Big Data on AWS v3.0: Lab 3 - Storing and Querying Data on DynamoDB

==================================================================================================================

Using this command reference.

==================================================================================================================

1. Locate the section you need. Each section in this file matches a section in the lab instructions.

2. Replace items in angle brackets - < > - with appropriate values. For example, in this command you would replace the value - <JobFlowID> - (including the angle brackets) with the parameter indicated in the lab instructions:

elastic-mapreduce --list <JobFlowID>. You can also use find and replace to change bracketed parameters in bulk.

3. Do NOT enable the Word Wrap feature in Windows Notepad or the text editor you use to view this file.

==================================================================================================================

Task 4: Creating External Table References for your Amazon S3 and DynamoDB Data

==================================================================================================================

+++ Step 4.1.2

sudo chown hadoop -R /var/log/hive

mkdir /var/log/hive/user/Hadoop

+++ Step 4.1.7

SET dynamodb.throughput.read.percent=1.0;

+++ Step 4.1.8

CREATE EXTERNAL TABLE orders\_s3\_export ( order\_id string, customer\_id string, order\_date int, total double ) PARTITIONED BY (year string, month string) ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LOCATION 's3://us-west-2-aws-training/awsu-ilt/AWS-200-BIG/v3.0/data/lab-3-dynamodb/data/ddb-orders';

+++ Step 4.1.9

CREATE EXTERNAL TABLE orders\_ddb\_2012\_01 ( order\_id string, customer\_id string, order\_date bigint, total double ) STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler' TBLPROPERTIES ("dynamodb.table.name" = "Orders-2012-01", "dynamodb.column.mapping" = "order\_id:Order\_ID,customer\_id:Customer ID,order\_date:Order Date,total:Total");

==================================================================================================================

Task 5: Using Hive to Load Data into DynamoDB

==================================================================================================================

+++ Step 5.1.1

ALTER TABLE orders\_s3\_export ADD PARTITION (year='2012', month='01');

+++ Step 5.1.2

INSERT OVERWRITE TABLE orders\_ddb\_2012\_01 SELECT order\_id, customer\_id, order\_date, total FROM orders\_s3\_export;

==================================================================================================================

Task 6: Querying Your Data

==================================================================================================================

+++ Step 6.1.1

SELECT customer\_id, sum(total) spend, count(\*) order\_count FROM orders\_ddb\_2012\_01 WHERE order\_date >= unix\_timestamp('2012-01-01', 'yyyy-MM-dd') AND order\_date < unix\_timestamp('2012-01-08', 'yyyy-MM-dd') GROUP BY customer\_id ORDER BY spend desc LIMIT 5;

+++ Step 6.1.2

MSCK REPAIR TABLE orders\_s3\_export;

+++ Step 6.1.3

SELECT year, month, customer\_id, sum(total) spend, count(\*) order\_count FROM orders\_s3\_export WHERE customer\_id = 'c-2cC5fF1bB' AND month >= 6 AND year = 2011 GROUP BY customer\_id, year, month ORDER by month desc;

==================================================================================================================

Task 7: Exporting DynamoDB Data to Amazon S3

==================================================================================================================

+++ Step 7.1.1

CREATE EXTERNAL TABLE orders\_s3\_new\_export ( order\_id string, customer\_id string, order\_date int, total double ) PARTITIONED BY (year string, month string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','LOCATION 's3://dynamo-bucket<yourinitials>/output';

+++ Step 7.1.2

INSERT OVERWRITE TABLE orders\_s3\_new\_export PARTITION (year='2012', month='01') SELECT \* from orders\_ddb\_2012\_01;

Â© 2017 Amazon Web Services, Inc. or its affiliates. All rights reserved. v3.0