



MUMBAI EDUCATIONAL TRUST

MET INSTITUTE OF COMPUTER SCIENCE



Program Number	
Roll Number	1546
Title of program	Hive Commands
Program	Hive Commands

start hive

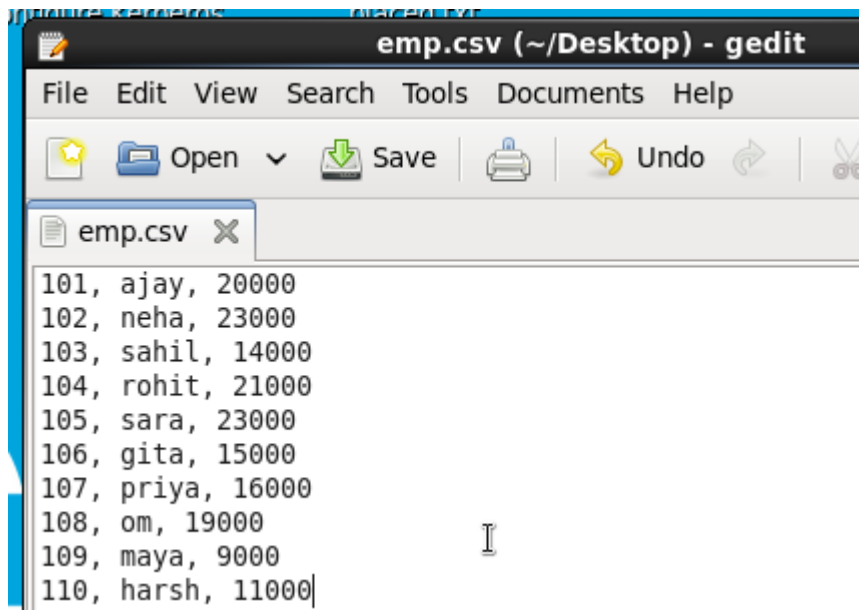
```
[cloudera@quickstart ~]$ hive
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create database EmployeeDB;
OK
Time taken: 5.322 seconds

hive> create database EmployeeDB;
OK
Time taken: 5.322 seconds
hive> use EmployeeDB;
OK
Time taken: 0.168 seconds
hive> set hive.cli.print.current.db= true;
```

describe a table

```
hive (EmployeeDB)> describe emp;
OK
empid          int
name           string
salary         float
Time taken: 0.925 seconds, Fetched: 3 row(s)
```

create a csv file and load the data from hdfs/local to table



```
hive (EmployeeDB)> load data local inpath '/home/cloudera/Desktop/emp.csv' into table emp;
Loading data to table employeeedb.emp
Table employeeedb.emp stats: [numFiles=1, totalSize=171]
OK
Time taken: 2.913 seconds
```

```
hive (EmployeeDB)> select * from emp
> ;
OK
101      ajay      20000.0
102      neha      23000.0
103      sahil      14000.0
104      rohit      21000.0
105      sara      23000.0
106      gita      15000.0
107      priya      16000.0
108      om        19000.0
109      maya      9000.0
110      harsh      11000.0
Time taken: 3.735 seconds, Fetched: 10 row(s)
```

display all records ascending order of salary

order by command is converted into mapreduce

```

hive (EmployeeDB)> select * from emp order by salary;
Query ID = cloudera_20250811010909_9b00a530-f566-474d-ac53-7dbd8b9717c8
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_1754892275457_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1754892275457_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1754892275457_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-08-11 01:11:28,074 Stage-1 map = 0%, reduce = 0%
2025-08-11 01:12:28,268 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.25 sec
2025-08-11 01:13:07,120 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.98 sec
MapReduce Total cumulative CPU time: 5 seconds 150 msec
Ended Job = job_1754892275457_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.15 sec HDFS Read: 7336 HDFS Write: 181 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 150 msec
OK
109      maya      9000.0
110      harsh     11000.0
103      sahil      14000.0
106      gita       15000.0
107      priya      16000.0
108      om         19000.0
101      ajay       20000.0
104      rohit      21000.0
105      sara       23000.0
102      neha       23000.0
Time taken: 195.13 seconds, Fetched: 10 row(s)

```

Activate
Go to Settings

descending order

```

hive (EmployeeDB)> select * from emp order by salary desc;
Query ID = cloudera_20250811011515_389f33f9-de41-4669-a314-09b79496494a
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_1754892275457_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1754892275457_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1754892275457_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-08-11 01:15:54,878 Stage-1 map = 0%, reduce = 0%
2025-08-11 01:16:34,679 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.2 sec
2025-08-11 01:17:18,138 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.11 sec
MapReduce Total cumulative CPU time: 5 seconds 370 msec
Ended Job = job_1754892275457_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.37 sec HDFS Read: 7399 HDFS Write: 181 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 370 msec
OK
105      sara       23000.0
102      neha       23000.0
104      rohit      21000.0
101      ajay       20000.0
108      om         19000.0
107      priya      16000.0
106      gita       15000.0
103      sahil      14000.0
110      harsh     11000.0
109      maya       9000.0
Time taken: 132.562 seconds, Fetched: 10 row(s)

```

Activate
Go to Settings

display maximum , minimum and average salary

```

hive (EmployeeDB)> select max(salary) as Maxsal , min(salary) as MinSal, avg(salary) as AvgSal from emp;
Query ID = cloudera_20250811012525_bde24be4-a03b-4988-9f0f-5405b35d172b
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_1754892275457_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1754892275457_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1754892275457_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-08-11 01:27:06,506 Stage-1 map = 0%, reduce = 0%
2025-08-11 01:27:50,617 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.39 sec
2025-08-11 01:28:23,762 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.86 sec
MapReduce Total cumulative CPU time: 4 seconds 860 msec
Ended Job = job_1754892275457_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.86 sec HDFS Read: 9453 HDFS Write: 23 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 860 msec
OK
23000.0 9000.0 17100.0
Time taken: 190.343 seconds, Fetched: 1 row(s)

```

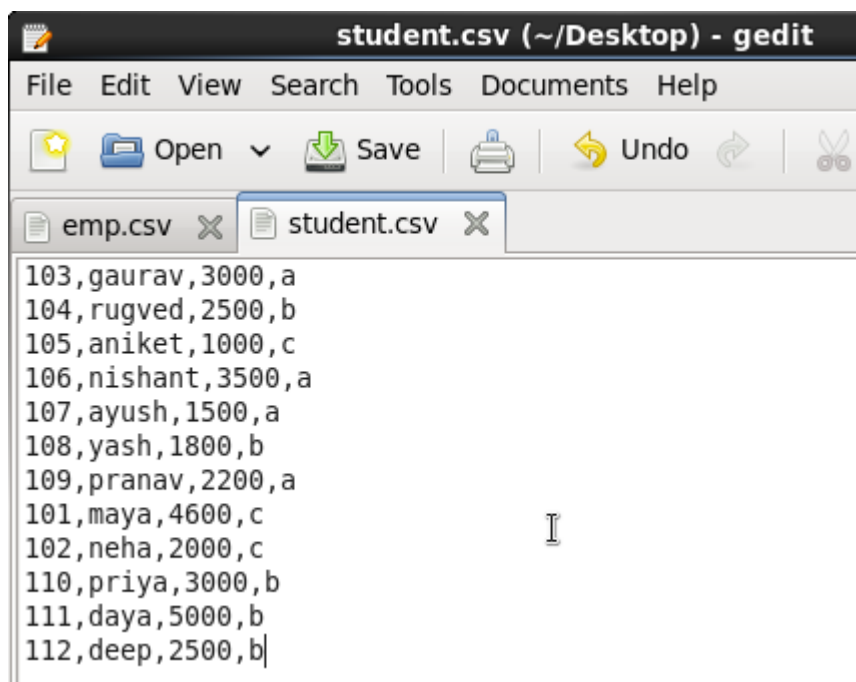
Activate
Go to Settings

create table and student.csv

```

hive (EmployeeDB)> create table students(sid int, sname string, salary float, division string) row format delimited by ',';
OK
Time taken: 0.382 seconds
hive (EmployeeDB)> load data local inpath '/home/cloudera/Desktop/student.csv' into table students;
FAILED: ParseException line 1:10 extraneous input 'local' expecting INPATH near '<EOF>'
hive (EmployeeDB)> load data local inpath '/home/cloudera/Desktop/student.csv' into table students;
Loading data to table employeeedb.students
Table employeeedb.students stats: [numFiles=1, totalSize=205]
OK
Time taken: 1.124 seconds
hive (EmployeeDB)> select * from students
> ;
OK
103    gaurav    3000.0    a
104    rugved    2500.0    b
105    aniket    1000.0    c
106    nishant    3500.0    a
107    ayush     1500.0    a
108    yash      1800.0    b
109    pranav    2200.0    a
101    maya      4600.0    c
102    neha      2000.0    c
110    priya     3000.0    b
111    daya      5000.0    b
112    deep      2500.0    b
Time taken: 0.318 seconds, Fetched: 12 row(s)

```



find count of students in each division

```
hive (EmployeeDB)> select division, count(*) from students group by division;
Query ID = cloudera_20250811023939_12c0ec9c-6de2-4d18-93e3-06776695f462
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_1754892275457_0007, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1754892275457_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1754892275457_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-08-11 02:39:51,521 Stage-1 map = 0%, reduce = 0%
2025-08-11 02:40:22,321 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.86 sec
2025-08-11 02:40:56,846 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.51 sec
MapReduce Total cumulative CPU time: 4 seconds 510 msec
Ended Job = job_1754892275457_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.51 sec HDFS Read: 8265 HDFS Write: 12 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 510 msec
OK
a      4
b      5
c      3
Time taken: 108.262 seconds. Fetched: 3 row(s)
```

Activate

find standard deviations

```

hive (EmployeeDB)> select stddev_samp(salary) from students;
Query ID = cloudera_20250811023434_7dddaac0-cla3-4256-8b88-9a4f6cf02419
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_1754892275457_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1754892275457_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1754892275457_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-08-11 02:35:28,109 Stage-1 map = 0%, reduce = 0%
2025-08-11 02:35:58,306 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.89 sec
2025-08-11 02:36:32,414 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.45 sec
MapReduce Total cumulative CPU time: 4 seconds 450 msec
Ended Job = job_1754892275457_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.45 sec HDFS Read: 8519 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 450 msec
OK
1193.7970996578588
Time taken: 104.092 seconds, Fetched: 1 row(s)

```

```

hive (EmployeeDB)> select stddev_pop(salary) from students;
Query ID = cloudera_20250811014646_64ac3c6d-cf96-49b7-898b-82522043f98b
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_1754892275457_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1754892275457_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1754892275457_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-08-11 01:46:56,826 Stage-1 map = 0%, reduce = 0%
2025-08-11 01:47:27,551 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.89 sec
2025-08-11 01:48:02,124 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.39 sec
MapReduce Total cumulative CPU time: 4 seconds 390 msec
Ended Job = job_1754892275457_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.39 sec HDFS Read: 8427 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 390 msec
OK
1142.9737043733285
Time taken: 110.689 seconds, Fetched: 1 row(s)

```

HIVE PARTITIONS

creating a partitioning table

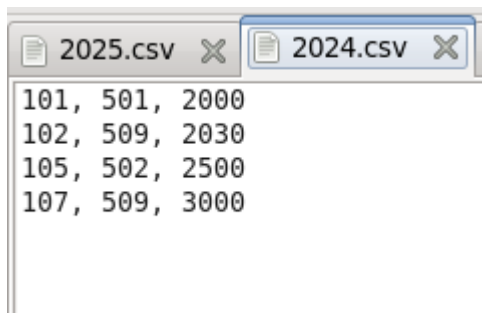
```

hive (EmployeeDB)> create table sales_data(pid int, cid int, amt double)partitioned by (year int)row format delimited fields
terminated by ',';
OK
Time taken: 0.422 seconds
hive (EmployeeDB)> describe sales_data;
OK
pid                int
cid                int
amt                double
year               int

# Partition Information
# col_name          data_type          comment
year                int
Time taken: 2.575 seconds, Fetched: 9 row(s)

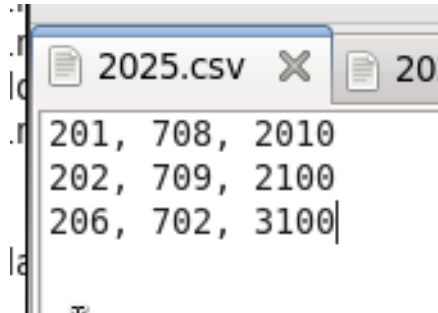
```

static partition



2025.csv x 2024.csv x

101,	501,	2000
102,	509,	2030
105,	502,	2500
107,	509,	3000



2025.csv x 2024.csv x

201,	708,	2010
202,	709,	2100
206,	702,	3100

```
hive (EmployeeDB)> load data local inpath '/home/cloudera/Desktop/2024.csv' into table sales_data partition(year=2024);
Loading data to table employeeedb.sales_data partition (year=2024)
Partition employeeedb.sales_data{year=2024} stats: [numFiles=1, numRows=0, totalSize=60, rawDataSize=0]
OK
Time taken: 2.394 seconds
hive (EmployeeDB)> load data local inpath '/home/cloudera/Desktop/2025.csv' into table sales_data partition(year=2025);
Loading data to table employeeedb.sales_data partition (year=2025)
Partition employeeedb.sales_data{year=2025} stats: [numFiles=1, numRows=0, totalSize=45, rawDataSize=0]
OK
Time taken: 1.853 seconds
hive (EmployeeDB)> select * from salaes_data;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'salaes_data'
hive (EmployeeDB)> select * from sales_data;
OK
101      NULL      2000.0  2024
102      NULL      2030.0  2024
105      NULL      2500.0  2024
107      NULL      3000.0  2024
201      NULL      2010.0  2025
202      NULL      2100.0  2025
206      NULL      3100.0  2025
Time taken: 0.54 seconds. Fetched: 7 row(s)
```

Activ
Go to

dynamic partitioning

```
hive (EmployeeDB)> describe students;
OK
sid              int
sname            string
salary          float
division        string
Time taken: 43.344 seconds, Fetched: 4 row(s)

hive (EmployeeDB)> create table student_part(sid int, sname string, salary float) partitioned by (division string) row format
delimited fields terminated by ',';
OK
Time taken: 0.272 seconds
hive (EmployeeDB)> set hive.exec.dynamic.partition.mode=nonstrict;
hive (EmployeeDB)> set hive.exec.dynamic.partition=true;
```

```

hive (EmployeeDB)> insert into table student_part partition(division) select sid,sname,salary,division from students;
Query ID = cloudera_20250811022424_28c67875-fbc1-4656-a5c6-d818ec772668
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_1754892275457_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1754892275457_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1754892275457_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2025-08-11 02:25:32,852 Stage-1 map = 0%, reduce = 0%
2025-08-11 02:26:16,058 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.45 sec
MapReduce Total cumulative CPU time: 2 seconds 450 msec
Ended Job = job_1754892275457_0005
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/employeeedb.db/student_part/.hive-staging_hive_2025-08-11_
02-24-50_039_2734795974958186603-1/-ext-10000
Loading data to table employeeedb.student_part partition (division=null)
    Time taken for load dynamic partitions : 1100
    Loading partition {division=c}
    Loading partition {division=b}
    Loading partition {division=a}
    Time taken for adding to write entity : 29
Partition employeeedb.student_part{division=a} stats: [numFiles=1, numRows=4, totalSize=72, rawDataSize=68]
Partition employeeedb.student_part{division=b} stats: [numFiles=1, numRows=5, totalSize=83, rawDataSize=78]
Partition employeeedb.student_part{division=c} stats: [numFiles=1, numRows=3, totalSize=50, rawDataSize=47]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 2.45 sec HDFS Read: 4678 HDFS Write: 401 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 450 msec
OK
Time taken: 92.303 seconds

```

Activate W

```

hive (EmployeeDB)> show partitions student_part;
OK
division=a
division=b
division=c
Time taken: 0.432 seconds, Fetched: 3 row(s)

```