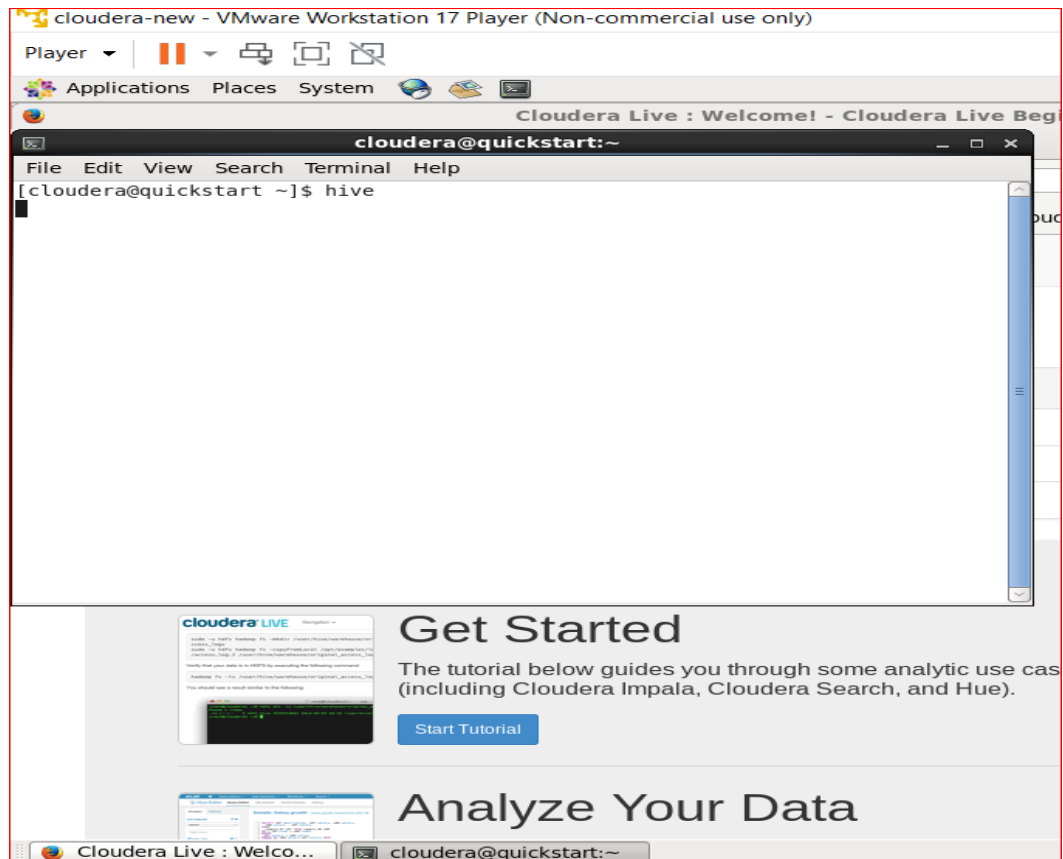


# HIVE PROJECT :-EMPLOYEE MANAGEMENT DATABASE

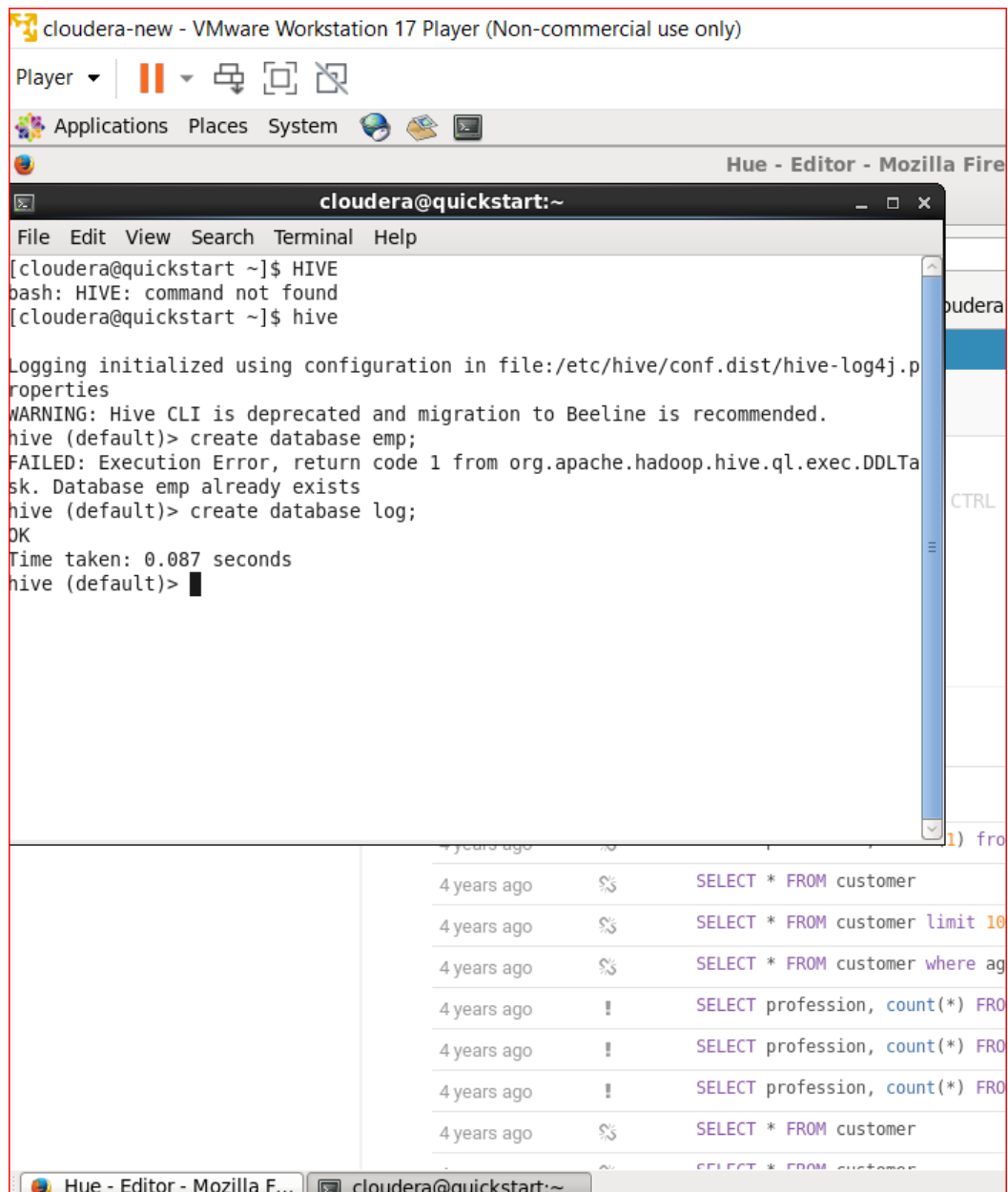
- **ACTIVATE HIVE INSIDE THE TERMINAL.**



## STEPS:- APPLICATIONS>SYSTEM TOOLS>TERMINAL

- Enter the Keyword “Hive” inside the Linux Terminal to Activate Hive.

- **CREATE DATABASE**



The screenshot shows a VMware Workstation 17 Player window titled "cloudera-new - VMware Workstation 17 Player (Non-commercial use only)". Inside the VM, there is a terminal window titled "cloudera@quickstart:~" and a Hue Editor window titled "Hue - Editor - Mozilla Fire".

The terminal window shows the following commands and output:

```
[cloudera@quickstart ~]$ HIVE
bash: HIVE: command not found
[cloudera@quickstart ~]$ hive

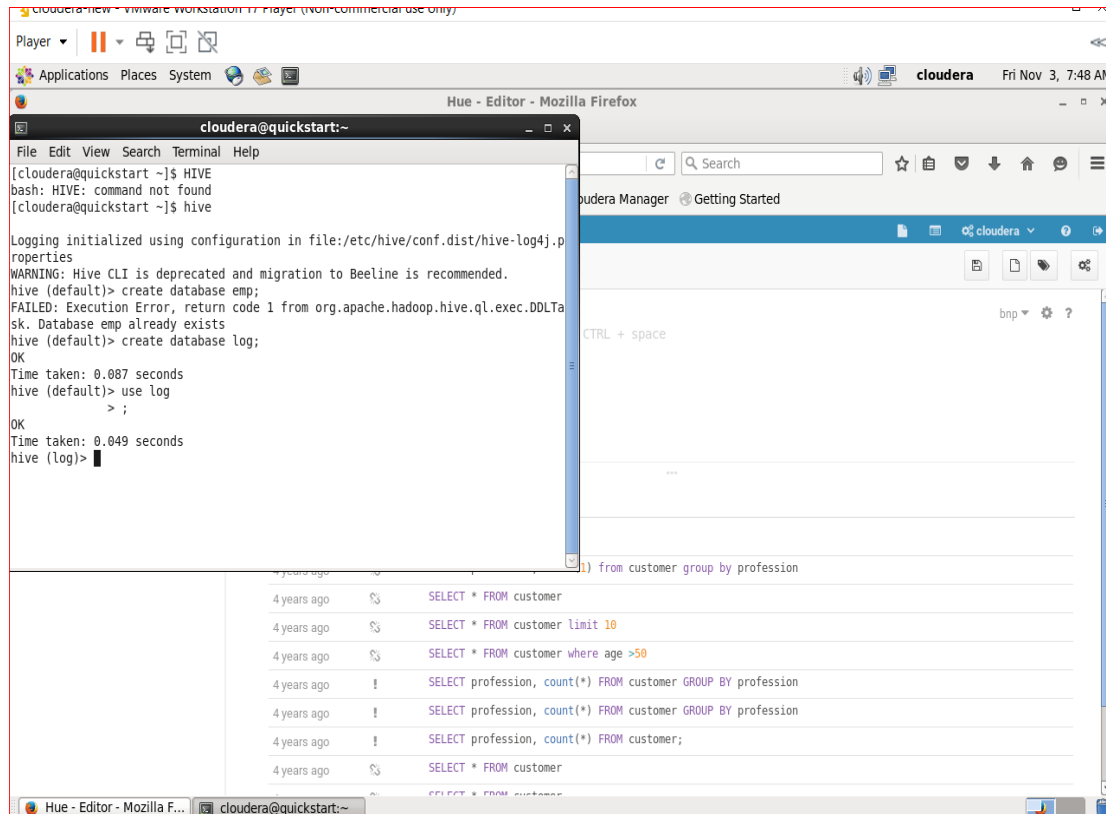
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive (default)> create database emp;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive ql.exec.DDLTa
sk. Database emp already exists
hive (default)> create database log;
OK
Time taken: 0.087 seconds
hive (default)> █
```

The Hue Editor window shows a list of SQL queries in the background, including:

- SELECT \* FROM customer
- SELECT \* FROM customer limit 10
- SELECT \* FROM customer where ag
- SELECT profession, count(\*) FRO
- SELECT profession, count(\*) FRO
- SELECT profession, count(\*) FRO
- SELECT \* FROM customer
- SELECT \* FROM customer

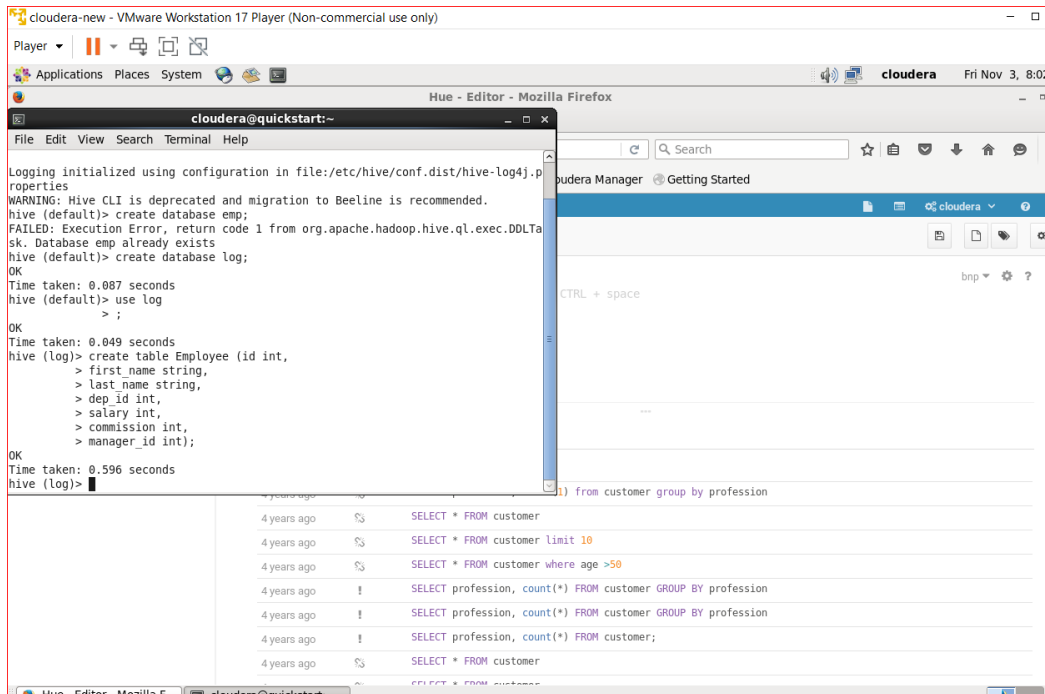
**SYNTAX :-CREATE DATABASE DATABASE\_NAME;**

- **USE DATABASE**



**SYNTAX :-USE DATABASE\_NAME;**

## • CREATE TABLE EMPLOYEE



The screenshot shows a VMware Workstation 17 Player window titled "cloudera-new - VMware Workstation 17 Player (Non-commercial use only)". Inside the VM, there is a terminal window titled "cloudera@quickstart:~" and a web browser window titled "Hue - Editor - Mozilla Firefox".

The terminal window displays the following commands and output:

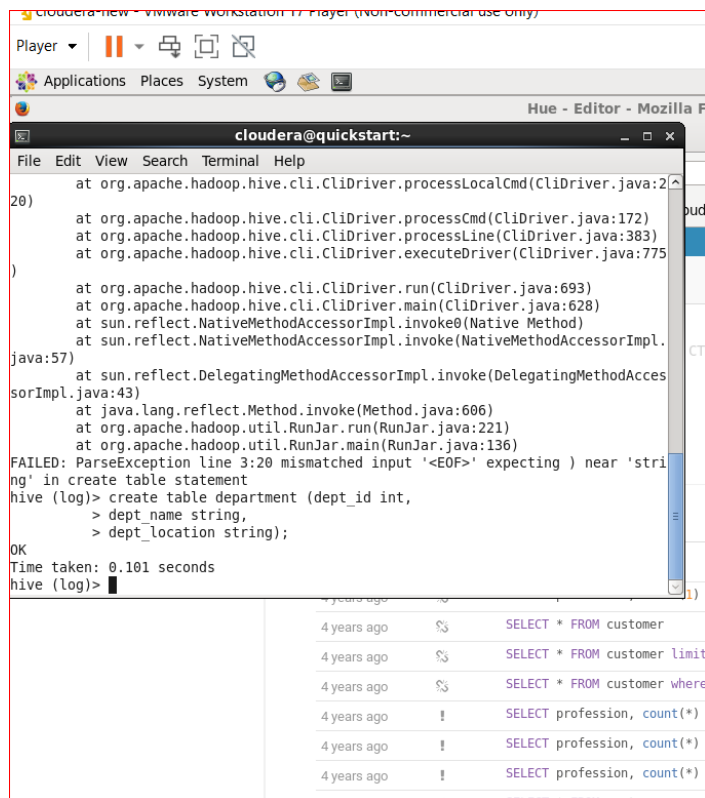
```
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive (default)> create database emp;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. Database emp already exists
hive (default)> create database log;
OK
Time taken: 0.087 seconds
hive (default)> use log
> ;
OK
Time taken: 0.049 seconds
hive (log)> create table Employee (id int,
> first_name string,
> last_name string,
> dep_id int,
> salary int,
> commission int,
> manager_id int);
OK
Time taken: 0.596 seconds
hive (log)>
```

The web browser window shows the Hue interface. The "cloudera" dropdown menu is open, showing a list of recent queries. The queries are as follows:

Time	Query
4 years ago	SELECT * FROM customer
4 years ago	SELECT * FROM customer limit 10
4 years ago	SELECT * FROM customer where age >50
4 years ago	SELECT profession, count(*) FROM customer GROUP BY profession
4 years ago	SELECT profession, count(*) FROM customer GROUP BY profession
4 years ago	SELECT profession, count(*) FROM customer;
4 years ago	SELECT * FROM customer

**SYNTAX :- CREATE TABLE TABLE\_NAME (COL\_NAME 1  
DATATYPE,COL\_NAME 2, DATATYPE);**

- **CREATE TABLE DEPARTMENT**



The screenshot shows a VMware Workstation 17 Player window titled "cloudera-new - VMware Workstation 17 Player (Non-Commercial use Only)". Inside the VM, a terminal window titled "cloudera@quickstart:~" displays the following output:

```
at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:2
20)
    at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:172)
    at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:383)
    at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:775
)
    at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:693)
    at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:628)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl
java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAcces
sorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 3:20 mismatched input '<EOF>' expecting ')' near 'stri
ng' in create table statement
hive (log)> create table department (dept_id int,
> dept_name string,
> dept_location string);
OK
Time taken: 0.101 seconds
hive (log)> █
```

Below the terminal window, a portion of the Hue interface is visible, showing a table with columns for time, status, and SQL queries. The visible rows are:

Time	Status	SQL
4 years ago	⌘	SELECT * FROM customer
4 years ago	⌘	SELECT * FROM customer limit
4 years ago	⌘	SELECT * FROM customer where
4 years ago	!	SELECT profession, count(*)
4 years ago	!	SELECT profession, count(*)
4 years ago	!	SELECT profession, count(*)

- CREATE TABLE PROJECT

The screenshot shows a VMware Workstation 17 Player window titled "cloudera-new - VMware Workstation 17 Player (Non-commercial use only)". Inside the VM, there is a terminal window titled "cloudera@quickstart:~" and a web browser window titled "Hue - Editor - Mozilla Firefox".

The terminal window displays the following output:

```
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:693)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:628)
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:606)
at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 3:20 mismatched input '<EOF>' expecting ')' near 'string' in create table statement
hive (log)> create table department (dept_id int,
> dept_name string,
> dept_location string);
OK
Time taken: 0.101 seconds
hive (log)> create table project (proj_id int,
> emp_id int,
> project_name string);
OK
Time taken: 0.133 seconds
hive (log)>
```

The web browser window shows the Hue interface. The "Table" tab is selected, and a list of tables is displayed. The tables are:

Table Name	Size	Created	Query
customer	4 years ago	4 years ago	SELECT * FROM customer
customer	4 years ago	4 years ago	SELECT * FROM customer limit 10
customer	4 years ago	4 years ago	SELECT * FROM customer where age >50
customer	4 years ago	4 years ago	SELECT profession, count(*) FROM customer
customer	4 years ago	4 years ago	SELECT profession, count(*) FROM customer
customer	4 years ago	4 years ago	SELECT profession, count(*) FROM customer
customer	4 years ago	4 years ago	SELECT * FROM customer

## • INSERT VALUES INSIDE TABLES

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
09  
bash: syntax error near unexpected token `('  
[cloudera@quickstart ~]$ hive  
  
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p  
roperties  
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.  
hive (default)> use log;  
OK  
Time taken: 0.844 seconds  
hive (log)> INSERT INTO TABLE EMPLOYEE VALUES  
      > (001,"SAMRIDDH","SHETTY",104,250000,75000,008),  
      > (002,"SAKTHESH","NADAR",103,110000,5000,005),  
      > (003,"PUSHPAK","BILAJE",104,155000,7500,006),  
      > (004,"SHARAYU","BANDRE",102,125000,25000,003),  
      > (005,"PRIYA","SHARMA",103,135000,5000,002),  
      > (006,"HIMESH","PATEL",101,100000,11000,001),  
      > (007,"ARKADEEP","MONDAL",102,125000,35000,009),  
      > (008,"SUJAY","VICHARE",101,150000,25000,003),  
      > (009,"RAJ","SHIDANE",102,135000,5000,001);  
Query ID = cloudera_20231103175252_aae553ea-e3f3-4418-94dd-e5dc21bb1b40  
Total jobs = 3  
Launching Job 1 out of 3  
Number of reduce tasks is set to 0 since there's no reduce operator
```

```
hive (log)> insert into department  
  > values (101,"HR","MUMBAI"),  
  > (102,"FINANCE","MUMBAI"),  
  > (103,"MANAGEMENT","MUMBAI"),  
  > (104,"IT","BANGALORE");  
Query ID = cloudera_20231103182121_3f58fa5d-c96f-485b-a9e0-ee33f04e9a74  
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_1699021878097_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1699021878097_0004/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0004  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0  
2023-11-03 18:22:11,224 Stage-1 map = 0%, reduce = 0%  
2023-11-03 18:22:32,176 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.78 sec  
MapReduce Total cumulative CPU time: 3 seconds 780 msec  
Ended Job = job_1699021878097_0004  
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/log.db/department/.hive-staging_hive_2023-11-03_18-21-49_740_7535572914910255679-1/-ext-1  
000  
Loading data to table log.department  
Table log.department stats: [numFiles=1, numRows=4, totalSize=72, rawDataSize=68]  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Cumulative CPU: 3.78 sec HDFS Read: 3669 HDFS Write: 142 SUCCESS  
Total MapReduce CPU Time Spent: 3 seconds 780 msec  
OK  
Time taken: 44.258 seconds  
hive (log)> SELECT * FROM DEPARTMENT;  
OK  
01 HR MUMBAI  
02 FINANCE MUMBAI  
03 MANAGEMENT MUMBAI  
04 IT BANGALORE  
Time taken: 0.205 seconds, Fetched: 4 row(s)  
hive (log)>
```

**SYNTAX:- INSERT INTO TABLE\_NAME VALUES (V1,V2,V3...Vn);**

- USING AGGREGATE FUNCTIONS

## FIND THE TOTAL EXPENSE OF THE COMPANY THAT GOES INTO EMPLOYEE SALARY AND COMMISSION.

```
hive (log)> SELECT SUM(SALARY+COMMISSION) FROM EMPLOYEE;
Query ID = ccloudera_20231104015252_96f35b5a-ela6-428b-90e0-7c7bfff69962
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_1699021878097_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1699021878097_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-11-04 01:52:47,345 Stage-1 map = 0%, reduce = 0%
2023-11-04 01:52:58,169 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.96 sec
2023-11-04 01:53:09,107 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.86 sec
MapReduce Total cumulative CPU time: 3 seconds 860 msec
Ended Job = job_1699021878097_0016
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.86 sec HDFS Read: 8786 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 860 msec
OK
1478500
Time taken: 32.52 seconds, Fetched: 1 row(s)
hive (log)>
```

## FIND THE TOTAL NO OF PROJECTS ASSIGNED BY THE COMPANY.

```
hive (log)> SELECT COUNT(PROJ_ID) FROM PROJECT;
Query ID = ccloudera_20231104015050_de4f428a-c126-4a20-b769-3c295747c429
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_1699021878097_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1699021878097_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-11-04 01:50:29,706 Stage-1 map = 0%, reduce = 0%
2023-11-04 01:50:39,718 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec
2023-11-04 01:50:50,646 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.06 sec
MapReduce Total cumulative CPU time: 3 seconds 60 msec
Ended Job = job_1699021878097_0015
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.06 sec HDFS Read: 6984 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 60 msec
OK
6
Time taken: 31.977 seconds, Fetched: 1 row(s)
hive (log)>
```



## FIND THE AVERAGE

```
hive (log)> SELECT AVG(SALARY) FROM EMPLOYEE;
Query ID = cloudera_20231104014747_436967d0-5876-4a94-af62-7b95400c6b69
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_1699021878097_0014, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1699021878097_0014/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-11-04 01:47:37,981 Stage-1 map = 0%, reduce = 0%
2023-11-04 01:47:46,775 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.34 sec
2023-11-04 01:47:57,593 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.03 sec
MapReduce Total cumulative CPU time: 3 seconds 30 msec
Ended Job = job_1699021878097_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.03 sec HDFS Read: 7821 HDFS Write: 19 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 30 msec
OK
142777.77777777778
Time taken: 31.844 seconds, Fetched: 1 row(s)
```

- HQL QUERIES WITH CONDITIONS.

## FULL NAME OF EMPLOYEES WHOSE SALARY IS GREATER THE 150000.

```
Time taken: 0.135 seconds
hive (log)> SELECT CONCAT(FIRSTNAME, ' ', LASTNAME) AS FULL_NAME FROM EMP WHERE
SALARY > 150000;
OK
SAMRIDDH SHETTY
PUSHPAK BILAJE
Time taken: 1.837 seconds, Fetched: 2 row(s)
hive (log)>
```

## TOP 3 EMPLOYEES WITH LOWEST SALARY

```
hive (log)> select * from emp order by salary limit 3;
Query ID = cloudera_20231104005252_a725cbad-078c-428a-ac92-bc600083f3f6
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_1699021878097_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1699021878097_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-11-04 00:52:52.055 Stage-1 map = 0%, reduce = 0%
2023-11-04 00:53:01.927 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec
2023-11-04 00:53:14.049 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.16 sec
MapReduce Total cumulative CPU time: 3 seconds 160 msec
Ended Job = job_1699021878097_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.16 sec HDFS Read: 7859 HDFS Write: 106 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 160 msec
OK
6      HIMESH PATEL      101      100000  11000  1
2      SAKTHESH         NADAR   103      110000  5000  5
7      ARKADEEP         MONDAL  102      125000  35000  9
Time taken: 33.691 seconds, Fetched: 3 row(s)
```

## EMPLOYEES WHOSE FIRST NAME OR LASTNAME STARTS WITH S.

```
hive (log)> select * from emp where firstname like "S%" OR lastname like "S%";
OK
1      SAMRIDDH         SHETTY  104      250000  75000  8
2      SAKTHESH         NADAR   103      110000  5000  5
4      SHARAYU BANDRE  102      125000  25000  3
5      PRIYA           SHARMA  103      135000  5000  2
8      SUJAY           VICHARE 101      150000  25000  3
9      RAJ             SHIDANE 102      135000  5000  1
Time taken: 0.164 seconds, Fetched: 6 row(s)
hive (log)> █
```

## EMPLOYEES WHOSE LOCATION IS NOT IN MUMBAI.

```
Time taken: 0.157 seconds, Fetched: 4 row(s)
hive (log)> select dept_name from department where dept_location not in("MUMBAI"
);
OK
IT
Time taken: 0.118 seconds, Fetched: 1 row(s)
hive (log)> █
```

- JOINS

## CROSS JOIN TO GET A CARTESIAN PRODUCT OF EMPLOYEE AND DEPARTMENT.

```
hive (log)> select e.firstname,d.dept_id,d.dept_name from emp e cross join department d;  
Warning: Map Join MAPJOIN[7][bigTable=e] in task 'Stage-3:MAPRED' is a cross product  
Query ID = cloudera_20231104003535_93502048-b9bb-4d70-8dca-e4d9adea59ae  
Total jobs = 1  
Execution log at: /tmp/cloudera/cloudera_20231104003535_93502048-b9bb-4d70-8dca-e4d9adea59ae.log  
2023-11-04 12:36:02 Starting to launch local task to process map join; maximum memory = 1013645312
```

```
MapReduce Jobs Launched:  
Stage-Stage-3: Map: 1 Cumulative CPU: 2.09 sec HDFS Read: 6287 HDFS Write: 633 SUCCESS  
Total MapReduce CPU Time Spent: 2 seconds 90 msec  
OK  
SAMRIDDH 101 HR  
SAMRIDDH 102 FINANCE  
SAMRIDDH 103 MANAGEMENT  
SAMRIDDH 104 IT  
SAKTHESH 101 HR  
SAKTHESH 102 FINANCE  
SAKTHESH 103 MANAGEMENT  
SAKTHESH 104 IT  
PUSHPAK 101 HR  
PUSHPAK 102 FINANCE  
PUSHPAK 103 MANAGEMENT  
PUSHPAK 104 IT  
SHARAYU 101 HR  
SHARAYU 102 FINANCE  
SHARAYU 103 MANAGEMENT  
SHARAYU 104 IT  
PRIYA 101 HR  
PRIYA 102 FINANCE  
PRIYA 103 MANAGEMENT  
PRIYA 104 IT  
HIMESH 101 HR  
HIMESH 102 FINANCE  
HIMESH 103 MANAGEMENT  
HIMESH 104 IT  
ARKADEEP 101 HR  
ARKADEEP 102 FINANCE  
ARKADEEP 103 MANAGEMENT  
ARKADEEP 104 IT  
SUJAY 101 HR  
SUJAY 102 FINANCE  
SUJAY 103 MANAGEMENT  
SUJAY 104 IT  
RAJ 101 HR  
RAJ 102 FINANCE  
RAJ 103 MANAGEMENT  
RAJ 104 IT  
Time taken: 43.902 seconds, Fetched: 36 row(s)  
hive (log)>
```

## INNER JOIN ON EMPLOYEE AND PROJECTS TO FIND THE NAME OF THE PROJECT THE PARTICULAR EMPLOYEE IS ASSIGNED TO.

```
hive (log)> select e.firstname,p.proj_id,p.project_name from emp e join project p on e.id=p.emp_id;
Query ID = cloudera_20231104004242_b7eb8e92-1ffd-4588-8056-dc9bb6f3fbd0
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20231104004242_b7eb8e92-1ffd-4588-8056-dc9bb6f3fbd0.log
2023-11-04 12:42:40 Starting to launch local task to process map join; maximum memory = 1013645312
2023-11-04 12:42:52 Dump the side-table for tag: 1 with group count: 4 into file: file:/tmp/cloudera/b5efe3c2-d7e2-4e21-9a75-10b6df265c55/hive_2023-11-04_00-42-40_662_4738445267525687051-1/-local-10003/le31--hashtable
2023-11-04 12:42:52 Uploaded 1 File to: file:/tmp/cloudera/b5efe3c2-d7e2-4e21-9a75-10b6df265c55/hive_2023-11-04_00-42-40_662_4738445267525687051-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile31--hash
2023-11-04 12:42:52 End of local task; Time Taken: 3.339 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_1699021878097_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1699021878097_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0010
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2023-11-04 00:43:05,015 Stage-3 map = 0%, reduce = 0%
2023-11-04 00:43:17,004 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.65 sec
MapReduce Total cumulative CPU time: 2 seconds 650 msec
Ended Job = job_1699021878097_0010
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.65 sec HDFS Read: 6490 HDFS Write: 203 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 650 msec
OK
SAMRIDDH 1004 MARKET TREND ANALYSIS
SAMRIDDH 1006 GLOBAL STOCK ANALYSIS
PUSHPAK 1001 RISK ASSESSMENT
PUSHPAK 1005 PRODUCTIVITY MANAGEMENT
SHARAYU 1002 COMPANY ACCOUNTS DATA
ARKADEEP 1003 PROFIT ANALYSIS
Time taken: 37.531 seconds, Fetched: 6 row(s)
hive (log)> █
```

## SELF JOIN TO FIND THE MANAGER NAME OF THE EMPLOYEE.

```
hive (log)> select e1.id,e1.firstname,e2.firstname,e2.manager_id FROM emp e1 JOIN
EMP e2 ON e1.id=e2.manager_id;
Query ID = cloudera_20231103192727_a3f8a642-c3f9-4b64-bd40-d61c1db0bdb3
Total jobs = 1
```

```
OK SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 610 msec
OK
8 SUJAY SAMRIDDH 8
5 PRIYA SAKTHESH 5
6 HIMESH PUSHPAK 6
3 PUSHPAK SHARAYU 3
2 SAKTHESH PRIYA 2
1 SAMRIDDH HIMESH 1
9 RAJ ARKADEEP 9
3 PUSHPAK SUJAY 3
1 SAMRIDDH RAJ 1
Time taken: 56.84 seconds, Fetched: 9 row(s)
hive (log)> █
```

## JOIN ON SUBQUERY TO FIND THE MAX SALARY OF THE EMPLOYEE.

```
time taken: 0.827 seconds
hive (log)> SELECT e.firstname, e.salary
      > FROM emp e
      > JOIN (SELECT MAX(salary) AS max_salary FROM emp) max_emp
      > ON e.salary = max_emp.max_salary;
Query ID = cloudera_20231103191010_4829a844-2573-4c00-b019-3ec0bb5d554d
Total jobs = 2
Launching Job 1 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>
```

```
Total MapReduce CPU Time Spent: 9 seconds 500 msec
OK
SAMRIDDH          250000
```

## THREE TABLE JOIN TO FIND NAME,SALARY,PROJECT AND LOCATION=MUMBAI

```
hive (log)> SELECT e.first name,e.salary,p.project name,d.dept name,d.dept_location from employee e join project p on e.id=p.emp_id join department d on e.dept_id=d.dept_id where d.dept_location="MUMBAI";
Query ID = cloudera_20231104020606_8b58a76b-c3f5-4d1d-b2c9-61adfa943615
Total jobs = 1
Execution Log at: /tmp/cloudera/cloudera_20231104020606_8b58a76b-c3f5-4d1d-b2c9-61adfa943615.log
2023-11-04 02:06:52 Starting to launch local task to process map join; maximum memory = 1013645312
2023-11-04 02:06:54 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/cloudera/b5efe3c2-d7e2-4e21-9a75-10b6df265c55/hive_2023-11-04_02-06-45_886_3853564915574554900-1/-Local-10005/HashTable-1-.hashtable
2023-11-04 02:06:54 Uploaded 1 File to: file:/tmp/cloudera/b5efe3c2-d7e2-4e21-9a75-10b6df265c55/hive_2023-11-04_02-06-45_886_3853564915574554900-1/-Local-10005/HashTable-Stage-5/MapJoin-mapfile81-.hashtable
2023-11-04 02:06:54 Dump the side-table for tag: 1 with group count: 4 into file: file:/tmp/cloudera/b5efe3c2-d7e2-4e21-9a75-10b6df265c55/hive_2023-11-04_02-06-45_886_3853564915574554900-1/-Local-10005/HashTable-2-.hashtable
2023-11-04 02:06:54 Uploaded 1 File to: file:/tmp/cloudera/b5efe3c2-d7e2-4e21-9a75-10b6df265c55/hive_2023-11-04_02-06-45_886_3853564915574554900-1/-Local-10005/HashTable-Stage-5/MapJoin-mapfile91-.hashtable
2023-11-04 02:06:54 End of local task; Time Taken: 1.909 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_1699021878097_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1699021878097_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0019
Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 0
2023-11-04 02:10:37,230 Stage-5 map = 0%, reduce = 0%
2023-11-04 02:11:22,173 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 2.41 sec
MapReduce Total cumulative CPU time: 2 seconds 410 msec
Ended Job = job_1699021878097_0019
MapReduce Jobs Launched:
Stage-Stage-5: Map: 1 Cumulative CPU: 2.41 sec HDFS Read: 8342 HDFS Write: 99 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 410 msec
OK
SHARAYU 125000 COMPANY ACCOUNTS DATA FINANCE MUMBAI
ARKADEEP 125000 PROFIT ANALYSIS FINANCE MUMBAI
Time taken: 277.41 seconds, Fetched: 2 row(s)
```

- **CREATE A VIRTUAL TABLE OUT OF EMPLOYEE, DEPARTMENT AND PROJECT TABLE.**

```
hive (log)> Create View vr as select e.first_name,p.project_name,d.dept_name from employee e join project p on e.id=p.id join department d on d.dept_id=e.dept_id;
FAILED: SemanticException [Error 10002]: Line 1:106 Invalid column reference 'id'
hive (log)> Create View vr as select e.first_name,p.project_name,d.dept_name from employee e join project p on e.id=p.emp_id join department d on d.dept_id=e.dept_id;
OK
Time taken: 0.174 seconds
hive (log)> select * from vr;
Query ID = cloudera_20231104024040_6de796d0-7de6-4d51-b535-9420d2c25569
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20231104024040_6de796d0-7de6-4d51-b535-9420d2c25569.log
2023-11-04 02:40:07 Starting to launch local task to process map join; maximum memory = 1013645312
2023-11-04 02:40:10 Dump the side-table for tag: 1 with group count: 4 into file: file:/tmp/cloudera/cc8b0781-c831-4b17-95cb-c2eb4c2d4019/hive_2023-11-04_02-40-00_097_1325717731799284
le61--.hashtable
2023-11-04 02:40:10 Uploaded 1 File to: file:/tmp/cloudera/cc8b0781-c831-4b17-95cb-c2eb4c2d4019/hive_2023-11-04_02-40-00_097_1325717731799284352-1/-local-10005/HashTable-Stage-5/MapJo
2023-11-04 02:40:10 Dump the side-table for tag: 1 with group count: 4 into file: file:/tmp/cloudera/cc8b0781-c831-4b17-95cb-c2eb4c2d4019/hive_2023-11-04_02-40-00_097_1325717731799284
le71--.hashtable
2023-11-04 02:40:10 Uploaded 1 File to: file:/tmp/cloudera/cc8b0781-c831-4b17-95cb-c2eb4c2d4019/hive_2023-11-04_02-40-00_097_1325717731799284352-1/-local-10005/HashTable-Stage-5/MapJo
2023-11-04 02:40:10 End of local task; Time Taken: 3.186 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_1699021878097_0024, Tracking URL = http://quickstart.cloudera:8080/proxy/application_1699021878097_0024/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1699021878097_0024
Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 0
2023-11-04 02:40:22,794 Stage-5 map = 0%, reduce = 0%
2023-11-04 02:40:33,899 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 2.22 sec
MapReduce Total cumulative CPU time: 2 seconds 220 msec
Ended Job = job_1699021878097_0024
MapReduce Jobs Launched:
Stage-Stage-5: Map: 1 Cumulative CPU: 2.22 sec HDFS Read: 8009 HDFS Write: 201 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 220 msec
OK
SAMRIDDH MARKET TREND ANALYSIS IT
SAMRIDDH GLOBAL STOCK ANALYSIS IT
PUSHPAK RISK ASSESSMENT IT
PUSHPAK PRODUCTIVITY MANAGEMENT IT
SHARAYU COMPANY ACCOUNTS DATA FINANCE
ARKADEEP PROFIT ANALYSIS FINANCE
Time taken: 35.961 seconds, Fetched: 6 row(s)
hive (log)> █
```

**SYNTAX :- CREATE VIEW View\_Name AS (condition.....)**