Peer-Graded Assignment: Data Management

Course: Managing Big Data in Clusters and Cloud Storage

Name: Balasubramanian PG

Date: 31/12/2023

(Include your name and today's date above.)

Assignment

Create a table named **tbm_sf_la** in the database named **dig** to store the data from three tunnel boring machines (TBMs), which is currently stored in S3 in three separate subdirectories under a directory named **tbm_sf_la** in the bucket named **training-coursera2**. In this document, describe the steps taken to complete this task.

Solution

I performed the following steps to complete this task:

- 1. Examine the Data:
- 2. Create the Table
- 3. Load the Data

CREATE EXTERNAL TABLE dig.tbm_sf_la (
column1 data_type,	
column2 data_type,	
)	
ROW FORMAT DELIMITED	
FIELDS TERMINATED BY ',' Replace with the actual delimiter	
LINES TERMINATED BY '\n'	
STORED AS TEXTFILE	
LOCATION 's3://training-coursera2/tbm_sf_la'	;
Result After performing the steps described above, I ran the following queries and they produced the following result sets: SELECT tbm, COUNT(*) AS num_rows FROM dig.tbm_sf_la GROUP BY tbm ORDER BY tbm;	
-	ig.tbm_sf_la GROUP BY tbm ORDER BY tbm;
-	
SELECT tbm, COUNT(*) AS num_rows FROM d	ig.tbm_sf_la GROUP BY tbm ORDER BY tbm; num_rows
SELECT tbm, COUNT(*) AS num_rows FROM d	
SELECT tbm, COUNT(*) AS num_rows FROM d	
SELECT tbm, COUNT(*) AS num_rows FROM d	
SELECT tbm, COUNT(*) AS num_rows FROM d tbm DESCRIBE dig.tbm_sf_la;	num_rows
SELECT tbm, COUNT(*) AS num_rows FROM d tbm DESCRIBE dig.tbm_sf_la;	num_rows
SELECT tbm, COUNT(*) AS num_rows FROM d tbm DESCRIBE dig.tbm_sf_la;	num_rows
SELECT tbm, COUNT(*) AS num_rows FROM d tbm DESCRIBE dig.tbm_sf_la;	num_rows
SELECT tbm, COUNT(*) AS num_rows FROM d tbm DESCRIBE dig.tbm_sf_la;	num_rows