**hive (niit)> create table NYSE(name STRING,stock STRING ,dates STRING ,open FLOAT , high FLOAT, low FLOAT, close FLOAT, volume BIGINT, adj\_close FLOAT)**

**row format delimited**

**fields terminated by ','**

**stored as textfile;**

OK

Time taken: 0.343 seconds

**hive (niit)> show tables;**

OK

tab\_name

**customer**

**nyse**

**txnrecords**

**hive (niit)> LOAD DATA LOCAL INPATH '/home/hduser/InputData/NYSE.csv' OVERWRITE INTO TABLE NYSE;**

Loading data to table niit.nyse

Table niit.nyse stats: [numFiles=1, numRows=0, totalSize=40990862, rawDataSize=0]

OK

Time taken: 0.568 seconds

**hive (niit)> select count(\*) from NYSE;**

Query ID = hduser\_20171222204552\_98f8d2dd-a4a4-4a30-933e-e2c70c41aa0a

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\_1513946090631\_0008, Tracking URL = http://rootuser:8088/proxy/application\_1513946090631\_0008/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1513946090631\_0008

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

2017-12-22 20:45:56,721 Stage-1 map = 0%, reduce = 0%

2017-12-22 20:46:00,877 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.23 sec

2017-12-22 20:46:06,094 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.47 sec

MapReduce Total cumulative CPU time: 3 seconds 470 msec

Ended Job = job\_1513946090631\_0008

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.47 sec HDFS Read: 41002323 HDFS Write: 7 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 470 msec

OK

\_c0

735026

Time taken: 14.819 seconds, Fetched: 1 row(s)

**hive (niit)> describe NYSE;**

OK

col\_name data\_type comment

name string

stock string

dates string

open float

high float

low float

close float

volume bigint

adj\_close float

Time taken: 0.055 seconds, Fetched: 9 row(s)

**hive (niit)> select stock,sum(volume)**

**> as totalvolume**

**> from NYSE**

**> GROUP BY stock**

**> ORDER BY totalvolume desc;**

Query ID = hduser\_20171222205303\_07176173-46f2-4a13-9e8a-87adca5ec8e8

Total jobs = 2

Launching Job 1 out of 2

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\_1513946090631\_0009, Tracking URL = http://rootuser:8088/proxy/application\_1513946090631\_0009/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1513946090631\_0009

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

2017-12-22 20:53:07,673 Stage-1 map = 0%, reduce = 0%

2017-12-22 20:53:12,823 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.38 sec

2017-12-22 20:53:16,931 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.85 sec

MapReduce Total cumulative CPU time: 4 seconds 850 msec

Ended Job = job\_1513946090631\_0009

Launching Job 2 out of 2

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\_1513946090631\_0010, Tracking URL = http://rootuser:8088/proxy/application\_1513946090631\_0010/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1513946090631\_0010

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2017-12-22 20:53:26,810 Stage-2 map = 0%, reduce = 0%

2017-12-22 20:53:30,918 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.89 sec

2017-12-22 20:53:36,061 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.54 sec

MapReduce Total cumulative CPU time: 2 seconds 540 msec

Ended Job = job\_1513946090631\_0010

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.85 sec HDFS Read: 41002418 HDFS Write: 5442 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.54 sec HDFS Read: 9996 HDFS Write: 2918 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 390 msec

OK

stock totalvolume

AMD 47252808500

AA 42061448400

AXP 40263020300

AET 30218027200

ABT 25664130200

-----

-----

Time taken: 33.333 seconds, Fetched: 203 row(s)

Finding TOP 5 variance

hive (niit)> insert overwrite directory '/niit/Top5Variance' row format delimited fields terminated by ','

> select stock, max((high-low)/low\*100) as MaxVariance from NYSE

> GROUP BY stock

> ORDER by MaxVariance desc

> LIMIT 5;

Query ID = hduser\_20171222210322\_be4c6a74-9bd1-492b-9daa-5a28bc3a0ae4

Total jobs = 2

Launching Job 1 out of 2

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\_1513946090631\_0011, Tracking URL = http://rootuser:8088/proxy/application\_1513946090631\_0011/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1513946090631\_0011

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

2017-12-22 21:03:26,545 Stage-1 map = 0%, reduce = 0%

2017-12-22 21:03:32,700 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.83 sec

2017-12-22 21:03:36,794 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 7.37 sec

MapReduce Total cumulative CPU time: 7 seconds 370 msec

Ended Job = job\_1513946090631\_0011

Launching Job 2 out of 2

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\_1513946090631\_0012, Tracking URL = http://rootuser:8088/proxy/application\_1513946090631\_0012/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1513946090631\_0012

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2017-12-22 21:03:46,274 Stage-2 map = 0%, reduce = 0%

2017-12-22 21:03:50,377 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 0.82 sec

2017-12-22 21:03:55,495 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.27 sec

MapReduce Total cumulative CPU time: 2 seconds 270 msec

Ended Job = job\_1513946090631\_0012

Moving data to: /niit/Top5Variance

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 7.37 sec HDFS Read: 41004808 HDFS Write: 6017 SUCCESS

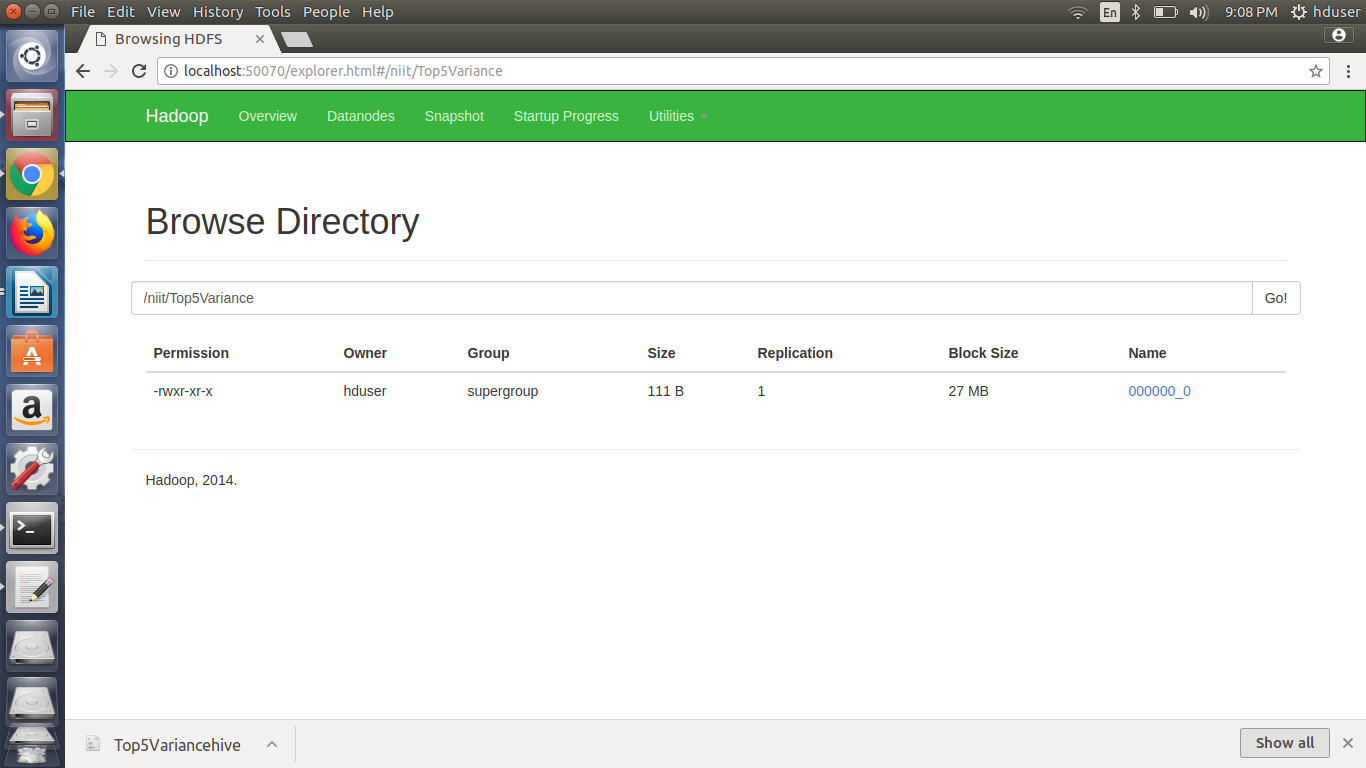
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.27 sec HDFS Read: 10576 HDFS Write: 111 SUCCESS

Total MapReduce CPU Time Spent: 9 seconds 640 msec

OK

stock maxvariance

Time taken: 34.479 seconds



ALY,324.999995343387

AIG,309.59999084472656

AFF,291.33857203413856

AI,233.33333954215064

AGM,187.9501661566204

Select top 10 customers from transaction data who has done maximum purchase.

hive (niit)> describe txnrecords;

OK

col\_name data\_type comment

txnno int

txndate string

custno int

amount double

category string

product string

city string

state string

spendby string

Time taken: 0.058 seconds, Fetched: 9 row(s)

hive (niit)> select custno, count(\*) as TotalPurchase from txnrecords group by custno order by TotalPurchase desc limit 10;

Query ID = hduser\_20171223091017\_1c2f9569-e7ea-4656-8926-917b5a1466ed

Total jobs = 2

Launching Job 1 out of 2

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\_1514000082488\_0004, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0004/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0004

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

2017-12-23 09:10:21,342 Stage-1 map = 0%, reduce = 0%

2017-12-23 09:10:25,466 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.04 sec

2017-12-23 09:10:30,644 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.19 sec

MapReduce Total cumulative CPU time: 4 seconds 190 msec

Ended Job = job\_1514000082488\_0004

Launching Job 2 out of 2

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\_1514000082488\_0005, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0005/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0005

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2017-12-23 09:10:42,070 Stage-2 map = 0%, reduce = 0%

2017-12-23 09:10:46,191 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.14 sec

2017-12-23 09:10:50,297 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.7 sec

MapReduce Total cumulative CPU time: 2 seconds 700 msec

Ended Job = job\_1514000082488\_0005

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.19 sec HDFS Read: 4425539 HDFS Write: 220670 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.7 sec HDFS Read: 225295 HDFS Write: 110 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 890 msec

OK

custno totalpurchase

4006606 16

4000815 14

4009485 14

4005991 14

4009673 13

4005751 13

4003528 13

4002572 13

4002033 13

4000221 13

Time taken: 33.482 seconds, Fetched: 10 row(s)

Select top 10 customers from transaction data who has spent maximum amount.

hive (niit)> select custno, sum(amount) as totalamount from txnrecords group by custno order by totalamount desc limit 10;

Query ID = hduser\_20171223092001\_52f17b1a-fe03-4e3b-b2b1-3718eb91acad

Total jobs = 2

Launching Job 1 out of 2

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\_1514000082488\_0006, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0006/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0006

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

2017-12-23 09:20:06,043 Stage-1 map = 0%, reduce = 0%

2017-12-23 09:20:10,181 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.17 sec

2017-12-23 09:20:14,299 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.71 sec

MapReduce Total cumulative CPU time: 3 seconds 710 msec

Ended Job = job\_1514000082488\_0006

Launching Job 2 out of 2

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\_1514000082488\_0007, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0007/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0007

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2017-12-23 09:20:24,282 Stage-2 map = 0%, reduce = 0%

2017-12-23 09:20:27,504 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.29 sec

2017-12-23 09:20:31,611 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.81 sec

MapReduce Total cumulative CPU time: 2 seconds 810 msec

Ended Job = job\_1514000082488\_0007

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.71 sec HDFS Read: 4425666 HDFS Write: 290839 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.81 sec HDFS Read: 295500 HDFS Write: 203 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 520 msec

OK

custno totalamount

4009485 1973.3

4006425 1732.09

4000221 1671.4700000000003

4003228 1640.63

4006606 1628.9399999999996

4006467 1605.9499999999998

4004927 1576.71

4008321 1560.79

4000815 1557.8200000000002

4001051 1488.67

Time taken: 30.855 seconds, Fetched: 10 row(s)

We can see some record has a lot of decimal places. Use the round() function to round off to a particular number of decimal places

hive (niit)> select custno, round(sum(amount),2) as totalamount from txnrecords group by custno order by totalamount desc limit 10;

Query ID = hduser\_20171223092429\_717ab9dd-3b24-44fa-b67c-33044dc1e56e

Total jobs = 2

Launching Job 1 out of 2

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\_1514000082488\_0008, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0008/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0008

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

2017-12-23 09:24:34,099 Stage-1 map = 0%, reduce = 0%

2017-12-23 09:24:38,218 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.05 sec

2017-12-23 09:24:43,385 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.85 sec

MapReduce Total cumulative CPU time: 4 seconds 850 msec

Ended Job = job\_1514000082488\_0008

Launching Job 2 out of 2

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\_1514000082488\_0009, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0009/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0009

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2017-12-23 09:24:53,161 Stage-2 map = 0%, reduce = 0%

2017-12-23 09:24:56,369 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.19 sec

2017-12-23 09:25:00,481 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.71 sec

MapReduce Total cumulative CPU time: 2 seconds 710 msec

Ended Job = job\_1514000082488\_0009

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.85 sec HDFS Read: 4425999 HDFS Write: 290839 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.71 sec HDFS Read: 295335 HDFS Write: 159 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 560 msec

OK

custno totalamount

4009485 1973.3

4006425 1732.09

4000221 1671.47

4003228 1640.63

4006606 1628.94

4006467 1605.95

4004927 1576.71

4008321 1560.79

4000815 1557.82

4001051 1488.67

Time taken: 31.737 seconds, Fetched: 10 row(s)

All the joins of MySQL will be applicable here.

hive (niit)> select a.custno,firstname, lastname, age, profession, round(sum(amount),2) as totalamt from txnrecords a join customer b on (a.custno=b.custno) group by custno order by totalamt desc limit 10;

FAILED: SemanticException Column custno Found in more than One Tables/Subqueries

hive (niit)> select a.custno,firstname, lastname, age, profession, round(sum(amount),2) as totalamt from txnrecords a join customer b on (a.custno=b.custno) group by a.custno order by totalamt desc limit 10;

FAILED: SemanticException [Error 10025]: Line 1:16 Expression not in GROUP BY key 'firstname'

This statement will work in SQL not in Hive ---- Reason: Mapreduce cannot figure out which columns are your key and which are your value.

**hive (niit)> select a.custno,firstname, lastname, age, profession, round(sum(amount),2) as totalamt from txnrecords a join customer b on (a.custno=b.custno) group by a.custno, firstname, lastname, age, profession order by totalamt desc limit 10;**

Query ID = hduser\_20171223093228\_4a1eba42-69eb-41b4-9995-0478e74f23e9

Total jobs = 2

17/12/23 09:32:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Execution log at: /tmp/hduser/hduser\_20171223093228\_4a1eba42-69eb-41b4-9995-0478e74f23e9.log

2017-12-23 09:32:30 Starting to launch local task to process map join; maximum memory = 477626368

2017-12-23 09:32:31 Dump the side-table for tag: 1 with group count: 10000 into file: file:/usr/local/hive/iotmp/0324a8f9-6bc1-49bf-bf2e-e8ecb0a58aeb/hive\_2017-12-23\_09-32-28\_799\_5171931717117834272-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile01--.hashtable

2017-12-23 09:32:31 Uploaded 1 File to: file:/usr/local/hive/iotmp/0324a8f9-6bc1-49bf-bf2e-e8ecb0a58aeb/hive\_2017-12-23\_09-32-28\_799\_5171931717117834272-1/-local-10005/HashTable-Stage-2/MapJoin-mapfile01--.hashtable (504636 bytes)

2017-12-23 09:32:31 End of local task; Time Taken: 0.869 sec.

Execution completed successfully

MapredLocal task succeeded

Launching Job 1 out of 2

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\_1514000082488\_0010, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0010/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0010

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2017-12-23 09:32:35,525 Stage-2 map = 0%, reduce = 0%

2017-12-23 09:32:40,658 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 4.22 sec

2017-12-23 09:32:44,789 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 7.17 sec

MapReduce Total cumulative CPU time: 7 seconds 170 msec

Ended Job = job\_1514000082488\_0010

Launching Job 2 out of 2

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\_1514000082488\_0011, Tracking URL = http://rootuser:8088/proxy/application\_1514000082488\_0011/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514000082488\_0011

Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1

2017-12-23 09:32:54,305 Stage-3 map = 0%, reduce = 0%

2017-12-23 09:32:57,384 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 1.35 sec

2017-12-23 09:33:01,498 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 2.89 sec

MapReduce Total cumulative CPU time: 2 seconds 890 msec

Ended Job = job\_1514000082488\_0011

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 7.17 sec HDFS Read: 4433262 HDFS Write: 573178 SUCCESS

Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 2.89 sec HDFS Read: 578714 HDFS Write: 468 SUCCESS

Total MapReduce CPU Time Spent: 10 seconds 60 msec

OK

a.custno firstname lastname age profession totalamt

4009485 Stuart House 58 Teacher 1973.3

4006425 Joe Burns 30 Economist 1732.09

4000221 Glenda Boswell 28 Civil engineer 1671.47

4003228 Elsie Newton 54 Accountant 1640.63

4006606 Jackie Lewis 66 Recreation and fitness worker 1628.94

4006467 Evelyn Monroe 37 Financial analyst 1605.95

4004927 Joan Lowry 30 Librarian 1576.71

4008321 Paul Carey 64 Human resources assistant 1560.79

4000815 Julie Galloway 53 Actor 1557.82

4001051 Arlene Higgins 62 Police officer 1488.67

Time taken: 34.814 seconds, Fetched: 10 row(s)

**Substring method in hive:**

**In hive there is no index value …. Therefore strings will start from 1 and not 0 (as in java)**

**hive (niit)> select substr('hadoop',1,2);**

OK

\_c0

ha

Time taken: 0.088 seconds, Fetched: 1 row(s)

**hive (niit)> select substr('hadoop',3,2);**

OK

\_c0

do

Time taken: 0.048 seconds, Fetched: 1 row(s)

**hive (niit)> select substr('hadoop',3,1);**

OK

\_c0

d

Time taken: 0.044 seconds, Fetched: 1 row(s)

**hive (niit)> select substr('2001-02-01 00:00:00', 6, 2);**

OK

\_c0

02

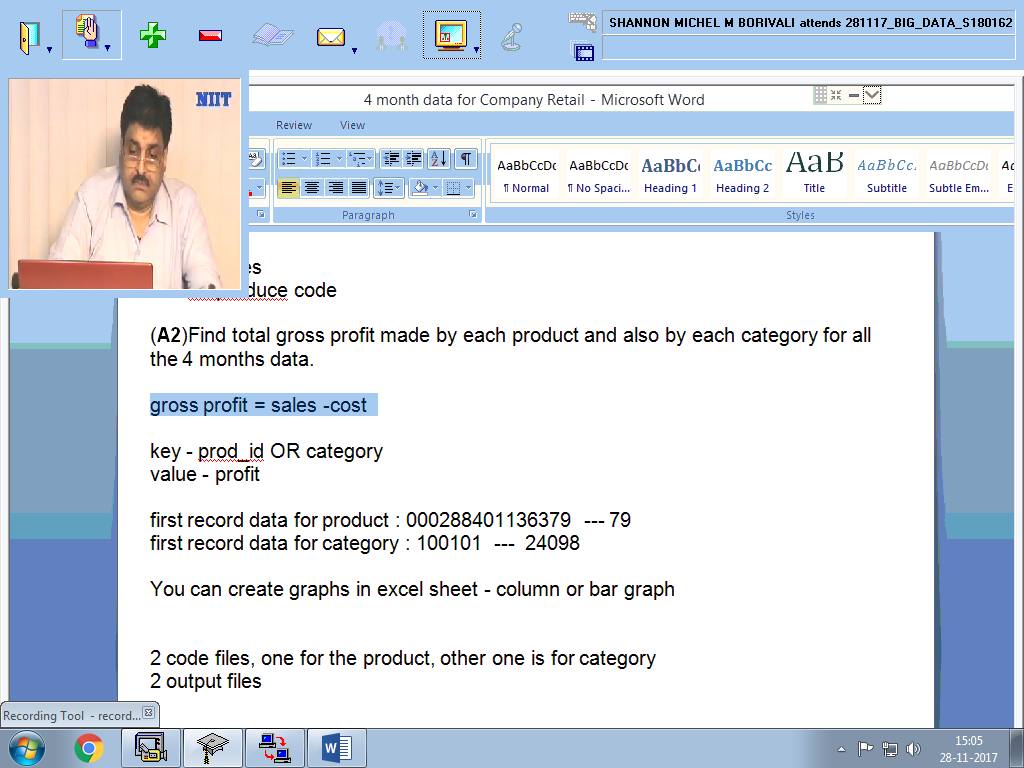
**hive (niit)> select month('2001-02-01 00:00:00');**

OK

\_c0

2

Time taken: 0.246 seconds, Fetched: 1 row(s)



hive (retail)> create table retailRecords(dates STRING, custid BIGINT, Age STRING, Residence STRING, ProductCategory BIGINT, ProductID BIGINT, Qty INT, Cost DOUBLE, Sales DOUBLE) row format delimited fields terminated by'\;' stored as textfile;

OK

Time taken: 0.092 seconds

hive (retail)> describe retailRecords;

OK

col\_name data\_type comment

dates string

custid bigint

age string

residence string

productcategory bigint

productid bigint

qty int

cost double

sales double

Time taken: 0.053 seconds, Fetched: 9 row(s)

**LOAD DATA LOCAL INPATH '//home/hduser/D11' OVERWRITE INTO TABLE retailRecords;**

**LOAD DATA LOCAL INPATH '/home/hduser/D12' OVERWRITE INTO TABLE retailRecords;**

**LOAD DATA LOCAL INPATH '/home/hduser/D01' OVERWRITE INTO TABLE retailRecords;**

**LOAD DATA LOCAL INPATH '/home/hduser/D02' OVERWRITE INTO TABLE retailRecords;**