hive> create table employee(id int,name string,salary int,gender string)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.089 seconds

hive> load data local inpath '/home/cloudera/dataemployee' into table employee;

Loading data to table tron.employee

Table tron.employee stats: [numFiles=1, totalSize=166]

OK

Time taken: 0.536 seconds

hive> select \* from employee;

OK

1 abhi 90000 m

2 prasad 90000 m

3 ram 89000 m

4 parth 91000 m

5 ravi 8900 m

6 shaam 89767 m

7 shivani 89000 f

8 kalyani 95000 f

10 jagruti 91000 f

11 jigisha 89000 f

Time taken: 0.08 seconds, Fetched: 10 row(s)

hive> select sum(salary) from employee;

Query ID = cloudera\_20220511073737\_3b0df593-0ae3-4e38-88d9-178cfda288ec

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\_1652265459133\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652265459133\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652265459133\_0003

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

2022-05-11 07:38:04,210 Stage-1 map = 0%, reduce = 0%

2022-05-11 07:38:11,646 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.76 sec

2022-05-11 07:38:19,097 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.84 sec

MapReduce Total cumulative CPU time: 3 seconds 840 msec

Ended Job = job\_1652265459133\_0003

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.84 sec HDFS Read: 6783 HDFS Write: 7 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 840 msec

OK

Aggregate function use for all data sum salary

**822667**

Time taken: 23.17 seconds, Fetched: 1 row(s)

hive> select gender,sum(salary) from employee

> group by gender;

Query ID = cloudera\_20220511073939\_3999cde4-b855-4575-8b3a-7ba07223a526

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\_1652265459133\_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652265459133\_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652265459133\_0004

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

2022-05-11 07:39:15,037 Stage-1 map = 0%, reduce = 0%

2022-05-11 07:39:22,485 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.07 sec

2022-05-11 07:39:29,957 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.14 sec

MapReduce Total cumulative CPU time: 5 seconds 140 msec

Ended Job = job\_1652265459133\_0004

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.14 sec HDFS Read: 7080 HDFS Write: 18 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 140 msec

OK

**f 364000**

Output of group by based on aggregate function of sum use in above query

**m 458667**

Time taken: 23.105 seconds, Fetched: 2 row(s)

hive> select salart,gender,sum(salary),min(salary),max(salary),count(\*) from employee

> group by salary,gender;

FAILED: SemanticException [Error 10025]: Line 1:7 Expression not in GROUP BY key 'salart'

hive> select salary,gender,sum(salary),min(salary),max(salary),count(\*) from employee

> group by salary,gender;

Query ID = cloudera\_20220511075252\_15c854c1-310f-4d0b-86db-cf274eb0cbff

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\_1652265459133\_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652265459133\_0006/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652265459133\_0006

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

2022-05-11 07:52:59,286 Stage-1 map = 0%, reduce = 0%

2022-05-11 07:53:05,628 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.45 sec

2022-05-11 07:53:13,072 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.48 sec

MapReduce Total cumulative CPU time: 3 seconds 480 msec

Ended Job = job\_1652265459133\_0006

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.48 sec HDFS Read: 8755 HDFS Write: 222 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 480 msec

OK

**8900 m 8900 8900 8900 1**

**89000 f 178000 89000 89000 2**

**89000 m 89000 89000 89000 1**

Group by multiple column aggregation function

**89767 m 89767 89767 89767 1**

**90000 m 180000 90000 90000 2**

**91000 f 91000 91000 91000 1**

**91000 m 91000 91000 91000 1**

**95000 f 95000 95000 95000 1**

Time taken: 21.889 seconds, Fetched: 8 row(s)

hive> select id,salary,gender,sum(salary),min(salary),max(salary),count(\*) from employee

> group by id,salary,gender;

Query ID = cloudera\_20220511075353\_b01e59d1-a1e6-4ec4-bb9f-072042e1aead

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\_1652265459133\_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652265459133\_0007/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652265459133\_0007

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

2022-05-11 07:53:46,198 Stage-1 map = 0%, reduce = 0%

2022-05-11 07:53:53,659 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.67 sec

2022-05-11 07:54:01,119 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.72 sec

MapReduce Total cumulative CPU time: 5 seconds 720 msec

Ended Job = job\_1652265459133\_0007

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.72 sec HDFS Read: 9049 HDFS Write: 298 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 720 msec

OK

1 90000 m 90000 90000 90000 1

2 90000 m 90000 90000 90000 1

3 89000 m 89000 89000 89000 1

4 91000 m 91000 91000 91000 1

5 8900 m 8900 8900 8900 1

6 89767 m 89767 89767 89767 1

7 89000 f 89000 89000 89000 1

8 95000 f 95000 95000 95000 1

10 91000 f 91000 91000 91000 1

11 89000 f 89000 89000 89000 1

Time taken: 23.336 seconds, Fetched: 10 row(s)

hive> select \* from sample;

OK

ravi 10

ram 30

pratik 20

ravi 10

ravi 10

pratik 20

ravi 10

ram 30

ram 30

pratik 20

Time taken: 0.123 seconds, Fetched: 10 row(s)

hive> describe formatted sample;

OK

# col\_name data\_type comment

name string

deptno int

# Detailed Table Information

Database: tron

Owner: cloudera

CreateTime: Sun May 15 06:50:29 PDT 2022

LastAccessTime: UNKNOWN

Protect Mode: None

Retention: 0

Location: hdfs://quickstart.cloudera:8020/user/hive/warehouse/tron.db/sample

Table Type: MANAGED\_TABLE

Table Parameters:

COLUMN\_STATS\_ACCURATE true

numFiles 1

totalSize 83

transient\_lastDdlTime 1652622657

# Storage Information

SerDe Library: org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe

InputFormat: org.apache.hadoop.mapred.TextInputFormat

OutputFormat: org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat

Compressed: No

Num Buckets: -1

Bucket Columns: []

Sort Columns: []

Storage Desc Params:

field.delim ,

serialization.format ,

Time taken: 0.181 seconds, Fetched: 31 row(s)

**hive> select name,deptno,count(\*) from sample**

**> group by name,deptno;**

Query ID = cloudera\_20220516060808\_257e3023-1c29-4383-b767-664cfc86e23c

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\_1652702731581\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652702731581\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652702731581\_0001

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

2022-05-16 06:08:47,446 Stage-1 map = 0%, reduce = 0%

2022-05-16 06:08:57,942 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.33 sec

2022-05-16 06:09:07,779 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.06 sec

MapReduce Total cumulative CPU time: 5 seconds 60 msec

Ended Job = job\_1652702731581\_0001

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.06 sec HDFS Read: 7027 HDFS Write: 31 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 60 msec

OK

**pratik 20 3**

**ram 30 3**

**ravi 10 4**

Time taken: 44.009 seconds, Fetched: 3 row(s)

**PARTITION**

101,aaa,4000,m,11

102,bbb,5000,f,12

103,ccc,6000,m,12

200,ddd,4000,f,13

201,eee,5000,m,12

202,fff,3000,f,11

203,ggg,4000,m,13

300,dfd,4500,f,11

301,gfd,5000,m,12

302,rtr,6000,f,13

hive> create table employee1(id int,name string,salary int,gender string,deptno int)

> row format delimited

> fields terminated by ',';

OK

Time taken: 0.118 seconds

hive> load data local inpath '/home/cloudera/empl' into table employee1;

Loading data to table tron.employee1

Table tron.employee1 stats: [numFiles=1, totalSize=180]

OK

Time taken: 0.763 seconds

**hive> select \* from employee1;**

**OK**

**101 aaa 4000 m 11**

**102 bbb 5000 f 12**

**103 ccc 6000 m 12**

**200 ddd 4000 f 13**

**201 eee 5000 m 12**

**202 fff 3000 f 11**

**203 ggg 4000 m 13**

**300 dfd 4500 f 11**

**301 gfd 5000 m 12**

**302 rtr 6000 f 13**

Time taken: 0.11 seconds, Fetched: 10 row(s)

hive**> create table empl1(id int,name string,salary int,gender string,deptno int)**

**> partitioned by (d int);**

OK

Time taken: 0.112 seconds

**hive> insert overwrite table empl**

**> partition (g='m')**

**> select \* from employee1 where gender='m';**

Query ID = cloudera\_20220516200606\_56d58668-9062-4048-b4f5-e3b117acd2e6

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\_1652755780647\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652755780647\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652755780647\_0001

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

2022-05-16 20:06:28,476 Stage-1 map = 0%, reduce = 0%

2022-05-16 20:06:37,227 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.14 sec

MapReduce Total cumulative CPU time: 3 seconds 140 msec

Ended Job = job\_1652755780647\_0001

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://quickstart.cloudera:8020/user/hive/warehouse/tron.db/empl/g=m/.hive-staging\_hive\_2022-05-16\_20-06-10\_144\_2119529557564872322-1/-ext-10000

Loading data to table tron.empl partition (g=m)

Partition tron.empl{g=m} stats: [numFiles=1, numRows=5, totalSize=90, rawDataSize=85]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 3.14 sec HDFS Read: 4098 HDFS Write: 159 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 140 msec

OK

Time taken: 31.463 seconds

**hive> select \* from empl;**

**OK**

**102 bbb 5000 f 12 f**

**200 ddd 4000 f 13 f**

**202 fff 3000 f 11 f**

**300 dfd 4500 f 11 f**

**302 rtr 6000 f 13 f**

**101 aaa 4000 m 11 m**

**103 ccc 6000 m 12 m**

**201 eee 5000 m 12 m**

**203 ggg 4000 m 13 m**

**301 gfd 5000 m 12 m**

**Time taken: 0.598 seconds, Fetched: 10 row(s)**

**hive> insert overwrite table empl1**

**> partition (d=11)**

**> select \* from employee1 where deptno=11;**

Query ID = cloudera\_20220516132929\_8f848d17-94d6-44f5-9b40-f985a93dc736

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\_1652730857821\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652730857821\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652730857821\_0003

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

2022-05-16 13:29:55,586 Stage-1 map = 0%, reduce = 0%

2022-05-16 13:30:05,378 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.91 sec

MapReduce Total cumulative CPU time: 3 seconds 910 msec

Ended Job = job\_1652730857821\_0003

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://quickstart.cloudera:8020/user/hive/warehouse/tron.db/empl1/d=11/.hive-staging\_hive\_2022-05-16\_13-29-45\_156\_6746455195203672047-1/-ext-10000

Loading data to table tron.empl1 partition (d=11)

Partition tron.empl1{d=11} stats: [numFiles=1, numRows=3, totalSize=54, rawDataSize=51]

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Cumulative CPU: 3.91 sec HDFS Read: 4120 HDFS Write: 125 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 910 msec

OK

Time taken: 22.703 seconds

hive> select \* from empl1;

OK

**101 aaa 4000 m 11 11**

**202 fff 3000 f 11 11**

**300 dfd 4500 f 11 11**

Time taken: 0.127 seconds, Fetched: 3 row(s)

**[cloudera@quickstart ~]$ hdfs dfs -cat /user/hive/warehouse/tron.db/empl/g=f/00000\***

**102bbb5000f12**

**200ddd4000f13**

**202fff3000f11**

**300dfd4500f11**

**302rtr6000f13**

[cloudera@quickstart ~]$ hdfs dfs -cat /user/hive/warehouse/tron.db/empl1

cat: `/user/hive/warehouse/tron.db/empl1': Is a directory

[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/warehouse/tron.db/empl1

Found 1 items

drwxrwxrwx - cloudera hive 0 2022-05-16 13:30 /user/hive/warehouse/tron.db/empl1/d=11

[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/warehouse/tron.db/empl1/d=11

Found 1 items

-rwxrwxrwx 1 cloudera hive 54 2022-05-16 13:30 /user/hive/warehouse/tron.db/empl1/d=11/000000\_0

^[[A[cloudera@quickstart ~]$ hdfs dfs -cat /user/hive/warehouse/tron.db/empl1/d=11/00000\*

22/05/16 19:54:39 WARN fs.FileSystem: "-cat" is a deprecated filesystem name. Use "hdfs://-cat/" instead.

/user/hive/warehouse/tron.db/empl1/d=11/00000\*: Unknown command

**[cloudera@quickstart ~]$ hdfs dfs -cat /user/hive/warehouse/tron.db/empl1/d=11/00000\***

**101aaa4000m11**

**202fff3000f11**

**300dfd4500f11**

**[cloudera@quickstart ~]$ hdfs dfs -cat /user/hive/warehouse/tron.db/empl/g='m'/00000\***

**101aaa4000m11**

**103ccc6000m12**

**201eee5000m12**

**203ggg4000m13**

**301gfd5000m12**

**51 SUCCESS**

**Total MapReduce CPU Time Spent: 2 seconds 270 msec**

**OK**

**Time taken: 35.901 seconds**

**hive> create table dypart(id int,name string,salary int,gender string,deptno int)**

**> partitioned by (d int,g string)**

**> row format delimited**

**> fields terminated by ',';**

**OK**

**Time taken: 1.585 seconds**

**hive> set hive.exec.dynamic.partition=true;**

**hive> set hive.exec.dynamic.partition.mode=nonstrict;**

**hive> insert overwrite table dypart**

**> partition (d,g)**

**> select id,name,salary,gender,deptno,deptno,gender from employee1;**

**Query ID = cloudera\_20220517132727\_7d26a8cf-a3cc-4333-9aad-40eb3874372e**

**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\_1652813413258\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1652813413258\_0001/**

**Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1652813413258\_0001**

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

**2022-05-17 13:27:24,036 Stage-1 map = 0%, reduce = 0%**

**2022-05-17 13:27:33,933 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.27 sec**

**MapReduce Total cumulative CPU time: 2 seconds 270 msec**

**Ended Job = job\_1652813413258\_0001**

**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://quickstart.cloudera:8020/user/hive/warehouse/tron.db/dypart/.hive-staging\_hive\_2022-05-17\_13-27-05\_615\_5333113919566778840-1/-ext-10000**

**Loading data to table tron.dypart partition (d=null, g=null)**

**Time taken for load dynamic partitions : 2853**

**Loading partition {d=12, g=f}**

**Loading partition {d=13, g=f}**

**Loading partition {d=11, g=m}**

**Loading partition {d=11, g=f}**

**Loading partition {d=13, g=m}**

**Loading partition {d=12, g=m}**

**Time taken for adding to write entity : 4**

**Partition tron.dypart{d=11, g=f} stats: [numFiles=1, numRows=2, totalSize=36, rawDataSize=34]**

**Partition tron.dypart{d=11, g=m} stats: [numFiles=1, numRows=1, totalSize=18, rawDataSize=17]**

**Partition tron.dypart{d=12, g=f} stats: [numFiles=1, numRows=1, totalSize=18, rawDataSize=17]**

**Partition tron.dypart{d=12, g=m} stats: [numFiles=1, numRows=3, totalSize=54, rawDataSize=51]**

**Partition tron.dypart{d=13, g=f} stats: [numFiles=1, numRows=2, totalSize=36, rawDataSize=34]**

**Partition tron.dypart{d=13, g=m} stats: [numFiles=1, numRows=1, totalSize=18, rawDataSize=17]**

**MapReduce Jobs Launched:**

**Stage-Stage-1: Map: 1 Cumulative CPU: 2.27 sec HDFS Read: 4060 HDFS Write: 451 SUCCESS**

**Total MapReduce CPU Time Spent: 2 seconds 270 msec**

**OK**

**Time taken: 35.901 seconds**

**hive> select\* from dypart;**

**OK**

**202 fff 3000 f 11 11 f**

**300 dfd 4500 f 11 11 f**

**101 aaa 4000 m 11 11 m**

**102 bbb 5000 f 12 12 f**

**103 ccc 6000 m 12 12 m**

**201 eee 5000 m 12 12 m**

**301 gfd 5000 m 12 12 m**

**200 ddd 4000 f 13 13 f**

**302 rtr 6000 f 13 13 f**

**203 ggg 4000 m 13 13 m**

**Time taken: 0.267 seconds, Fetched: 10 row(s)**