**1. Create a customer\_hive table on the top of 'customer' table created in the last session.**

**Calculate the maximum and minimum age of customer from the table.**

hive>create external table customer\_hive(id int, name string, location string,age int) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' with serdeproperties ("hbase.columns.mapping" = ":key,details:name,details:location,details:age") tblproperties("hbase.table.name" = "customer");

hive> select \* from customer\_hive;

1 Amit IND 18

2 Sumit PAK 20

3 Rohit AUS 26

4 Namit UK 24

hive> select MAX(age) as max\_age from customer\_hive;

Query ID = acadgild\_20170505195858\_555aff74-716f-4dc7-8344-f21a958a7411

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job\_1493966073243\_0007, Tracking URL = http://localhost:8088/proxy/application\_1493966073243\_0007/

Kill Command = /home/acadgild/hadoop-2.6.0/bin/hadoop job -kill job\_1493966073243\_0007

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1

2017-05-05 19:58:48,702 Stage-1 map = 0%, reduce = 0%

2017-05-05 19:59:48,926 Stage-1 map = 0%, reduce = 0%

2017-05-05 20:00:05,458 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.06 sec

2017-05-05 20:01:05,741 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.06 sec

2017-05-05 20:01:38,662 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.62 sec

MapReduce Total cumulative CPU time: 10 seconds 620 msec

Ended Job = job\_1493966073243\_0007

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.62 sec HDFS Read: 255 HDFS Write: 3 SUCCESS

Total MapReduce CPU Time Spent: 10 seconds 620 msec

OK

26

Time taken: 220.443 seconds, Fetched: 1 row(s)

hive> select MIN(age) as min\_age from customer\_hive;

**18**

**2. Access the customer hbase table from pig and compute the maximum and minimum age**

**among all the customers along with their corresponding name and id.**

**Step 1:create an empty table in hbase shell.**

hbase(main):004:0> create 'customer',{NAME=>'details'}

0 row(s) in 2.7310 seconds

=> Hbase::Table - customer

hbase(main):005:0> scan 'customer'

ROW COLUMN+CELL

**Step 2:Run pig shell and run the below query to load the dataset from hdfs location to hbase.**

grunt> raw\_data= load '/customers.dat' using PigStorage(',') as (id:int,name:chararray,location:chararray,age:int);

grunt> store raw\_data into 'hbase://customer' using org.apache.pig.backend.hadoop.hbase.HBaseStorage('details:name,details:location,details:age');

**Step 3:Login to hbase command line and scan the table created to cross check if the data has been loaded.**

hbase(main):006:0> scan 'customer'

hbase(main):003:0> scan 'customer'

ROW COLUMN+CELL

1 column=details:age, timestamp=1494086673487, value=18

1 column=details:location, timestamp=1494086673487, value=IN

D

1 column=details:name, timestamp=1494086673487, value=Amit

2 column=details:age, timestamp=1494086673557, value=20

2 column=details:location, timestamp=1494086673557, value=PA

K

2 column=details:name, timestamp=1494086673557, value=Sumit

3 column=details:age, timestamp=1494086673558, value=26

3 column=details:location, timestamp=1494086673558, value=AU

S

3 column=details:name, timestamp=1494086673558, value=Rohit

4 column=details:age, timestamp=1494086673558, value=24

4 column=details:location, timestamp=1494086673558, value=UK

4 column=details:name, timestamp=1494086673558, value=Namit

4 row(s) in 0.2910 seconds

**Step 4: Login to PIG and run the below queries to find out maximum age and minimum age;**

**MAX\_AGE:**

A =order raw\_data by age desc;

B= LIMIT A 1;

C= foreach B generate $0,$1,$3;

dump C;

(3,Rohit,26)

**MIN\_AGE:**

A =order raw\_data by age asc;

B= LIMIT A 1;

C= foreach B generate $0,$1,$3;

dump C;

(1,Amit,18)