LAB 6 BIG DATA

Submitted by:
ABHISHEK ARYA
NET ID: axa220149

Answer: The desirable outputs of LAB 6 are as follows in the 3 immediate screenshots below:

Player ▾

localhost:8888/beeswax/execute/query/57#query/results

Google

Most Visited ▾ Cloudera Hue YARN RM Spark UI (local) Spark Doc Solr Admin UI Kite SDK Doc

Hue

Query Editors ▾   Data Browsers ▾   Workflows ▾   Search   File Browser   Job Browser   training ▾

Hive Editor   **Query Editor**   My Queries   Saved Queries   History

Assist   Settings

DATABASE

default ▾

Table name...

accounts_avro
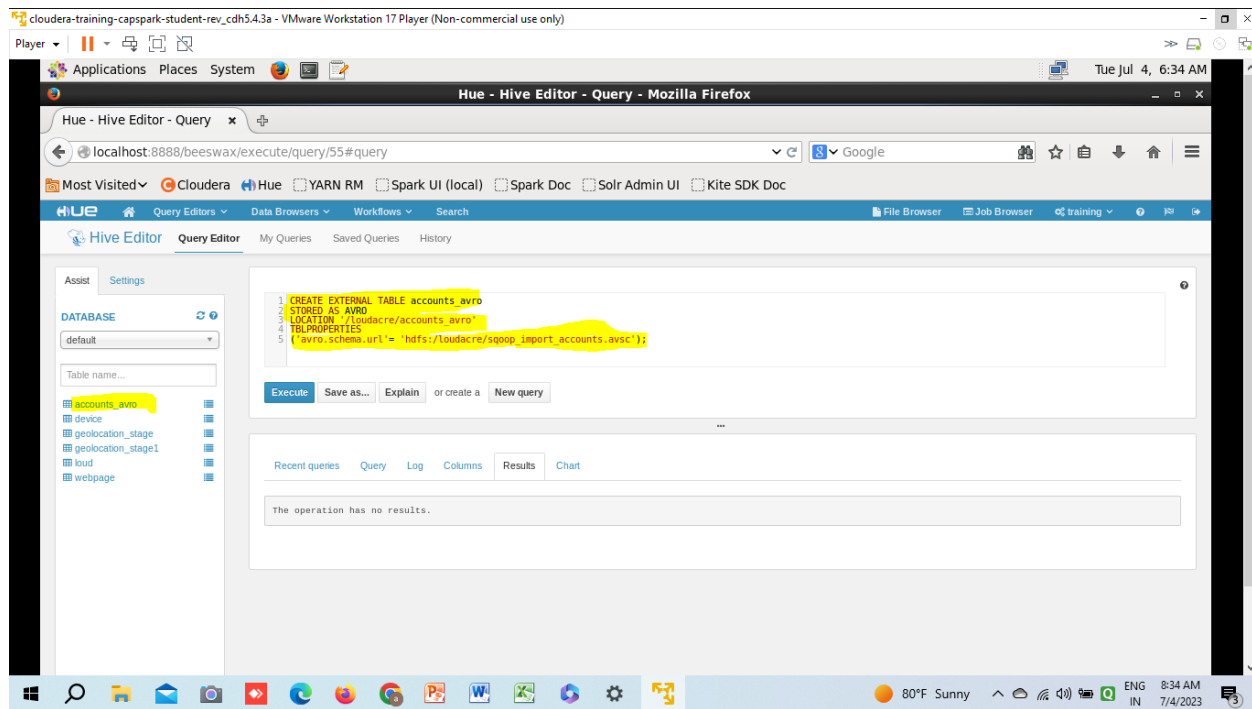device
geolocation_stage
geolocation_stage1
loud
webpage

1 describe formatted accounts_avro

Execute   Save as...   Explain   or create a   New query

Recent queries   Query   Log   Columns   **Results**   Chart

	col_name	data_type	comment
0	# col_name	data_type	comment
1	NULL	NULL	NULL
2	acct_num	int	from deserializer
3	acct_create_dt	bigint	from deserializer
4	acct_close_dt	bigint	from deserializer
5	first_name	string	from deserializer
6	last_name	string	from deserializer
7	address	string	from deserializer
8	city	string	from deserializer
9	state	string	from deserializer

training@localhost:~   Hue - Hive Editor - Qu...

82°F Mostly sunny   ENG IN   8:40 AM 7/4/2023

I am adding all the steps which led me to the above outputs just for the safer side; please find the screenshots of all the steps outputs below:

Player ▾

Applications Places System   Tue Jul 4, 6:44 AM

training@localhost:~/training_materials/dev1/exercises/data-format

File Edit View Search Terminal Help

```
[training@localhost ~]$ cd $DEV1
[training@localhost dev1]$ cd scripts/
[training@localhost scripts]$ ./training_setup_dev1.sh
[Dev] Setting up for Dev1Training
* Setting up Eclipse workspace for Dev1
* /home/training/workspace already exists, moving to /home/training/workspace.sa
ve.dev1
* ERROR: /home/training/workspace.save.dev1 already exists. Delete it and retry.
* Enabling and starting services required for Dev1 Training
Started Impala Catalog Server (catalogd) :                    [  OK  ]
Stopping Apache Archiva...
Apache Archiva was not running.
Stopped Hive Server2:                                         [  OK  ]
* Disabling services not required for Dev1 Training
no secondarynamenode to stop
Stopped Hadoop secondarynamenode:                             [  OK  ]
Flume agent is not running                                    [  OK  ]
Stopping Hadoop HBase regionserver daemon: no regionserver to stop because no pi
d file /var/run/hbase/hbase-hbase-regionserver.pid
hbase-regionserver.
no master to stop because no pid file /var/run/hbase/hbase-hbase-master.pid
Stopped HBase master daemon:                                  [  OK  ]
no rest to stop because no pid file /var/run/hbase/hbase-hbase-rest.pid
Stopped HBase rest daemon:                                    [  OK  ]
no thrift to stop because no pid file /var/run/hbase/hbase-hbase-thrift.pid
Stopped HBase thrift daemon:                                  [  OK  ]
Stopped Kafka Server:                                         [  OK  ]
Solr server daemon is not running                             [  OK  ]
[Dev] Done setting up for Dev1 Training
[training@localhost scripts]$ ls -l
total 20
-rwxr-xr-x 1 training training  876 Aug 27  2015 dev1_eclipse_workspace.sh
-rwxr-xr-x 1 training training 2143 Aug 27  2015 dev1_toggle_services.sh
-rwxr-xr-x 1 training training  864 Aug 27  2015 impala_debug.sh
-rwxr-xr-x 1 training training  883 Aug 27  2015 impala_nodebug.sh
```

82°F Mostly sunny   ENG IN 8:44 AM 7/4/2023

Player ▾

Applications Places System   Tue Jul 4, 6:45 AM

training@localhost:~/training_materials/dev1/exercises/data-format

File Edit View Search Terminal Help

```
[Dev] Done setting up for Dev1 Training
[training@localhost scripts]$ ls -l
total 20
-rwxr-xr-x 1 training training  876 Aug 27  2015 dev1_eclipse_workspace.sh
-rwxr-xr-x 1 training training 2143 Aug 27  2015 dev1_toggle_services.sh
-rwxr-xr-x 1 training training  864 Aug 27  2015 impala_debug.sh
-rwxr-xr-x 1 training training  883 Aug 27  2015 impala_nodebug.sh
-rwxr-xr-x 1 training training  302 Aug 27  2015 training_setup_dev1.sh
[training@localhost scripts]$ pwd
/home/training/training_materials/dev1/scripts
[training@localhost scripts]$ cd exercises/
bash: cd: exercises/: No such file or directory
[training@localhost scripts]$ cd ..
[training@localhost dev1]$ cd exercises/
[training@localhost exercises]$ cd datas-format
bash: cd: datas-format: No such file or directory
[training@localhost exercises]$ cd data-format
[training@localhost data-format]$ ls -l
total 44
-rw-rw-r-- 1 training training 34334 Jul  4 03:48 accounts.java
-rwxr-xr-x 1 training training   209 Jul  4 05:56 sqoop_accounts_avro.sh
-rw-rw-r-- 1 training training  1574 Jul  4 05:57 sqoop_import_accounts.avsc
[training@localhost data-format]$ ./sqoop_accounts_avro.sh
23/07/04 06:43:39 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
23/07/04 06:43:39 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/07/04 06:43:40 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/07/04 06:43:40 INFO tool.CodeGenTool: Beginning code generation
23/07/04 06:43:40 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `accounts` AS t LIMIT 1
23/07/04 06:43:40 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `accounts` AS t LIMIT 1
23/07/04 06:43:40 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-training/compile/d0884c07e7375973cf025ab9686d3da8/accounts.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/07/04 06:43:45 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-training/compile/d0884c07e7375973cf025ab9686d3da8/accounts.jar
23/07/04 06:43:45 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/07/04 06:43:45 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
```

82°F Mostly sunny   ENG IN 8:45 AM 7/4/2023

```
JMX enabled by default
Using config: /etc/zookeeper/conf/zoo.cfg
Starting zookeeper ... already running as process 1705.
[training@localhost ~]$ sudo service hive-server2 start
Started Hive Server2 (hive-server2):              [  OK  ]
[training@localhost ~]$ hdfs dfs -ls /loudacre
Found 6 items
drwxrwxrwx   - training supergroup          0 2023-06-15 08:20 /loudacre/account
s
drwxrwxrwx   - training supergroup          0 2023-07-04 06:05 /loudacre/account
s_avro
drwxrwxrwx   - training supergroup          0 2023-06-28 06:19 /loudacre/input_f
ile
drwxrwxrwx   - training supergroup          0 2023-06-13 00:40 /loudacre/kb
-rw-rw-rw-   1 training supergroup       1574 2023-07-04 06:21 /loudacre/sqoop_i
mport_accounts.avsc
drwxrwxrwx   - training supergroup          0 2023-06-15 07:57 /loudacre/webpage
[training@localhost ~]$ hdfs dfs -ls /loudacre
Found 6 items
drwxrwxrwx   - training supergroup          0 2023-06-15 08:20 /loudacre/accounts
drwxrwxrwx   - training supergroup          0 2023-07-04 06:05 /loudacre/accounts_avro
drwxrwxrwx   - training supergroup          0 2023-06-28 06:19 /loudacre/input_file
drwxrwxrwx   - training supergroup          0 2023-06-13 00:40 /loudacre/kb
-rw-rw-rw-   1 training supergroup       1574 2023-07-04 06:21 /loudacre/sqoop_import_accounts.avsc
drwxrwxrwx   - training supergroup          0 2023-06-15 07:57 /loudacre/webpage
[training@localhost ~]$
```