## Question 1 – Answer

hive> show databases;

OK

custom

default

Time taken: 0.843 seconds, Fetched: 2 row(s)

hive> use custom;

OK

Time taken: 0.051 seconds

hive> show tables;

OK

temperature\_data

Time taken: 0.099 seconds, Fetched: 1 row(s)

hive> describe temperature\_data;

OK

date string

zip\_code int

temperature int

Time taken: 0.523 seconds, Fetched: 3 row(s)

hive> select date, temperature from temperature\_data where zip\_code BETWEEN 300000 AND 399999;

OK

10-03-1990 15

10-01-1991 22

12-02-1990 9

10-03-1991 16

10-01-1990 23

12-02-1991 10

10-03-1993 16

10-01-1994 23

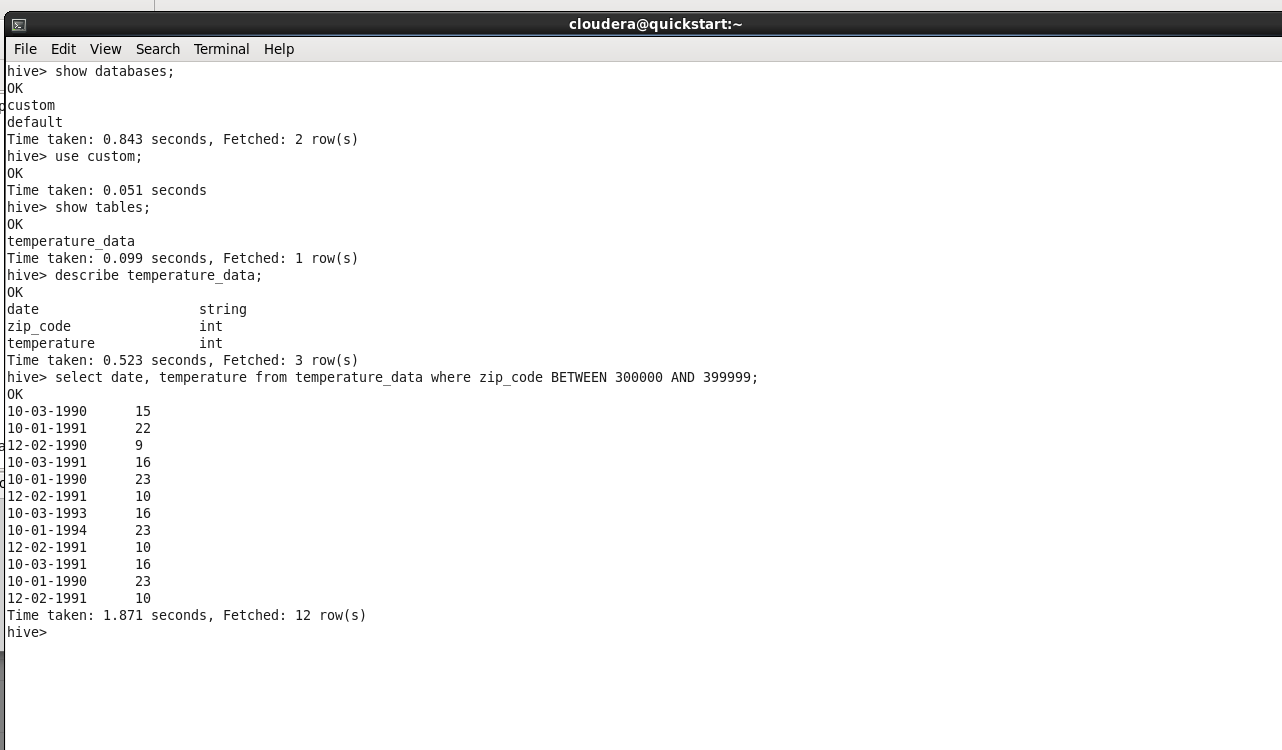
12-02-1991 10

10-03-1991 16

10-01-1990 23

12-02-1991 10

Time taken: 1.871 seconds, Fetched: 12 row(s)



## Question 2 – Answer

hive> select YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date)), max(temperature)

> from temperature\_data

> group by YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date));

Query ID = cloudera\_20171220081414\_d30e541b-5c2a-4f76-a636-6addb41a5df8

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 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\_1513783240235\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1513783240235\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1513783240235\_0001

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

2017-12-20 08:21:04,121 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:27:14,985 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:28:20,748 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:28:23,149 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.47 sec

2017-12-20 08:28:37,773 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.27 sec

MapReduce Total cumulative CPU time: 5 seconds 270 msec

Ended Job = job\_1513783240235\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.27 sec HDFS Read: 9112 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 270 msec

OK

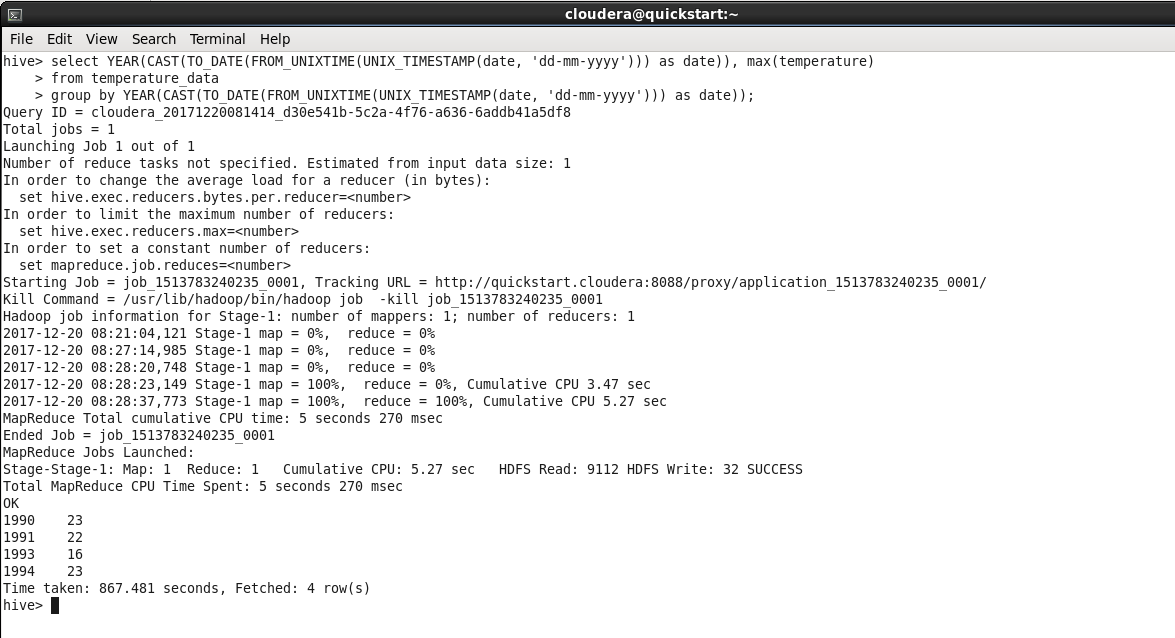
1990 23

1991 22

1993 16

1994 23

Time taken: 867.481 seconds, Fetched: 4 row(s)



## Question 3 – Answer

hive> select YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date)), max(temperature)

> from temperature\_data

> group by YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date))

> having COUNT(YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date))) > 1;

Query ID = cloudera\_20171220083535\_2599de5a-2b84-4945-929b-4f2bbb93bf04

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 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\_1513783240235\_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1513783240235\_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1513783240235\_0002

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

2017-12-20 08:35:29,194 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:35:40,698 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.2 sec

2017-12-20 08:35:53,022 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.95 sec

MapReduce Total cumulative CPU time: 4 seconds 950 msec

Ended Job = job\_1513783240235\_0002

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.95 sec HDFS Read: 10597 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 950 msec

OK

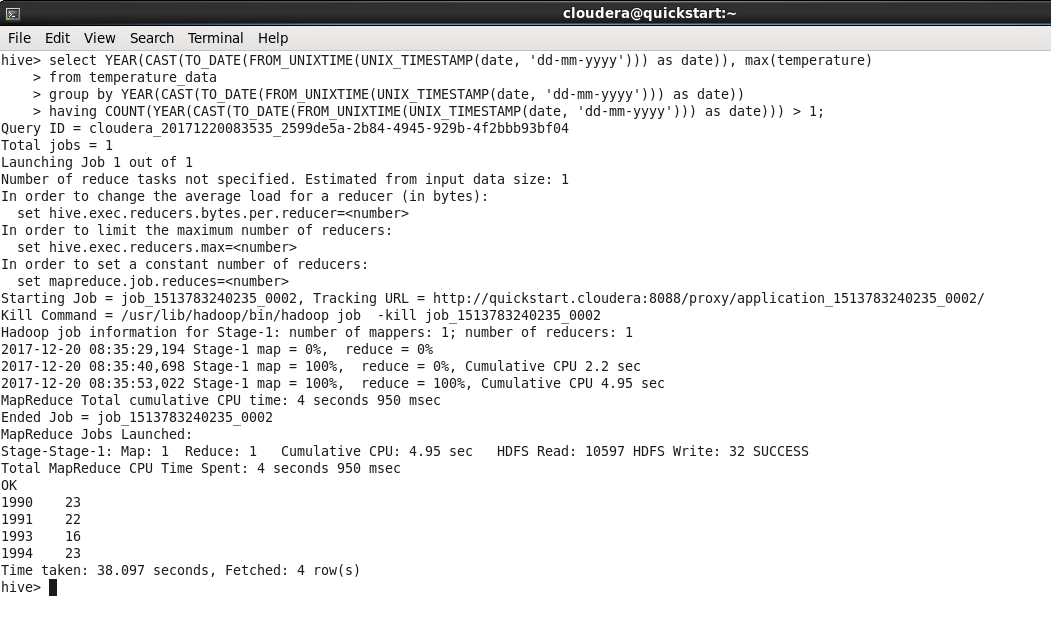
1990 23

1991 22

1993 16

1994 23

Time taken: 38.097 seconds, Fetched: 4 row(s)



## Question 4 – Answer

hive> create view temperature\_data\_vw as

> select YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date)), max(temperature)

> from temperature\_data

> group by YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date))

> having COUNT(YEAR(CAST(TO\_DATE(FROM\_UNIXTIME(UNIX\_TIMESTAMP(date, 'dd-mm-yyyy'))) as date))) > 1;

OK

Time taken: 0.629 seconds

hive> show tables;

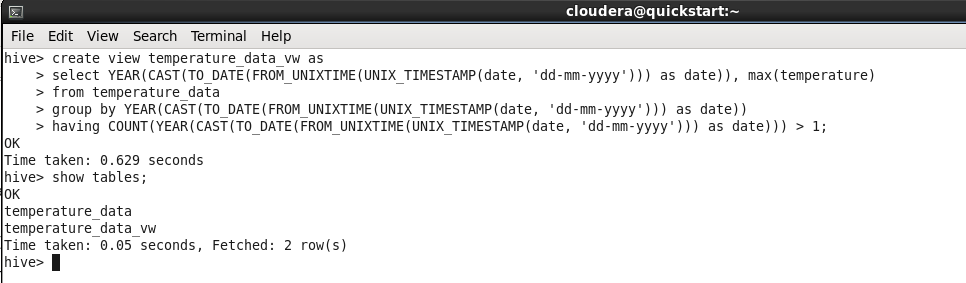
OK

temperature\_data

temperature\_data\_vw

Time taken: 0.05 seconds, Fetched: 2 row(s)

hive>



## Question 5 – Answer

hive> INSERT OVERWRITE LOCAL DIRECTORY '/home/cloudera/Downloads' ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'

> select \* from temperature\_data\_vw;

Query ID = cloudera\_20171220084545\_c26f338e-65a0-4801-894f-a64e964c00ba

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 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\_1513783240235\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1513783240235\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1513783240235\_0003

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

2017-12-20 08:46:23,480 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:53:16,434 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:56:30,974 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:57:32,731 Stage-1 map = 0%, reduce = 0%

2017-12-20 08:58:02,074 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.9 sec

2017-12-20 08:58:15,468 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.74 sec

MapReduce Total cumulative CPU time: 5 seconds 740 msec

Ended Job = job\_1513783240235\_0003

Copying data to local directory /home/cloudera/Downloads

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.74 sec HDFS Read: 10456 HDFS Write: 32 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 740 msec

OK

Time taken: 749.105 seconds

