ArrayFile:

1,abc,40000,a$b$c,city1

2,def,3000,d$f,city2

------------------------------------------------------------------------------------------------------------------------

hive (complexDatatypes)> create table arrayType(id int,name string,sal bigint,sub array<string>,city string) row format delimited fields terminated by ',' collection items terminated by '$';

OK

Time taken: 0.473 seconds

------------------------------------------------------------------------------------------------------------------------

hive (complexDatatypes)> load data local inpath '/home/hduser/arrayfile' overwrite into table arrayType;

Loading data to table complexdatatypes.arraytype

Table complexdatatypes.arraytype stats: [numFiles=1, numRows=0, totalSize=45, rawDataSize=0]

OK

Time taken: 1.014 seconds

-----------------------------------------------------------------------------------------------------------------------

hive (complexDatatypes)> select \* from arrayType;

OK

arraytype.id arraytype.name arraytype.sal arraytype.sub arraytype.city

1 abc 40000 ["a","b","c"] city1

2 def 3000 ["d","f"] city2

Time taken: 0.346 seconds, Fetched: 2 row(s)

hive (complexDatatypes)> select \* from arrayType where id=1;

OK

arraytype.id arraytype.name arraytype.sal arraytype.sub arraytype.city

1 abc 40000 ["a","b","c"] city1

Time taken: 0.209 seconds, Fetched: 1 row(s)

hive (complexDatatypes)> select id, name, sal, sub[2] from arrayType;

OK

id name sal \_c3

1 abc 40000 c

2 def 3000 NULL

Time taken: 0.049 seconds, Fetched: 2 row(s)

hive (complexDatatypes)> select \* from arrayType where array\_contains(sub, 'c');

OK

arraytype.id arraytype.name arraytype.sal arraytype.sub arraytype.city

1 abc 40000 ["a","b","c"] city1

Time taken: 0.092 seconds, Fetched: 1 row(s)

MapFile:

1,abc,40000,a$b$c,bonus#500$insurance#200,city1

2,def,3000,d$f,bonus#500,city2

hive (complexDatatypes)> create table MapType(id int,name string,sal bigint,sub array<string>,dud map<string,int>,city string)

> row format delimited

> fields terminated by ','

> collection items terminated by '$'

> map keys terminated by '#';

OK

Time taken: 0.054 seconds

hive (complexDatatypes)> describe MapType;

OK

col\_name data\_type comment

id int

name string

sal bigint

sub array<string>

dud map<string,int>

city string

Time taken: 0.053 seconds, Fetched: 6 row(s)

hive (complexDatatypes)> select \* from MapType;

OK

maptype.id maptype.name maptype.sal maptype.sub maptype.dud maptype.city

1 abc 40000 ["a","b","c"] {"bonus":500,"insurance":200} city1

2 def 3000 ["d","f"] {"bonus":500} city2

Time taken: 0.038 seconds, Fetched: 2 row(s)

hive (complexDatatypes)> select \* from MapType where dud["bonus"]>0;

OK

maptype.id maptype.name maptype.sal maptype.sub maptype.dud maptype.city

1 abc 40000 ["a","b","c"] {"bonus":500,"insurance":200} city1

2 def 3000 ["d","f"] {"bonus":500} city2

Time taken: 0.047 seconds, Fetched: 2 row(s)

hive (complexDatatypes)> select \* from MapType where dud["insurance"] is not null;

OK

maptype.id maptype.name maptype.sal maptype.sub maptype.dud maptype.city

1 abc 40000 ["a","b","c"] {"bonus":500,"insurance":200} city1

Time taken: 0.082 seconds, Fetched: 1 row(s)

hive (complexDatatypes)> select sum(dud["bonus"]) from MapType where dud["bonus"] is not null;

Query ID = hduser\_20171224120826\_99c6449f-fe3e-43d2-a925-4e2bdfc59aaa

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\_1514094803278\_0001, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0001/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0001

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

2017-12-24 12:08:33,127 Stage-1 map = 0%, reduce = 0%

2017-12-24 12:08:37,269 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.83 sec

2017-12-24 12:08:42,403 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.74 sec

MapReduce Total cumulative CPU time: 3 seconds 740 msec

Ended Job = job\_1514094803278\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.74 sec HDFS Read: 7715 HDFS Write: 5 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 740 msec

OK

\_c0

1000

Time taken: 16.927 seconds, Fetched: 1 row(s)

Struct File:

1,abc,40000,a$b$c,bonus#500$insurance#200,city1$state1$111

2,def,3000,d$f,bonus#500,city2$state2$222

hive (complexDatatypes)> create table StructType(id int,name string,sal bigint,sub array<string>,dud map<string,int>,addr struct<city:string,state:string,pin:bigint>)

> row format delimited

> fields terminated by ','

> collection items terminated by '$'

> map keys terminated by '#';

OK

Time taken: 0.048 seconds

hive (complexDatatypes)> describe StructType;

OK

col\_name data\_type comment

id int

name string

sal bigint

sub array<string>

dud map<string,int>

addr struct<city:string,state:string,pin:bigint>

Time taken: 0.035 seconds, Fetched: 6 row(s)

hive (complexDatatypes)> load data local inpath '/home/hduser/structfile' into table StructType;

Loading data to table complexdatatypes.structtype

Table complexdatatypes.structtype stats: [numFiles=1, totalSize=101]

OK

Time taken: 0.104 seconds

hive (complexDatatypes)> select \* from StructType;

OK

structtype.id structtype.name structtype.sal structtype.sub structtype.dud structtype.addr

1 abc 40000 ["a","b","c"] {"bonus":500,"insurance":200} {"city":"city1","state":"state1","pin":111}

2 def 3000 ["d","f"] {"bonus":500} {"city":"city2","state":"state2","pin":222}

Time taken: 0.045 seconds, Fetched: 2 row(s)

hive (complexDatatypes)> select addr.city, addr.state, addr.pin from StructType;

OK

city state pin

city1 state1 111

city2 state2 222

Time taken: 0.035 seconds, Fetched: 2 row(s)

hive (complexDatatypes)> select \* from StructType where addr.pin=222;

OK

structtype.id structtype.name structtype.sal structtype.sub structtype.dud structtype.addr

2 def 3000 ["d","f"] {"bonus":500} {"city":"city2","state":"state2","pin":222}

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

find sales done in each payment mode and their percentage

hive (niit)> create table totalsales (total bigint)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.029 seconds

hive (niit)> insert overwrite table totalsales

> select sum(amount) from txnrecords;

Query ID = hduser\_20171224122612\_faea6bbc-9ca3-496e-9bce-1a5015c0674f

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\_1514094803278\_0002, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0002/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0002

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

2017-12-24 12:26:15,577 Stage-1 map = 0%, reduce = 0%

2017-12-24 12:26:20,736 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.0 sec

2017-12-24 12:26:24,859 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.51 sec

MapReduce Total cumulative CPU time: 4 seconds 510 msec

Ended Job = job\_1514094803278\_0002

Loading data to table niit.totalsales

Table niit.totalsales stats: [numFiles=1, numRows=1, totalSize=8, rawDataSize=7]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.51 sec HDFS Read: 4426300 HDFS Write: 79 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 510 msec

OK

\_col0

Time taken: 15.036 seconds

hive (niit)> select \* from totalsales;

OK

totalsales.total

5110820

Time taken: 0.032 seconds, Fetched: 1 row(s)

hive (niit)> select a.spendby, round(sum(a.amount),2) as typesales, round((sum(a.amount)/total\*100),2) as salespercent from txnrecords a, totalsales b group by a.spendby, b.total ;

Warning: Map Join MAPJOIN[13][bigTable=a] in task 'Stage-2:MAPRED' is a cross product

Query ID = hduser\_20171224122720\_e91f5c0e-1be3-430f-9753-381f9701c75d

Total jobs = 1

17/12/24 12:27:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Execution log at: /tmp/hduser/hduser\_20171224122720\_e91f5c0e-1be3-430f-9753-381f9701c75d.log

2017-12-24 12:27:22 Starting to launch local task to process map join; maximum memory = 477626368

2017-12-24 12:27:22 Dump the side-table for tag: 1 with group count: 1 into file: file:/usr/local/hive/iotmp/7e6d14e2-7dd3-4b49-bed7-0af9a715daa6/hive\_2017-12-24\_12-27-20\_306\_2056072787579906996-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile01--.hashtable

2017-12-24 12:27:22 Uploaded 1 File to: file:/usr/local/hive/iotmp/7e6d14e2-7dd3-4b49-bed7-0af9a715daa6/hive\_2017-12-24\_12-27-20\_306\_2056072787579906996-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile01--.hashtable (281 bytes)

2017-12-24 12:27:22 End of local task; Time Taken: 0.605 sec.

Execution completed successfully

MapredLocal task succeeded

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\_1514094803278\_0003, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0003/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0003

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

2017-12-24 12:27:28,152 Stage-2 map = 0%, reduce = 0%

2017-12-24 12:27:32,247 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.6 sec

2017-12-24 12:27:37,382 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.3 sec

MapReduce Total cumulative CPU time: 5 seconds 300 msec

Ended Job = job\_1514094803278\_0003

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.3 sec HDFS Read: 4432119 HDFS Write: 44 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 300 msec

OK

a.spendby typesales salespercent

cash 187685.61 3.67

credit 4923134.93 96.33

Time taken: 18.137 seconds, Fetched: 2 row(s)

.find sales based on age group with the % on totalsales

hduser@rootuser:~$ hadoop fs -put /home/hduser/custs.txt /user/hive/warehouse/customer;

17/12/24 12:40:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

hduser@rootuser:~$ hadoop fs -put /home/hduser/txns1.txt /user/hive/warehouse/customer;

17/12/24 12:41:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

hive (niit)> create table out1 (custno int,firstname string,age int,profession string,amount double,product string)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.026 seconds

hive (niit)> insert overwrite table out1

> select a.custno,a.firstname,a.age,a.profession,b.amount,b.product

> from customer a JOIN txnrecords b ON a.custno = b.custno;

Query ID = hduser\_20171224125107\_a2c55901-f73b-42b6-a7b5-8de6736c58ee

Total jobs = 1

17/12/24 12:51:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Execution log at: /tmp/hduser/hduser\_20171224125107\_a2c55901-f73b-42b6-a7b5-8de6736c58ee.log

2017-12-24 12:51:09 Starting to launch local task to process map join; maximum memory = 477626368

2017-12-24 12:51:10 Dump the side-table for tag: 0 with group count: 9999 into file: file:/usr/local/hive/iotmp/8c68e5a2-4bec-422a-a153-d6eca81e2e2c/hive\_2017-12-24\_12-51-07\_222\_5005940508844468911-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile00--.hashtable

2017-12-24 12:51:10 Uploaded 1 File to: file:/usr/local/hive/iotmp/8c68e5a2-4bec-422a-a153-d6eca81e2e2c/hive\_2017-12-24\_12-51-07\_222\_5005940508844468911-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile00--.hashtable (433919 bytes)

2017-12-24 12:51:10 End of local task; Time Taken: 0.837 sec.

Execution completed successfully

MapredLocal task succeeded

Launching Job 1 out of 1

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job\_1514094803278\_0007, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0007/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0007

Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0

2017-12-24 12:51:15,000 Stage-4 map = 0%, reduce = 0%

2017-12-24 12:51:20,177 Stage-4 map = 100%, reduce = 0%, Cumulative CPU 5.02 sec

MapReduce Total cumulative CPU time: 5 seconds 20 msec

Ended Job = job\_1514094803278\_0007

Loading data to table niit.out1

Table niit.out1 stats: [numFiles=1, numRows=49995, totalSize=2532933, rawDataSize=2482938]

MapReduce Jobs Launched:

Stage-Stage-4: Map: 1 Cumulative CPU: 5.02 sec HDFS Read: 4425788 HDFS Write: 2533006 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 20 msec

OK

a.custno a.firstname a.age a.profession b.amount b.product

Time taken: 14.238 seconds

hive (niit)> select \* from out1 limit 5;

OK

out1.custno out1.firstname out1.age out1.profession out1.amount out1.product

4007024 Cameron 59 Actor 40.33 Cardio Machine Accessories

4006742 Gregory 36 Accountant 198.44 Weightlifting Gloves

4009775 Ruby 44 Designer 5.58 Weightlifting Machine Accessories

4002199 Keith 44 Police officer 198.19 Gymnastics Rings

4002613 Hugh 43 Engineering technician 98.81 Field Hockey

Time taken: 0.034 seconds, Fetched: 5 row(s)

hive (niit)> create table out2 (custno int,firstname string,age int,profession string,amount double,product string, level string)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.098 seconds

hive (niit)> insert overwrite table out2

> select \* , case when age<30 then 'low' when age>=30 and age < 50 then 'middle' when age>=50 then 'old'

> else 'others' end

> from out1;

Query ID = hduser\_20171224125546\_815f7b89-47ca-486d-80f0-3dcbeb335098

Total jobs = 3

Launching Job 1 out of 3

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job\_1514094803278\_0008, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0008/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0008

Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0

2017-12-24 12:55:50,176 Stage-1 map = 0%, reduce = 0%

2017-12-24 12:55:55,333 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.93 sec

MapReduce Total cumulative CPU time: 3 seconds 930 msec

Ended Job = job\_1514094803278\_0008

Stage-4 is selected by condition resolver.

Stage-3 is filtered out by condition resolver.

Stage-5 is filtered out by condition resolver.

Moving data to: hdfs://localhost:54310/user/hive/warehouse/niit.db/out2/.hive-staging\_hive\_2017-12-24\_12-55-46\_280\_4736573903430112973-1/-ext-10000

Loading data to table niit.out2

Table niit.out2 stats: [numFiles=1, numRows=49995, totalSize=2787462, rawDataSize=2737467]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 3.93 sec HDFS Read: 2537696 HDFS Write: 2787535 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 930 msec

OK

out1.custno out1.firstname out1.age out1.profession out1.amount out1.product \_c1

Time taken: 10.243 seconds

hive (niit)> describe out2;

OK

col\_name data\_type comment

custno int

firstname string

age int

profession string

amount double

product string

level string

Time taken: 0.053 seconds, Fetched: 7 row(s)

hive (niit)> select \* from out2 limit 5;

OK

out2.custno out2.firstname out2.age out2.profession out2.amount out2.product out2.level

4007024 Cameron 59 Actor 40.33 Cardio Machine Accessories old

4006742 Gregory 36 Accountant 198.44 Weightlifting Gloves middle

4009775 Ruby 44 Designer 5.58 Weightlifting Machine Accessories middle

4002199 Keith 44 Police officer 198.19 Gymnastics Rings middle

4002613 Hugh 43 Engineering technician 98.81 Field Hockey middle

Time taken: 0.032 seconds, Fetched: 5 row(s)

hive (niit)> create table out3 (level string, amount double, salespercent double)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.032 seconds

hive (niit)> describe out3;

OK

col\_name data\_type comment

level string

amount double

salespercent double

Time taken: 0.034 seconds, Fetched: 3 row(s)

hive (niit)> insert overwrite table out3

> select a.level, round(sum(a.amount),2) as totalspent, round((sum(a.amount)/total\*100),2) as salespercent from out2 a, totalsales b group by a.level, b.total;

Warning: Map Join MAPJOIN[13][bigTable=a] in task 'Stage-2:MAPRED' is a cross product

Query ID = hduser\_20171224125742\_59caf64e-dc8e-45d0-91b1-6124df69e286

Total jobs = 1

17/12/24 12:57:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Execution log at: /tmp/hduser/hduser\_20171224125742\_59caf64e-dc8e-45d0-91b1-6124df69e286.log

2017-12-24 12:57:44 Starting to launch local task to process map join; maximum memory = 477626368

2017-12-24 12:57:45 Dump the side-table for tag: 1 with group count: 1 into file: file:/usr/local/hive/iotmp/8c68e5a2-4bec-422a-a153-d6eca81e2e2c/hive\_2017-12-24\_12-57-42\_983\_2131670345597502795-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile11--.hashtable

2017-12-24 12:57:45 Uploaded 1 File to: file:/usr/local/hive/iotmp/8c68e5a2-4bec-422a-a153-d6eca81e2e2c/hive\_2017-12-24\_12-57-42\_983\_2131670345597502795-1/-local-10003/HashTable-Stage-2/MapJoin-mapfile11--.hashtable (281 bytes)

2017-12-24 12:57:45 End of local task; Time Taken: 0.564 sec.

Execution completed successfully

MapredLocal task succeeded

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\_1514094803278\_0009, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0009/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0009

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

2017-12-24 12:57:49,063 Stage-2 map = 0%, reduce = 0%

2017-12-24 12:57:54,199 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.72 sec

2017-12-24 12:57:58,309 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 5.42 sec

MapReduce Total cumulative CPU time: 5 seconds 420 msec

Ended Job = job\_1514094803278\_0009

Loading data to table niit.out3

Table niit.out3 stats: [numFiles=1, numRows=3, totalSize=65, rawDataSize=62]

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.42 sec HDFS Read: 2801407 HDFS Write: 130 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 420 msec

OK

a.level totalspent salespercent

Time taken: 17.53 seconds

hive (niit)> select \* from out3;

OK

out3.level out3.amount out3.salespercent

low 725221.34 14.19

middle 1855861.67 36.31

old 2529100.31 49.49

Time taken: 0.03 seconds, Fetched: 3 row(s)

find sales based on profession with % on totalsales.

hive (niit)> create table out4(profession string, amount double, salespercent double) row format delimited fields terminated by ',';

OK

Time taken: 0.029 seconds

hive (niit)> insert overwrite table out4

> select a.profession, round(sum(a.amount),2) as totalspent, round((sum(a.amount)/total\*100),2) as salespercent from out1 a, totalsales b group by a.profession, b.total order by salespercent desc;

Warning: Map Join MAPJOIN[17][bigTable=a] in task 'Stage-2:MAPRED' is a cross product

Query ID = hduser\_20171224130442\_44443f84-6589-4643-9dee-35c98fc395ab

Total jobs = 2

17/12/24 13:04:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Execution log at: /tmp/hduser/hduser\_20171224130442\_44443f84-6589-4643-9dee-35c98fc395ab.log

2017-12-24 13:04:44 Starting to launch local task to process map join; maximum memory = 477626368

2017-12-24 13:04:44 Dump the side-table for tag: 1 with group count: 1 into file: file:/usr/local/hive/iotmp/8c68e5a2-4bec-422a-a153-d6eca81e2e2c/hive\_2017-12-24\_13-04-42\_846\_2650433760880178383-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile21--.hashtable

2017-12-24 13:04:44 Uploaded 1 File to: file:/usr/local/hive/iotmp/8c68e5a2-4bec-422a-a153-d6eca81e2e2c/hive\_2017-12-24\_13-04-42\_846\_2650433760880178383-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile21--.hashtable (281 bytes)

2017-12-24 13:04:44 End of local task; Time Taken: 0.531 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\_1514094803278\_0010, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0010/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0010

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

2017-12-24 13:04:51,601 Stage-2 map = 0%, reduce = 0%

2017-12-24 13:04:54,749 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.07 sec

2017-12-24 13:04:59,874 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.47 sec

MapReduce Total cumulative CPU time: 4 seconds 470 msec

Ended Job = job\_1514094803278\_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\_1514094803278\_0011, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0011/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0011

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

2017-12-24 13:05:09,732 Stage-3 map = 0%, reduce = 0%

2017-12-24 13:05:13,839 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 0.77 sec

2017-12-24 13:05:18,945 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 2.27 sec

MapReduce Total cumulative CPU time: 2 seconds 270 msec

Ended Job = job\_1514094803278\_0011

Loading data to table niit.out4

Table niit.out4 stats: [numFiles=1, numRows=51, totalSize=1417, rawDataSize=1366]

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.47 sec HDFS Read: 2545933 HDFS Write: 2482 SUCCESS

Stage-Stage-3: Map: 1 Reduce: 1 Cumulative CPU: 2.27 sec HDFS Read: 7513 HDFS Write: 1484 SUCCESS

Total MapReduce CPU Time Spent: 6 seconds 740 msec

OK

a.profession totalspent salespercent

Time taken: 37.213 seconds

hive (niit)> select \* from out4 limit 10;

OK

out4.profession out4.amount out4.salespercent

Firefighter 116516.99 2.28

Politician 114633.27 2.24

Librarian 114152.24 2.23

Computer support specialist 114138.49 2.23

Pilot 112438.32 2.2

Human resources assistant 112682.97 2.2

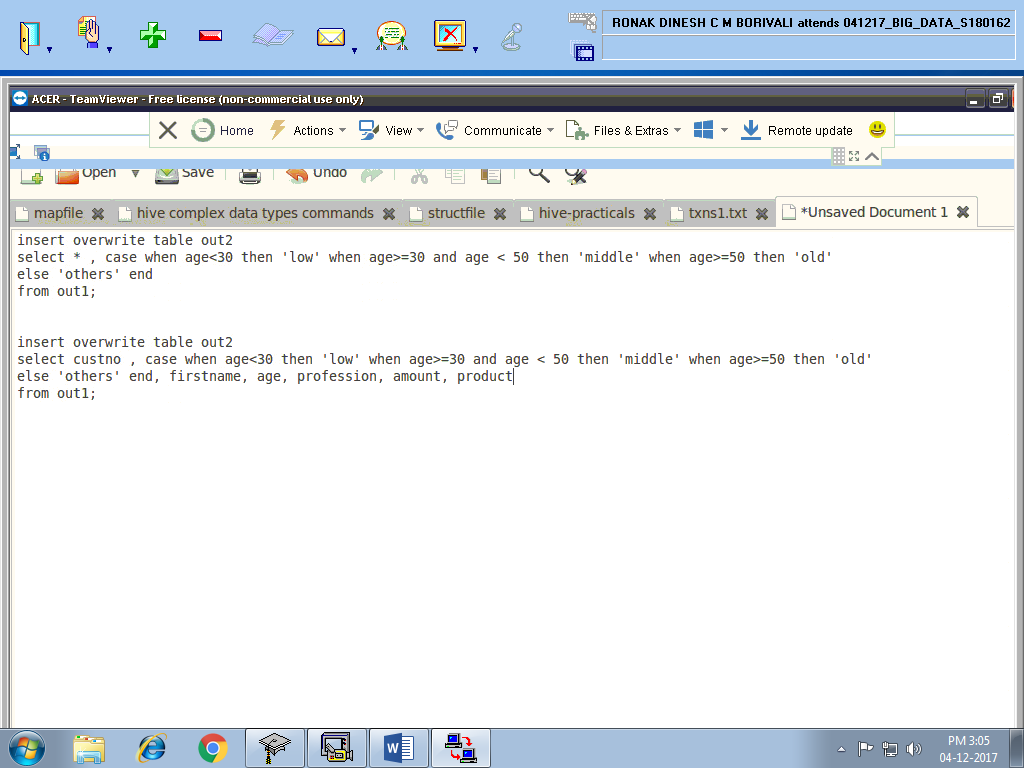
Photographer 111930.18 2.19

Police officer 111577.42 2.18

Loan officer 110836.27 2.17

Actor 109969.35 2.15

Time taken: 0.028 seconds, Fetched: 10 row(s)



create an index on customer (earlier created) table on profession column

\*\*\* deferred rebuild will create an empty index

hive (niit)> create index prof\_index on table customer(profession) as 'compact' with deferred rebuild;

OK

Time taken: 1.022 seconds

\*\*\*\* alter index will actually create the index

hive (niit)> alter index prof\_index on customer rebuild;

Query ID = hduser\_20171224130915\_1b54918e-3178-4957-b3d5-d6585ee059df

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\_1514094803278\_0012, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0012/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0012

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

2017-12-24 13:09:19,031 Stage-1 map = 0%, reduce = 0%

2017-12-24 13:09:23,131 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.43 sec

2017-12-24 13:09:27,217 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.32 sec

MapReduce Total cumulative CPU time: 3 seconds 320 msec

Ended Job = job\_1514094803278\_0012

Loading data to table niit.niit\_\_customer\_prof\_index\_\_

Table niit.niit\_\_customer\_prof\_index\_\_ stats: [numFiles=1, numRows=51, totalSize=71410, rawDataSize=71359]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.32 sec HDFS Read: 401159 HDFS Write: 71501 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 320 msec

OK

Time taken: 12.833 seconds

hive (niit)> show tables;

OK

tab\_name

afe\_45plus

avro\_table

csv\_table

customer

customerext

niit\_\_customer\_prof\_index\_\_

nyse

out1

out2

out3

out4

totalsales

txnrecords

txnrecsbycat

txnrecsbycat2

txnrecsbycat3

txnrecsbycat4

txnrecsbycat5

Time taken: 0.014 seconds, Fetched: 18 row(s)

\*\*\*\*\*\*list all the indexes on the table

hive (niit)> show indexes on customer;

OK

idx\_name tab\_name col\_names idx\_tab\_name idx\_type comment

prof\_index customer profession niit\_\_customer\_prof\_index\_\_ compact

Time taken: 0.039 seconds, Fetched: 1 row(s)

hive (niit)> select \* from niit\_\_customer\_prof\_index\_\_ where profession="Pilot";

OK

niit\_\_customer\_prof\_index\_\_.profession niit\_\_customer\_prof\_index\_\_.\_bucketnameniit\_\_customer\_prof\_index\_\_.\_offsets

Pilot hdfs://localhost:54310/user/hive/warehouse/niit.db/customer/custs.txt [0,226,9039,9069,9899,10636,12622,12817,13175,15369,15781,16556,16622,17202,18176,21290,22070,23542,23725,25611,28227,30350,34908,35647,36051,37505,38424,39263,40593,42720,42846,48997,52647,55760,55924,58345,58521,64934,67802,70966,73891,75310,76054,77318,78587,80628,80915,81759,82529,83108,86404,88673,92535,92940,95094,96906,96938,98418,100202,103224,106317,109501,112655,113018,113721,114165,117257,118361,118767,118882,122929,127432,131288,132859,133241,133439,139869,141929,143186,144137,147578,148767,150490,156016,156157,156485,156556,157288,158740,163765,168659,169361,169814,169884,170588,170909,173748,175576,175843,181504,181935,185686,187671,194506,194611,195402,196010,196276,197784,201228,209060,210663,211853,212360,217935,219591,221334,222276,225341,225907,226184,230427,230921,230949,231075,232661,233689,236726,238319,242454,242708,242776,244873,245308,249616,251345,251727,254101,254332,257422,257453,259837,261478,262981,265880,266880,266995,267971,268179,269954,269985,270653,271258,272525,280009,280668,281176,283972,284553,287232,289657,290712,291039,291729,292174,292339,292717,296187,297350,297739,298266,300367,301179,305516,305940,307931,309615,310745,312149,313130,317550,327075,329552,332754,333996,336863,338143,341073,345573,350628,350952,354543,359367,359399,361395,364641,364868,366842,366942,368286,371203,372556,375041,377627,378693,379357,382757,384064,387807,389026,389946]

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

\*\*\*\*\*schema of the index

hive (niit)> describe niit\_\_customer\_prof\_index\_\_;

OK

col\_name data\_type comment

profession string

\_bucketname string

\_offsets array<bigint>

Time taken: 0.039 seconds, Fetched: 3 row(s)

\*\*\*\*Time taken without index

hive (niit)> select profession, count(\*) from customer group by profession;

Query ID = hduser\_20171224131746\_bae93020-e04a-4fd6-b034-a78e7d6c81de

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\_1514094803278\_0013, Tracking URL = http://rootuser:8088/proxy/application\_1514094803278\_0013/

Kill Command = /usr/local/hadoop/bin/hadoop job -kill job\_1514094803278\_0013

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

2017-12-24 13:17:50,318 Stage-1 map = 0%, reduce = 0%

2017-12-24 13:17:54,408 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.45 sec

2017-12-24 13:17:59,497 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.74 sec

MapReduce Total cumulative CPU time: 2 seconds 740 msec

Ended Job = job\_1514094803278\_0013

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.74 sec HDFS Read: 399270 HDFS Write: 885 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 740 msec

OK

profession \_c1

Accountant 199

Actor 202

Agricultural and food scientist 195

Architect 203

Artist 175

Athlete 196

Automotive mechanic 193

Carpenter 181

Chemist 209

Childcare worker 207

Civil engineer 193

Coach 201

Computer hardware engineer 204

Computer software engineer 216

Computer support specialist 222

Dancer 185

Designer 205

Doctor 197

Economist 189

Electrical engineer 192

Electrician 194

Engineering technician 204

Environmental scientist 176

Farmer 201

Financial analyst 198

Firefighter 217

Human resources assistant 212

Judge 196

Lawyer 212

Librarian 218

Loan officer 221

Musician 205

Nurse 192

Pharmacist 213

Photographer 222

Physicist 201

Pilot 211

Police officer 210

Politician 228

Psychologist 194

Real estate agent 191

Recreation and fitness worker 210

Reporter 200

Secretary 200

Social Worker 1

Social worker 212

Statistician 196

Teacher 204

Therapist 187

Veterinarian 208

Writer 101

Time taken: 13.608 seconds, Fetched: 51 row(s)

\*\*\*\*Time taken with index

hive (niit)> select profession, SIZE(`\_offsets`) from niit\_\_customer\_prof\_index\_\_;

OK

profession \_c1

Accountant 199

Actor 202

Agricultural and food scientist 195

Architect 203

Artist 175

Athlete 196

Automotive mechanic 193

Carpenter 181

Chemist 209

Childcare worker 207

Civil engineer 193

Coach 201

Computer hardware engineer 204

Computer software engineer 216

Computer support specialist 222

Dancer 185

Designer 205

Doctor 197

Economist 189

Electrical engineer 192

Electrician 194

Engineering technician 204

Environmental scientist 176

Farmer 201

Financial analyst 198

Firefighter 217

Human resources assistant 212

Judge 196

Lawyer 212

Librarian 218

Loan officer 221

Musician 205

Nurse 192

Pharmacist 213

Photographer 222

Physicist 201

Pilot 211

Police officer 210

Politician 228

Psychologist 194

Real estate agent 191

Recreation and fitness worker 210

Reporter 200

Secretary 200

Social Worker 1

Social worker 212

Statistician 196

Teacher 204

Therapist 187

Veterinarian 208

Writer 101

Time taken: 0.039 seconds, Fetched: 51 row(s)