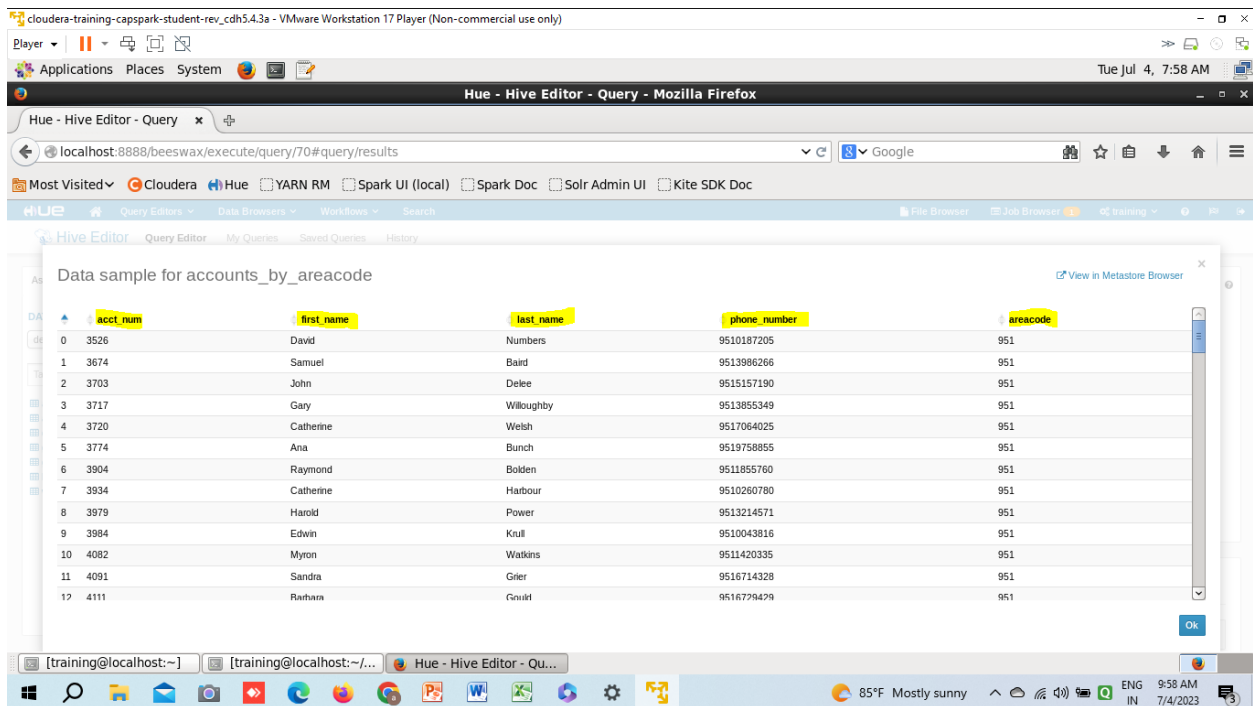# LAB 7 BIG DATA

Submitted by: ABHISHEK ARYA

NET ID: axa220149

Answer: The desirable outputs of LAB 7 are as follows in the 2 immediate screenshots below:

```
cloudera-training-capspark-student-rev_cdh5.4.3a - VMware Workstation 17 Player (Non-commercial use only)
Player ▾   ‖ ▾   ⊕  ⊡  ⊠                                                                    ⇒ ⌷ ⊙ ⊡

Applications  Places  System                                                         Tue Jul  4,  8:01 AM

                                    training@localhost:~

File  Edit  View  Search  Terminal  Help
[training@localhost ~]$ sudo service hive-server2 start
Started Hive Server2 (hive-server2):                     [  OK  ]
[training@localhost ~]$ hdfs dfs -ls /loudacre/account_by_areacode
ls: `/loudacre/account_by_areacode': No such file or directory
[training@localhost ~]$ hdfs dfs -ls /loudacre/accounts_by_areacode
[training@localhost ~]$ hdfs dfs -ls /loudacre
Found 7 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 2023-07-04 07:58 /loudacre/accounts_by_areacode
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 ~]$ hdfs dfs -ls /loudacre/accounts_by_areacode
Found 31 items
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/.hive-staging_hive_2023-07-04_07-57-11_833_447535822374462858-1
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=209
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=213
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=310
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=408
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=415
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=424
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=503
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=510
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=530
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=541
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=559
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=562
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=619
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=626
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=650
drwxrwxrwx   - training supergroup          0 2023-07-04 07:58 /loudacre/accounts_by_areacode/areacode=657
```

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:

```sql
SET hive.exec.dynamic.partition=true;
SET hive.exec.dynamic.partition.mode=nonstrict;


CREATE EXTERNAL TABLE accounts_by_areacode (
acct_num INT,
first_name STRING,
last_name STRING,
phone_number STRING)
PARTITIONED BY (areacode STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION '/loudacre/accounts_by_areacode';

SELECT
acct_num,
first_name,
last_name
```



```sql
last_name STRING,
phone_number STRING)
PARTITIONED BY (areacode STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
LOCATION '/loudacre/accounts_by_areacode';

SELECT
acct_num,
first_name,
last_name,
phone_number,
SUBSTR(phone_number,1,3) AS areacode
FROM accounts_avro
```

	acct_num	first_name	last_name	phone_number	areacode
0	1	Donald	Becton	5100032418	510
1	2	Donna	Jones	4150835799	415
2	3	Dorthy	Chalmers	6506877757	650
3	4	Leila	Spencer	6503198619	650

Player

Applications Places System

Tue Jul 4, 8:04 AM

Hue - Hive Editor - Query - Mozilla Firefox

Hue - Hive Editor - Query

localhost:8888/beeswax/execute/query/70#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
accounts_by_areacode
device
geolocation_stage
geolocation_stage1
loud
webpage

```
16  LOCATION '/loudacre/accounts_by_areacode';
17
18  SELECT
19  acct_num,
20  first_name,
21  last_name,
22  phone_number,
23  SUBSTR(phone_number,1,3) AS areacode
24
25  FROM accounts_avro
26
27  INSERT INTO TABLE accounts_by_areacode
28  PARTITION (areacode)
29  SELECT
30  acct_num,
31  first_name,
32  last_name,
33  phone_number,
34  SUBSTR(phone_number,1,3) AS areacode
35  FROM accounts_avro
```

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

Recent queries   Query   Log   Columns   Results   Chart

The operation has no results.

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

85°F Mostly sunny   ENG IN   10:04 AM 7/4/2023