

**1. Create database 'company' located at /user/hive/mywarehouse/.  
 Create database if not exists company  
 Create database company with dbproperties("transaction"="true");**

```

hive> create database if not exists company
> comment 'this is the data base for assignment 2'
> location '/user/hive/mywarehouse/'
> with dbproperties ("transaction"="true")
> ;
OK
Time taken: 0.16 seconds
hive> show databases;
OK
company
default
retail_db
Time taken: 0.025 seconds, Fetched: 3 row(s)
hive>
  
```

**2. Create a hive managed table partitioned table 'employee' in 'company' database which can store the following data:  
 Create table employee (empid int , name string, designation string, salary int)  
 Partitioned by (dept char(1))  
 Row format delimited  
 Fields terminated by ','  
 Lines terminated by '\n'  
 Stored as textfile;**

**Alter table employee  
 Add partition (dept='A')  
 Partition (dept='B')  
 Partition (dept='C')**

**Show partitions employee;**

```

hive> describe employee
> ;
OK
empid      int
name       string
designation string
salary    int
dept       char(1)
# Partition Information
# col_name  data_type  comment
dept       char(1)
Time taken: 0.115 seconds, Fetched: 9 row(s)
hive>
  
```

**3. Save the above data in a text file on local file system and HDFS. Load data into 'employee' table from both LOCAL & HDFS filesystem.  
 Load data local inpath 'emp\_a.txt' into table employee  
 Partition(dept='A');**

Load data local inpath 'emp\_b.txt' into table employee  
 Partition(dept='B');  
 Load data local inpath 'emp\_c.txt' into table employee  
 Partition(dept='C');  
 set hive.exec.dynamic.partition=true;  
 set hive.exec.dynamic.partition.mode=nonstrict;  
 for dynamic portioning

```

Activities  Terminal  Jan 12 12:11
osboxes@osboxes: -
name          string
designation    string
salary        int
dept          char(1)

# Partition Information
# col_name      data_type      comment
dept           char(1)
Time taken: 0.115 seconds, Fetched: 9 row(s)
hive> load data local inpath 'emp_A.txt' into table employee
> partition(dept='A');
FAILED: ParseException line 1:22 extraneous input '=' expecting StringLiteral near '<EOF>'
hive> load data local inpath 'emp_A.txt' into table employee
> partition(dept='A');
Loading data to table company.employee partition (dept=A)
OK
Time taken: 1.967 seconds
hive> load data local inpath 'emp_B.txt' into table employee
> partition(dept='B');
Loading data to table company.employee partition (dept=B)
OK
Time taken: 0.219 seconds
hive> load data local inpath 'emp_C.txt' into table employee
> partition(dept='C');
Loading data to table company.employee partition (dept=C)
OK
Time taken: 0.607 seconds
hive> select * from employee;
OK
1  Anne  AdmIn  50000  A
3  Janet  Sales  60000  A
6  Margaret  Tech  12000  A
9  Nancy  AdmIn  50000  A
10 Andrew  Manager 40000  A
13 Robert  Manager 40000  A
14 Laura  Engineer 45000  A
2  Gokul  AdmIn  50000  B
7  Nirmal  Tech  12000  B
8  Jinju  Engineer 45000  B
11 Arun  Manager 40000  B
12 Harish  Sales  60000  B
15 Anju  CEO  100000  B
4  Harl  AdmIn  50000  C
5  Sanker  AdmIn  50000  C
Time taken: 1.217 seconds, Fetched: 15 row(s)
hive>

```

4. Add the following records to the text file and load the modified data into 'employee' table using OVERWRITE.  
 Load data local inpath 'employee\_more' into table employee;

```

Activities  Terminal  Jan 12 12:17
osboxes@osboxes: -
Starting Job = job_1610469059330_0001, Tracking URL = http://osboxes:8088/proxy/application_1610469059330_0001/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610469059330_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 12:17:00,577 Stage-1 map = 0%, reduce = 0%
2021-01-12 12:17:07,100 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.31 sec
2021-01-12 12:17:11,354 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.51 sec
MapReduce Total cumulative CPU time: 2 seconds 510 msec
Ended Job = job_1610469059330_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 directory hdfs://localhost:9000/user/hive/warehouse/employee/.hive-staging_hive_2021-01-12_12-16-49_423_74844758266568657-1/-ext-10000
Loading data to table company.employee partition (dept=null)

Time taken to load dynamic partitions: 0.306 seconds
Time taken for adding to write entity : 0.002 seconds
MapReduce Jobs Launched:
Stage:Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.51 sec HDFS Read: 17720 HDFS Write: 746 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 510 msec
OK
Time taken: 25.103 seconds
hive> select * from employee;
OK
19 Steven  Engineer  45000  A
20 Michael  CEO  100000  A
1  Anne  AdmIn  50000  A
3  Janet  Sales  60000  A
6  Margaret  Tech  12000  A
9  Nancy  AdmIn  50000  A
10 Andrew  Manager 40000  A
13 Robert  Manager 40000  A
14 Laura  Engineer 45000  A
16 Aarathi  Manager 40000  B
17 Parvathi  Engineer  45000  B
18 Gopika  AdmIn  50000  B
2  Gokul  AdmIn  50000  B
7  Nirmal  Tech  12000  B
8  Jinju  Engineer 45000  B
11 Arun  Manager 40000  B
12 Harish  Sales  60000  B
15 Anju  CEO  100000  B
4  Harl  AdmIn  50000  C
5  Sanker  AdmIn  50000  C
Time taken: 0.31 seconds, Fetched: 20 row(s)
hive>

```

5. Create another table 'new\_employees' with the following records:  
 Create table new\_employee (empid int , name string, designation string, salary int, dept char(1))  
 Row format delimited

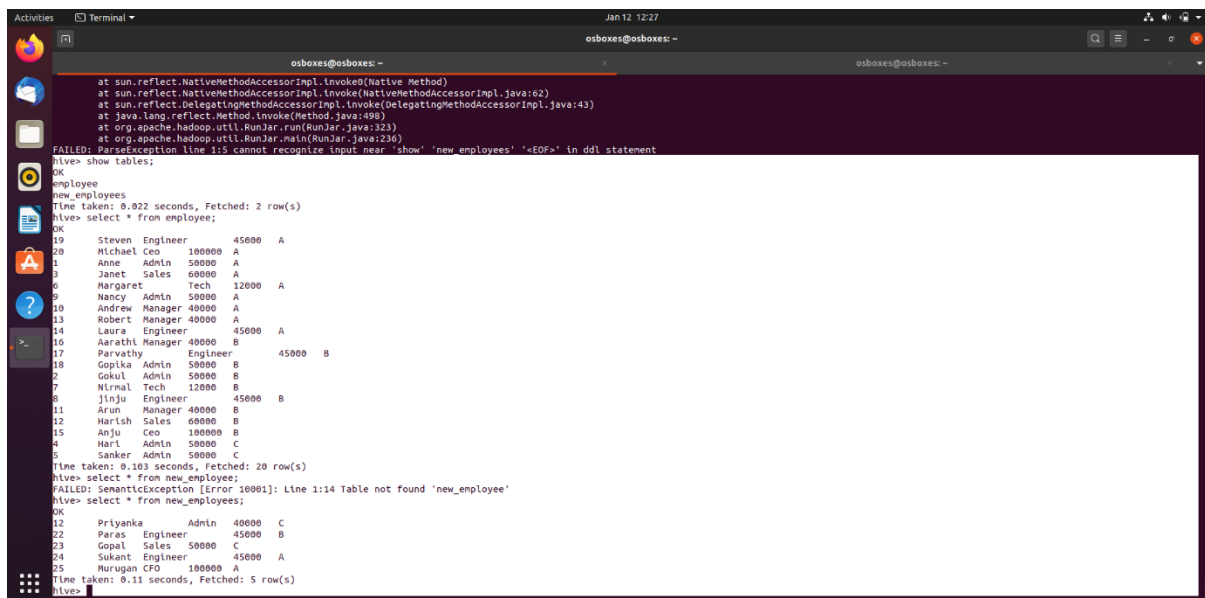
Fields terminated by ','  
Lines terminated by '\n'  
Stored as textfile;

Insert into table employee values (12,'Priyanka','Admin',40000,'C');  
Inserting single row

Insert into table employee values (12,'Priyanka','Admin',40000,'C'),(  
22,'Paras','Engineer',45000,'B'),(23,'Gopal','Sales',50000,'C'),(  
24,'Sukant','Engineer',45000,'A'),( 25,'Murugan','CFO',100000,'A');  
Inserting multi rows

Or load data from file  
Load data local inpath 'employee\_more' into table employee;

From new\_employee  
Insert overwrite table employee  
Partition (dept)  
Select empid , name, designation, salary, dept;  
#requires dynamic partitioning



```
at sun.reflect.NativeMethodAccessorImpl.invoke(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:498)
at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
FAILED: ParseException line 1:5 cannot recognize input near 'show' 'new_employees' '<EOF>' in ddl statement
hive> show tables;
OK
employee
new_employees
Time taken: 0.022 seconds, Fetched: 2 row(s)
hive> select * from employee;
OK
19 Steven Engineer 45000 A
20 Michael Ceo 100000 A
1 Anne Admin 50000 A
3 Janet Sales 60000 A
6 Margaret Tech 12000 A
9 Nancy Admin 50000 A
10 Andrew Manager 40000 A
13 Robert Manager 40000 A
14 Laura Engineer 45000 A
16 Aarathi Manager 40000 B
17 Parvathy Engineer 45000 B
18 Gopika Admin 50000 B
2 Gokul Admin 50000 B
7 Nirral Tech 12000 B
8 Jinju Engineer 45000 B
11 Arun Manager 40000 B
12 Harish Sales 60000 B
15 Anju Ceo 100000 B
4 Harl Admin 50000 C
5 Sanker Admin 50000 C
Time taken: 0.103 seconds, Fetched: 20 row(s)
hive> select * from new_employee;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'new_employee'
hive> select * from new_employees;
OK
12 Priyanka Admin 40000 C
22 Paras Engineer 45000 B
23 Gopal Sales 50000 C
24 Sukant Engineer 45000 A
25 Murugan CFO 100000 A
Time taken: 0.11 seconds, Fetched: 5 row(s)
hive>
```

6. List the employees having salary>50000  
Select \* from employee where salary>50000;

```
htive> select * from employee where salary > 50000;
OK
20 Michael Ceo 100000 A
25 Murugan CFO 100000 A
3 Janet Sales 60000 A
12 Harish Sales 60000 B
15 Anju Ceo 100000 B
Time taken: 0.357 seconds, Fetched: 5 row(s)
htive>
```

**7. Select the list of employees whose names start from ‘A’.**  
**Select name from employee where name like ‘A%’;**

```
htive> select * from employee where salary > 50000;
OK
20 Michael Ceo 100000 A
25 Murugan CFO 100000 A
3 Janet Sales 60000 A
12 Harish Sales 60000 B
15 Anju Ceo 100000 B
Time taken: 0.357 seconds, Fetched: 5 row(s)
htive> select name from employee where name like 'A%'
> ;
OK
Anne
Andrew
Aarathi
Arun
Anju
Time taken: 0.108 seconds, Fetched: 5 row(s)
htive>
```

**8. List the number of employees for each Designation.**  
**Select designation . count(\*) from employee group by dept**  
**OR**

Select \* from employee where dept in ('A','B','C');

```
Activities Terminal Jan 12 12:50
osboxes@osboxes: -
OK
20 Michael Ceo 100000 A
25 Murugan CFO 100000 A
3 Janet Sales 60000 A
12 Harish Sales 60000 B
15 Anju Ceo 100000 B
Time taken: 0.357 seconds, Fetched: 5 row(s)
hive> select name from employee where name like 'A%'
> ;
OK
Anne
Andrew
Aarathi
Arun
Anju
Time taken: 0.108 seconds, Fetched: 5 row(s)
hive> select * from employee where dept in ('A','B','C')
> ;
OK
19 Steven Engineer 45000 A
20 Michael Ceo 100000 A
24 Sukant Engineer 45000 A
25 Murugan CFO 100000 A
1 Anne Admin 50000 A
3 Janet Sales 60000 A
6 Margaret Tech 12000 A
9 Nancy Admin 50000 A
10 Andrew Manager 40000 A
13 Robert Manager 40000 A
14 Laura Engineer 45000 A
16 Aarathi Manager 40000 B
17 Parvathy Engineer 45000 B
18 Gopika Admin 50000 B
22 Paras Engineer 45000 B
2 Gokul Admin 50000 B
7 Nirmla Tech 12000 B
8 Jinju Engineer 45000 B
11 Arun Manager 40000 B
12 Harish Sales 60000 B
15 Anju Ceo 100000 B
12 Priyanka Admin 40000 C
23 Gopal Sales 50000 C
4 Hari Admin 50000 C
5 Sanker Admin 50000 C
Time taken: 0.139 seconds, Fetched: 25 row(s)
hive>
```

9. Order the list of employees according to their names.  
Select name from employee order by name;

```
Activities Terminal Jan 12 12:52
osboxes@osboxes: -
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.reducers=<number>
Starting Job = Job_1610473164347_0002, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0002/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 12:51:44,946 Stage-1 map = 0%, reduce = 0%
2021-01-12 12:51:54,744 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.28 sec
MapReduce Total cumulative CPU time: 2 seconds 520 msec
Ended Job = Job_1610473164347_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.52 sec HDFS Read: 12289 HDFS Write: 556 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 520 msec
OK
Aarathi
Andrew
Anju
Anne
Arun
Gokul
Gopal
Gopika
Hari
Harish
Janet
Laura
Margaret
Michael
Murugan
Nancy
Nirmla
Paras
Parvathy
Priyanka
Robert
Sanker
Steven
Sukant
Jinju
Time taken: 19.725 seconds, Fetched: 25 row(s)
hive>
```

10. Partition the employees table based upon the salary.  
Select \* from employee order by salary desc;

```
Activities Terminal Jan 12 12:55
osboxes@osboxes: -
osboxes@osboxes: -
osboxes@osboxes: -
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_1610473164347_0004, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0004/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 12:54:52,502 Stage-1 map = 0%, reduce = 0%
2021-01-12 12:54:56,593 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.58 sec
2021-01-12 12:55:02,704 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.02 sec
MapReduce Total cumulative CPU time: 2 seconds 820 msec
Ended Job = job_1610473164347_0004
MapReduce Jobs Launched:
  Stage:Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.02 sec HDFS Read: 13629 HDFS Write: 993 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 820 msec
OK
20 Michael Ceo 100000 A
15 Anju Ceo 100000 B
25 Murugan CFO 100000 A
3 Janet Sales 60000 A
12 Harish Sales 60000 B
5 Sanker Admin 50000 C
9 Nancy Admin 50000 A
4 Harl Admin 50000 C
1 Anne Admin 50000 A
2 Gokul Admin 50000 B
23 Gopal Sales 50000 C
18 Gopika Admin 50000 B
17 Parvathy Engineer 45000 B
8 Jinju Engineer 45000 B
22 Paras Engineer 45000 B
14 Laura Engineer 45000 A
24 Sukant Engineer 45000 A
19 Steven Engineer 45000 A
16 Aarathi Manager 40000 B
11 Arun Manager 40000 B
12 Priyanka Admin 40000 C
10 Andrew Manager 40000 A
13 Robert Manager 40000 A
6 Margaret Tech 12000 A
7 Nirmal Tech 12000 B
Time taken: 17.78 seconds, Fetched: 25 row(s)
hive>
```

11. List to top 5 highly paid employees.  
Select \* from employee order by salary desc limit 5;

```
Activities Terminal Jan 12 12:57
osboxes@osboxes: -
osboxes@osboxes: -
osboxes@osboxes: -
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_1610473164347_0004, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0004/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 12:54:52,502 Stage-1 map = 0%, reduce = 0%
2021-01-12 12:54:56,593 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.58 sec
2021-01-12 12:55:02,704 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.02 sec
MapReduce Total cumulative CPU time: 2 seconds 820 msec
Ended Job = job_1610473164347_0004
MapReduce Jobs Launched:
  Stage:Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.02 sec HDFS Read: 13629 HDFS Write: 993 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 820 msec
OK
20 Michael Ceo 100000 A
15 Anju Ceo 100000 B
25 Murugan CFO 100000 A
3 Janet Sales 60000 A
12 Harish Sales 60000 B
5 Sanker Admin 50000 C
9 Nancy Admin 50000 A
4 Harl Admin 50000 C
1 Anne Admin 50000 A
2 Gokul Admin 50000 B
23 Gopal Sales 50000 C
18 Gopika Admin 50000 B
17 Parvathy Engineer 45000 B
8 Jinju Engineer 45000 B
22 Paras Engineer 45000 B
14 Laura Engineer 45000 A
24 Sukant Engineer 45000 A
19 Steven Engineer 45000 A
16 Aarathi Manager 40000 B
11 Arun Manager 40000 B
12 Priyanka Admin 40000 C
10 Andrew Manager 40000 A
13 Robert Manager 40000 A
6 Margaret Tech 12000 A
7 Nirmal Tech 12000 B
Time taken: 17.78 seconds, Fetched: 25 row(s)
hive>
```

13. Compute the average salary paid by the company to its employees  
select avg(salary),count(\*) from employee;

```
Activities Terminal Jan 12 13:10
osboxes@osboxes: ~
osboxes@osboxes: ~
osboxes@osboxes: ~
hive> select avg(salary),count(*) from emp;
FAILED: SemanticException [Error 10001]: line 1:33 'table not found 'emp''
hive> select avg(salary),count(*) from employee;
Query ID = osboxes_20210112130933_07907de7-9368-45c7-bede-926a51c40f97
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_1610473164347_0006, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0006/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:09:38,722 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:09:42,826 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.81 sec
2021-01-12 13:09:48,937 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.83 sec
MapReduce Total cumulative CPU time: 3 seconds 830 msec
Ended Job = job_1610473164347_0006
MapReduce Jobs Launched:
Stage:Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.83 sec HDFS Read: 17739 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 830 msec
OK
50560.0 25
Time taken: 16.52 seconds, Fetched: 1 row(s)
hive>
```

**14. Compute the total salary paid by the company per month.**  
**select sum(salary),count(\*) from employee;**

```
Activities Terminal Jan 12 13:11
osboxes@osboxes: ~
osboxes@osboxes: ~
osboxes@osboxes: ~
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_1610473164347_0006, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0006/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:09:38,722 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:09:42,826 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.81 sec
2021-01-12 13:09:48,937 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.83 sec
MapReduce Total cumulative CPU time: 3 seconds 830 msec
Ended Job = job_1610473164347_0006
MapReduce Jobs Launched:
Stage:Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.83 sec HDFS Read: 17739 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 830 msec
OK
50560.0 25
Time taken: 16.52 seconds, Fetched: 1 row(s)
hive> select sum(salary),count(*) from employee;
Query ID = osboxes_20210112131125_ba728dcf-d865-4634-962a-c724ea973b94
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_1610473164347_0007, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0007/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:11:32,448 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:11:36,545 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.59 sec
2021-01-12 13:11:42,018 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.86 sec
MapReduce Total cumulative CPU time: 2 seconds 860 msec
Ended Job = job_1610473164347_0007
MapReduce Jobs Launched:
Stage:Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.86 sec HDFS Read: 16338 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 860 msec
OK
1264000 25
Time taken: 18.774 seconds, Fetched: 1 row(s)
hive>
```

**15. List the employee with the highest salary. List the employee with the minimum salary.**  
**select max(salary), min(salary) from employee;**  
**OR**

Select \* from employee where salary in (select min(salary)from employee);  
Select \* from employee where salary in (select max(salary)from employee);

```
Activities Terminal Jan 12 13:13
osboxes@osboxes: -
osboxes@osboxes: -
osboxes@osboxes: -

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.reducers=<number>
Starting Job = Job_1610473164347_0007, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0007/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0007
Hadoop Job Information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:11:32,440 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:11:36,545 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.59 sec
2021-01-12 13:11:42,010 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.80 sec
MapReduce Total cumulative CPU time: 2 seconds 800 msec
Ended Job = Job_1610473164347_0007
MapReduce Jobs Launched:
Stage:Stage-1 Map: 1 Reduce: 1 Cumulative CPU: 2.80 sec HDFS Read: 16338 HDFS Write: 110 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 800 msec
OK
1264000 25
Time taken: 18.774 seconds, Fetched: 1 row(s)
hive> select max(salary), min(salary) from employee;
FAILED: ParseException line 1:31 missing ')' at 'from' near '<EOF>'
hive> select max(salary), min(salary) from employee;
Query ID = osboxes_20210112131319_b58f0f4f-7ba9-45c8-80eb-c20a513f9513
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.reducers=<number>
Starting Job = Job_1610473164347_0008, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0008/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0008
Hadoop Job Information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:13:23,930 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:13:28,023 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 0.98 sec
2021-01-12 13:13:32,110 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.23 sec
MapReduce Total cumulative CPU time: 2 seconds 230 msec
Ended Job = Job_1610473164347_0008
MapReduce Jobs Launched:
Stage:Stage-1 Map: 1 Reduce: 1 Cumulative CPU: 2.23 sec HDFS Read: 15980 HDFS Write: 112 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 230 msec
OK
100000 12000
Time taken: 15.364 seconds, Fetched: 1 row(s)
hive>
```

16. Display DISTINCT salaries paid by the company.  
select count (distinct salary) from employee;

```
Activities Terminal Jan 12 13:17
osboxes@osboxes: -
osboxes@osboxes: -
osboxes@osboxes: -

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.reducers=<number>
Starting Job = Job_1610473164347_0008, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0008/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0008
Hadoop Job Information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:13:23,930 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:13:28,023 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 0.98 sec
2021-01-12 13:13:32,110 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.23 sec
MapReduce Total cumulative CPU time: 2 seconds 230 msec
Ended Job = Job_1610473164347_0008
MapReduce Jobs Launched:
Stage:Stage-1 Map: 1 Reduce: 1 Cumulative CPU: 2.23 sec HDFS Read: 15980 HDFS Write: 112 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 230 msec
OK
100000 12000
Time taken: 15.364 seconds, Fetched: 1 row(s)
hive> select count (distinct salary) from employee;
Query ID = osboxes_20210112131538_54b99d8a-c7cd-4f5c-9ebd-b4d2dba9d17
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.reducers=<number>
Starting Job = Job_1610473164347_0009, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0009/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0009
Hadoop Job Information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:15:43,450 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:15:47,550 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.6 sec
2021-01-12 13:15:52,666 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.92 sec
MapReduce Total cumulative CPU time: 2 seconds 920 msec
Ended Job = Job_1610473164347_0009
MapReduce Jobs Launched:
Stage:Stage-1 Map: 1 Reduce: 1 Cumulative CPU: 2.92 sec HDFS Read: 11066 HDFS Write: 101 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 920 msec
OK
0
Time taken: 16.583 seconds, Fetched: 1 row(s)
hive> select * (distinct salary) from employee;
FAILED: ParseException line 1:9 missing EOF at '(' near '*'
hive>
```

17. List the employees with increasing order of salary paid.  
select \* from employee order by salary desc;



```

Activities  Terminal  Jan 12 13:18
osboxes@osboxes: -
osboxes@osboxes: -
osboxes@osboxes: -
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_1610473164347_0011, Tracking URL = http://osboxes:8088/proxy/application_1610473164347_0011/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610473164347_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-12 13:18:37,533 Stage-1 map = 0%, reduce = 0%
2021-01-12 13:18:41,650 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.19 sec
2021-01-12 13:18:45,780 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.41 sec
MapReduce Total cumulative CPU time: 2 seconds 410 msec
Ended Job = job_1610473164347_0011
MapReduce Jobs Launched:
  Stage:Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.41 sec HDFS Read: 13629 HDFS Write: 993 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 410 msec
OK
20 Michael_Ceo 100000 A
15 Anju_Ceo 100000 B
25 Murugan_CFO 100000 A
3 Janet_Sales 60000 A
12 Harish_Sales 60000 B
5 Sanker_Admin 50000 C
9 Nancy_Admin 50000 A
4 Harl_Admin 50000 C
1 Anne_Admin 50000 A
2 Gokul_Admin 50000
23 Gopal_Sales 50000
18 Gopika_Admin 50000
17 Parvathy_Engineer
8 Jinju_Engineer
22 Paras_Engineer
14 Laura_Engineer
24 Sukant_Engineer
19 Steven_Engineer
16 Aarathi_Manager 40000
11 Arun_Manager 40000
12 Priyanka_Admin
10 Andrew_Manager 40000 A
13 Robert_Manager 40000 A
6 Margaret_Tech 12000 A
7 Nitral_Tech 12000 B
Time Taken: 14.616 seconds, Fetched: 25 row(s)
hive>

```

**18. Make two partitions of the table with the following criteria:**

**a. Partition 1 consisting of employees having salary =50000**

**Create table employee\_part1 (empid int , name string, designation string, dept char(1))**

**Partitioned by (salary int)**

**Row format delimited**

**Fields terminated by ','**

**Lines terminated by '\n'**

**Stored as textfile;**

**Set hive.exec.dynamic.partition.mode=nonstrict;**

**Insert into employee\_part1**

**partition(salary)**

**select empid, name, designation, dept, salary from employee**

**where salary = 50000;**

**Select \* from employee\_part1**

**OR**

**create table employee\_part1 (eid int, name string, designation string, dept char(1))**

**partitioned by (salary int)**

**row format delimited**

**fields terminate by ','**

**lines terminated by '\n'**

**stored as textfile;**

**alter table employee\_part1**

**add partition (salary=50000)**

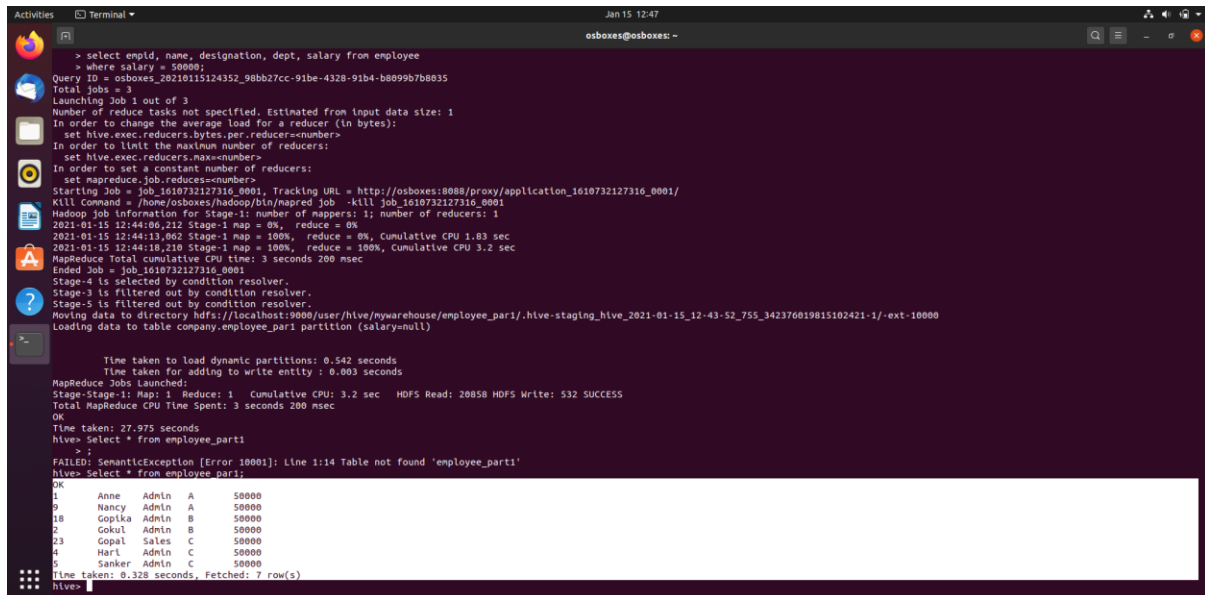
**partition(salary=60000);**

**insert into table employee\_part1**

**partition (salary)**

**select eid , name , dept, designation from employee**

**where salary = 50000;**



```
> select empid, name, designation, dept, salary from employee
> where salary = 50000;
Query ID = osboxes_20210115124352_98bb27cc-91be-4328-91b4-b8099b7b8035
Total Jobs = 3
Launching Job 1 out of 3
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_1610732127316_0001, Tracking URL = http://osboxes:8088/proxy/application_1610732127316_0001/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610732127316_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-15 12:44:06,212 Stage-1 map = 0%, reduce = 0%
2021-01-15 12:44:13,062 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.83 sec
2021-01-15 12:44:18,210 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.2 sec
MapReduce Total cumulative CPU time: 3 seconds 200 msec
Ended Job = job_1610732127316_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 directory hdfs://localhost:9000/user/hive/mywarehouse/employee_part1/.hive-staging_hive_2021-01-15_12-43-52_755_342376019815102421-1/-ext-10000
Loading data to table company.employee_part1 partition (salary=null)

Time taken to load dynamic partitions: 0.542 seconds
Time taken for adding to write entity : 0.003 seconds
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.2 sec HDFS Read: 2085B HDFS Write: 532 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 200 msec
OK
Time taken: 27.975 seconds
hive> select * from employee_part1
> ;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'employee_part1'
hive> select * from employee_part1;
OK
1      Anne  AdmIn  A      50000
9      Nancy AdmIn  A      50000
18     Gopika  AdmIn  B      50000
2      Gokul  AdmIn  B      50000
23     Gopal  Sales  C      50000
4      Harl    AdmIn  C      50000
5      Sanker  AdmIn  C      50000
Time taken: 0.328 seconds, Fetched: 7 row(s)
hive>
```

**b. Partition 2 consisting of employees having salary =60000**

**Insert into employee\_part1**

**partition(salary)**

**select empid, name, designation, dept, salary from employee**

**where salary = 60000;**

**Select \* from employee\_part1**

```
Activities Terminal Jan 15 12:50 osboxes@osboxes: -
Launching Job 1 out of 3
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.reducers=<number>
Starting Job = job_1610732127316_0002, Tracking URL = http://osboxes:8088/proxy/application_1610732127316_0002/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610732127316_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-15 12:49:17,934 Stage-1 map = 0%, reduce = 0%
2021-01-15 12:49:24,097 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.89 sec
2021-01-15 12:49:29,212 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.23 sec
MapReduce Total cumulative CPU time: 3 seconds 230 msec
Ended Job = job_1610732127316_0002
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 directory hdfs://localhost:9000/user/hive/mywarehouse/employee_part1/.hive-staging_hive_2021-01-15_12-49-09_519_4176115946026385220-1/-ext-10000
Loading data to table company.employee_part1 partition (salary=null)

Time taken to load dynamic partitions: 0.15 seconds
Time taken for adding to write entity : 0.001 seconds
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.23 sec HDFS Read: 20866 HDFS Write: 383 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 230 msec
OK
Time taken: 21.178 seconds
hive> select * from employee_part1;
OK
1 Anne Admln A 50000
9 Nancy Admln A 50000
18 Gopika Admln B 50000
2 Gokul Admln B 50000
23 Gopal Sales C 50000
4 Harl Admln C 50000
5 Sanker Admln C 50000
3 Janet Sales A 60000
12 Harish Sales B 60000
Time taken: 0.189 seconds, Fetched: 9 row(s)
hive> select * from employee_part1 where salary = 60000;
OK
3 Janet Sales A 60000
12 Harish Sales B 60000
Time taken: 0.184 seconds, Fetched: 2 row(s)
hive>
```

## 19. Create partition 3 for all Engineer having salary =45000.

Insert into employee\_part1  
partition(salary)

select empid, name, designation, dept, salary from employee  
where salary = 60000 and designation = 'Engineer';

```
Activities Terminal Jan 15 12:55 osboxes@osboxes: -
Query ID = osboxes_20210115125348_babbf3e2-2607-42ea-bb9c-49a64a34a38a
Total Jobs = 3
Launching Job 1 out of 3
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.reducers=<number>
Starting Job = job_1610732127316_0003, Tracking URL = http://osboxes:8088/proxy/application_1610732127316_0003/
Kill Command = /home/osboxes/hadoop/bin/mapred job -kill job_1610732127316_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-01-15 12:53:57,410 Stage-1 map = 0%, reduce = 0%
2021-01-15 12:54:02,729 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.59 sec
2021-01-15 12:54:07,840 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.93 sec
MapReduce Total cumulative CPU time: 3 seconds 930 msec
Ended Job = job_1610732127316_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 directory hdfs://localhost:9000/user/hive/mywarehouse/employee_part1/.hive-staging_hive_2021-01-15_12-53-48_575_8449353108486058390-1/-ext-10000
Loading data to table company.employee_part1 partition (salary=null)

Time taken to load dynamic partitions: 0.999 seconds
Time taken for adding to write entity : 0.0 seconds
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.93 sec HDFS Read: 21384 HDFS Write: 89 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 930 msec
OK
Time taken: 20.443 seconds
hive> select * from employee_part1;
OK
1 Anne Admln A 50000
9 Nancy Admln A 50000
18 Gopika Admln B 50000
23 Gopal Sales C 50000
4 Harl Admln C 50000
5 Sanker Admln C 50000
3 Janet Sales A 60000
12 Harish Sales B 60000
Time taken: 0.111 seconds, Fetched: 9 row(s)
hive> select * from employee_part1 where salary = 60000 and designation = 'Engineer';
OK
Time taken: 0.149 seconds
hive>
```