitioned
  VERSION AS OF 0
```

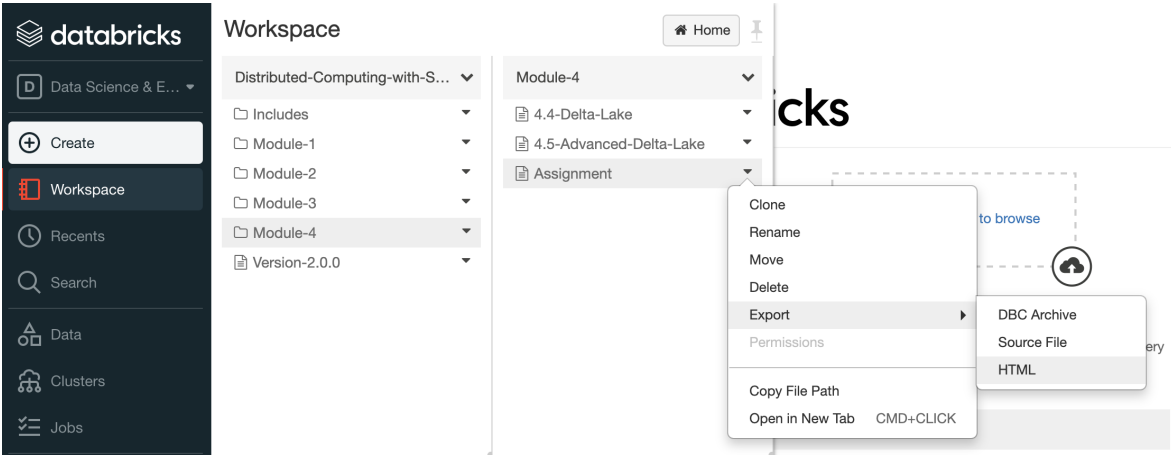
Table	
	count(1) ▲
1	417419

1 row

- Congrats on finishing your last assignment notebook!
- Now you will have to upload this notebook to Coursera for peer reviewing.
- Make sure that all your code will run without errors
  - Check this by clicking the "Clear State & Run All" dropdown option at the top of your notebook



- Click on the "Workspace" icon on the side bar
- Next to the notebook you're working in right now, click on the dropdown arrow
- In the dropdown, click on "Export" then "HTML"



- On the Coursera platform, upload this HTML file to Week 4's Peer Review Assignment

Go back onto the Coursera platform for the free response portion of this assignment and for instructions on how to review your

© 2021 Databricks, Inc. All rights reserved.

Apache, Apache Spark, Spark and the Spark logo are trademarks of the Apache Software Foundation (<http://www.apache.org/>).

Privacy Policy (<https://databricks.com/privacy-policy>) | Terms of Use (<https://databricks.com/terms-of-use>) | Support