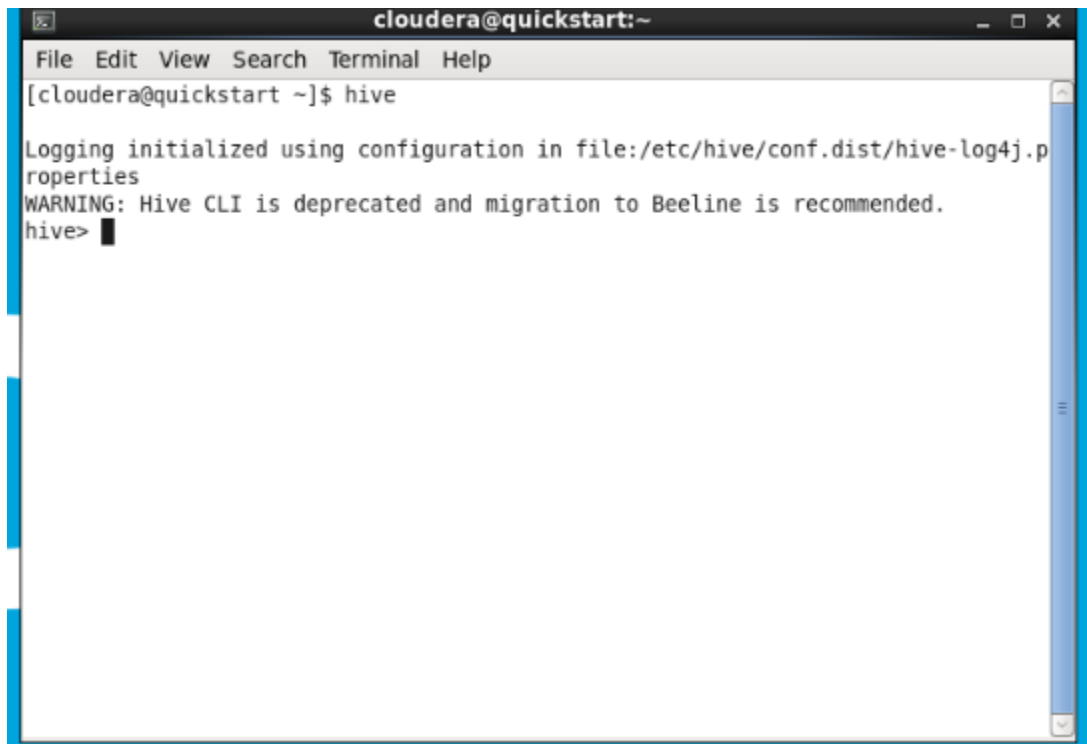


Kaushal Phutane
roll no 19

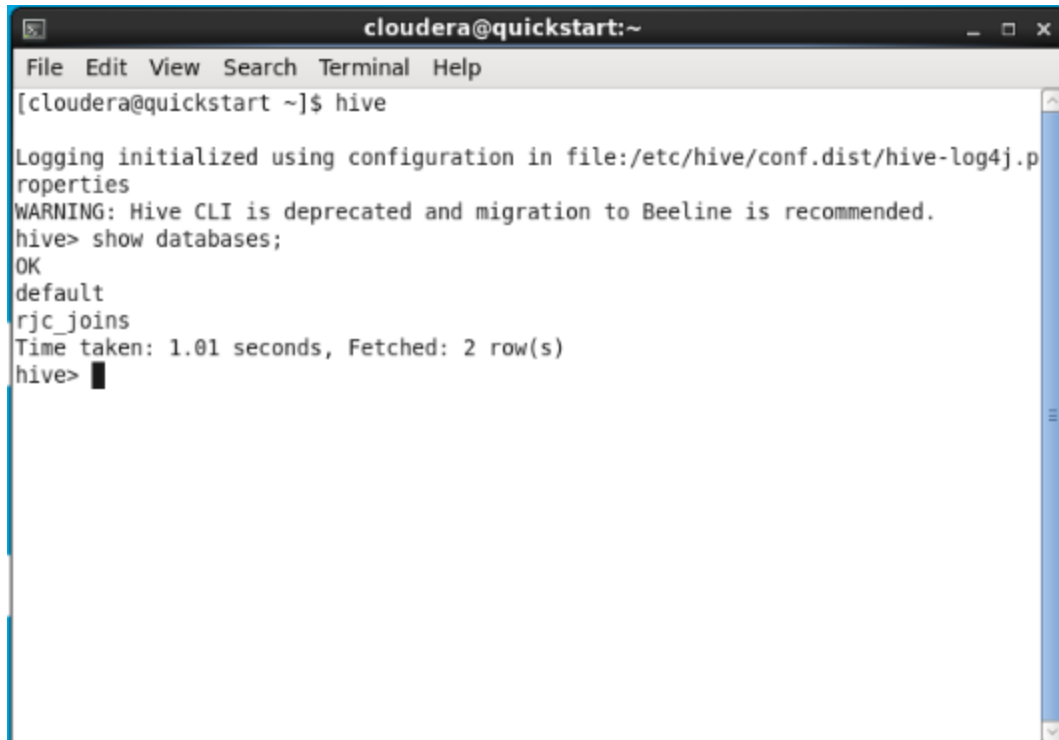
practical no 7

Hive command

A screenshot of a terminal window titled "cloudera@quickstart:~". The terminal has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The command prompt shows "[cloudera@quickstart ~]\$ hive". The output of the command is "Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties" followed by a warning: "WARNING: Hive CLI is deprecated and migration to Beeline is recommended." The prompt then changes to "hive>".

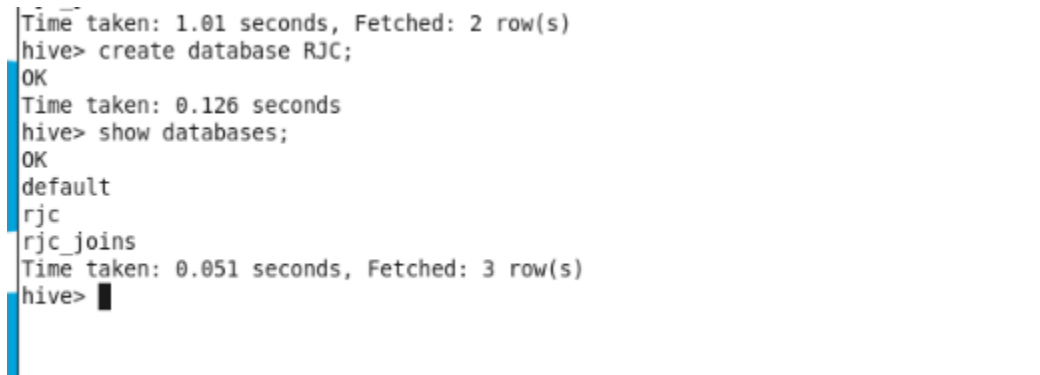
```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hive  
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p  
roperties  
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.  
hive>
```

Show databases;

A terminal window titled 'cloudera@quickstart:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The session shows the user running 'hive', which initializes logging and shows a warning about the deprecated Hive CLI. The user then runs 'show databases;', which returns 'OK' and lists 'default' and 'rjc_joins'.

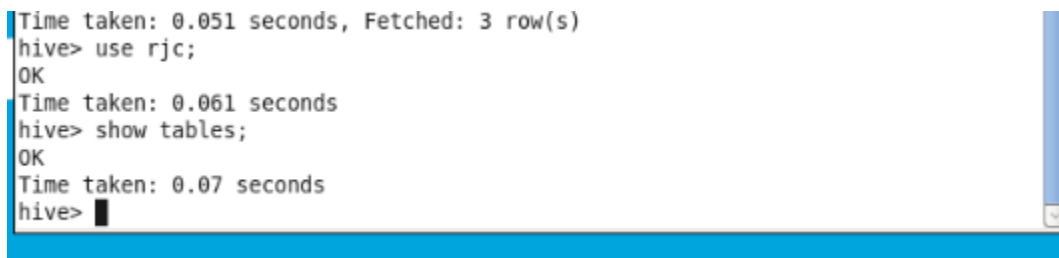
```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ hive  
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p  
roperties  
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.  
hive> show databases;  
OK  
default  
rjc_joins  
Time taken: 1.01 seconds, Fetched: 2 row(s)  
hive> █
```

Create database RJC;

A terminal window showing the continuation of the Hive CLI session. The user runs 'create database RJC;', which returns 'OK'. Then, the user runs 'show databases;', which returns 'OK' and lists 'default', 'rjc', and 'rjc_joins'.

```
Time taken: 1.01 seconds, Fetched: 2 row(s)  
hive> create database RJC;  
OK  
Time taken: 0.126 seconds  
hive> show databases;  
OK  
default  
rjc  
rjc_joins  
Time taken: 0.051 seconds, Fetched: 3 row(s)  
hive> █
```

Use rjc;

A terminal window showing the continuation of the Hive CLI session. The user runs 'use rjc;', which returns 'OK'. Then, the user runs 'show tables;', which returns 'OK'.

```
Time taken: 0.051 seconds, Fetched: 3 row(s)  
hive> use rjc;  
OK  
Time taken: 0.061 seconds  
hive> show tables;  
OK  
Time taken: 0.07 seconds  
hive> █
```

drop rjc;

```
hive> drop database rjc;  
OK  
Time taken: 0.183 seconds  
hive> █
```

Create database RJC;

```
Time taken: 0.183 seconds  
hive> create database rjc;  
OK  
Time taken: 0.072 seconds  
hive> █
```

Use rjc;

```
Time taken: 0.072 seconds  
hive> use rjc;  
OK  
Time taken: 0.03 seconds  
hive> █
```

create table employee

```
Time taken: 0.03 seconds  
hive> create table employee(ID int,name string,salary float,age int)  
  > row format delimited  
  > fields terminated by ',';  
OK  
Time taken: 0.288 seconds  
hive> █
```

describe employee;

```
Time taken: 0.288 seconds  
hive> describe employee;  
OK  
id                int  
name              string  
salary            float  
age               int  
Time taken: 0.15 seconds, Fetched: 4 row(s)
```

describe formatted employee;

```
hive> describe formatted employee;
OK
# col_name          data_type          comment
id                  int
name                string
salary              float
age                 int

# Detailed Table Information
Database:           rjc
Owner:              cloudera
CreateTime:         Mon Mar 21 00:05:22 PDT 2022
LastAccessTime:     UNKNOWN
Protect Mode:       None
Retention:          0
Location:           hdfs://quickstart.cloudera:8020/user/hive/warehouse/rjc.db/employee
Table Type:         MANAGED_TABLE
Table Parameters:
    transient_lastDdlTime    1647846322

# 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: 0.174 seconds, Fetched: 30 row(s)
hive>
```

create external table employee2

```
hive> create external table employee2(ID int,name string,salary float,age int)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 0.162 seconds
hive>
```

describe employee2;

```
Time taken: 0.102 seconds
hive> describe employee2;
OK
id                  int
name                string
salary              float
age                 int
Time taken: 0.267 seconds, Fetched: 4 row(s)
hive>
```

describe formatted employee2;

```
hive> describe employee2;
OK
id                int
name              string
salary            float
age               int
Time taken: 0.267 seconds, Fetched: 4 row(s)
hive> describe formatted employee2;
OK
# col_name          data_type          comment

id                  int
name                string
salary              float
age                 int

# Detailed Table Information
Database:            rjc
Owner:               cloudera
CreateTime:          Mon Mar 21 00:15:09 PDT 2022
LastAccessTime:      UNKNOWN
Protect Mode:        None
Retention:           0
Location:             hdfs://quickstart.cloudera:8020/user/hive/warehouse/rjc.db/employee2
Table Type:          EXTERNAL_TABLE
Table Parameters:
    EXTERNAL              TRUE
    transient_lastDdlTime 1647846909

# 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

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: 0.242 seconds, Fetched: 31 row(s)
hive> █
```

Open the browser

Kaushal Phutane
roll no 19

practical no 7

Cloudera Live : Welcome! - Cloudera Live Beginner Tutorial - Mozilla Firefox

Cloudera Live : Welcom... x

quickstart.cloudera/#/

Search

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

cloudera LIVE Navigation

Welcome to Your Cloudera QuickStart VM!

Your Cluster	
Node	Address
Manager Node	127.0.0.1
Worker Node 1	127.0.0.1

Get Started

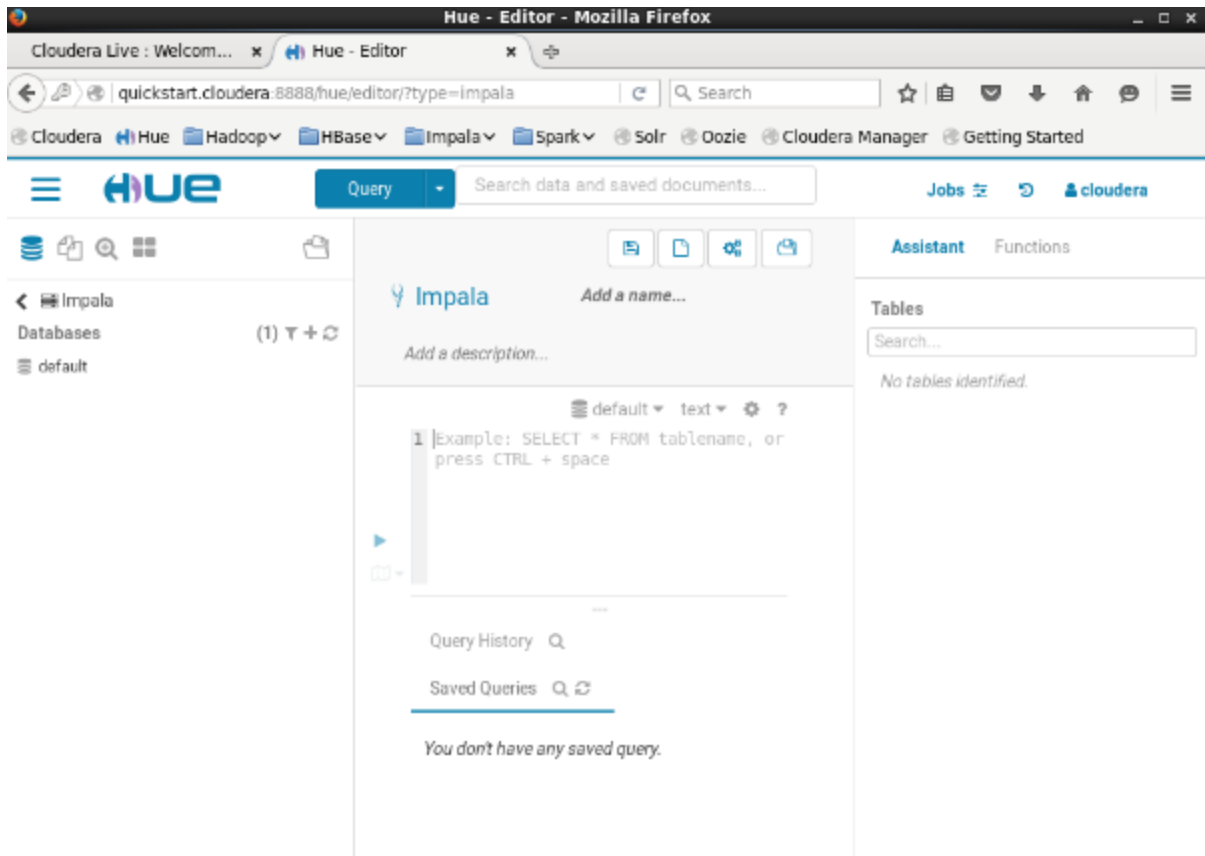
The tutorial below guides you through some analytic use cases, using the most popular open source tools included with CDH (including Cloudera Impala, Cloudera Search, and Hue).

[Start Tutorial](#)

Analyze Your Data

Kaushal Phutane
roll no 19

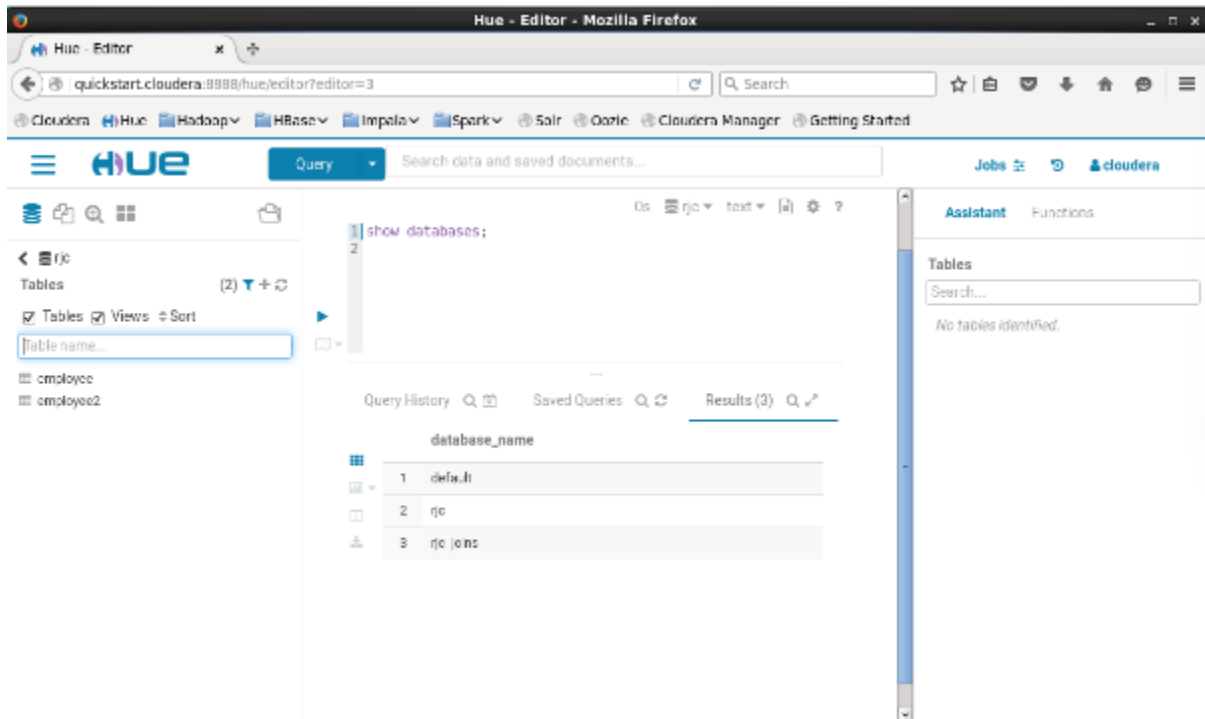
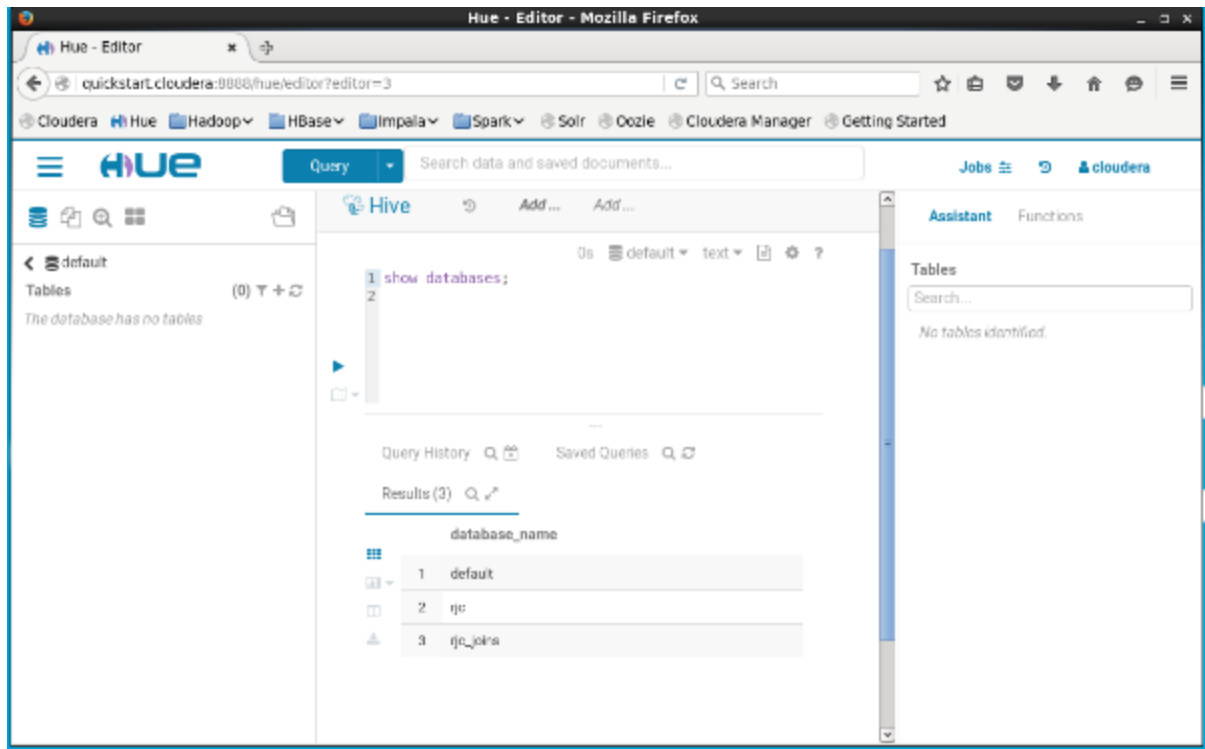
practical no 7



Show databases

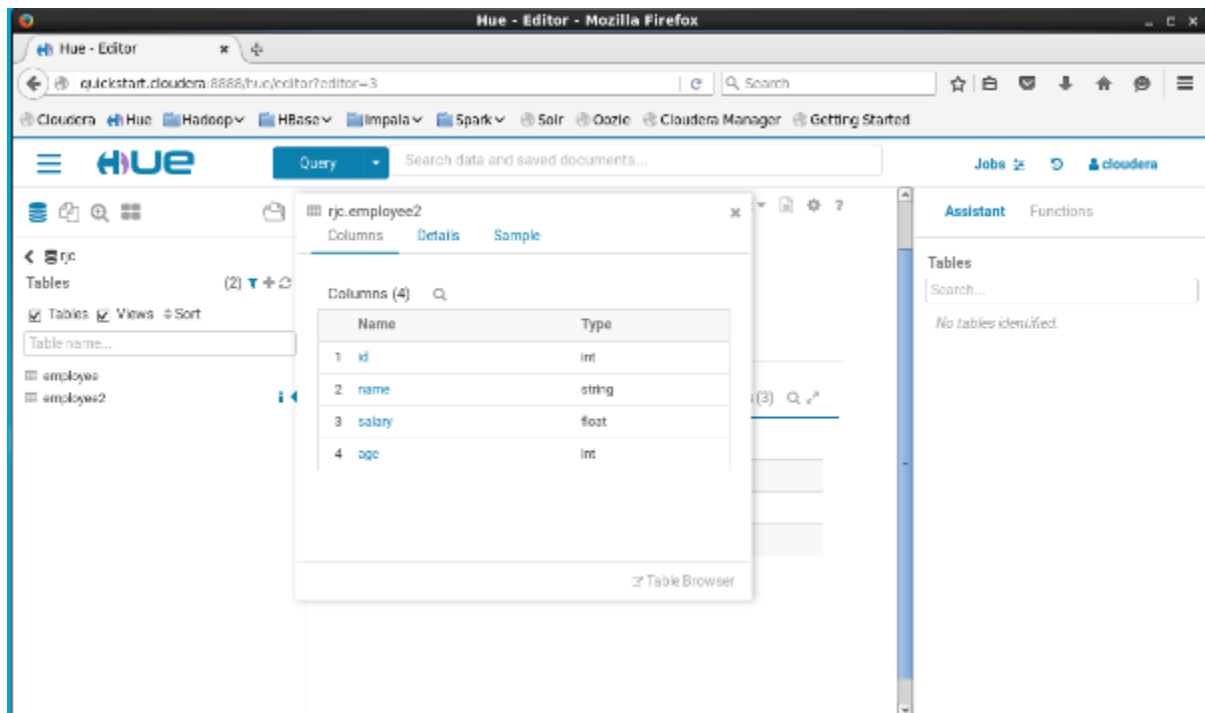
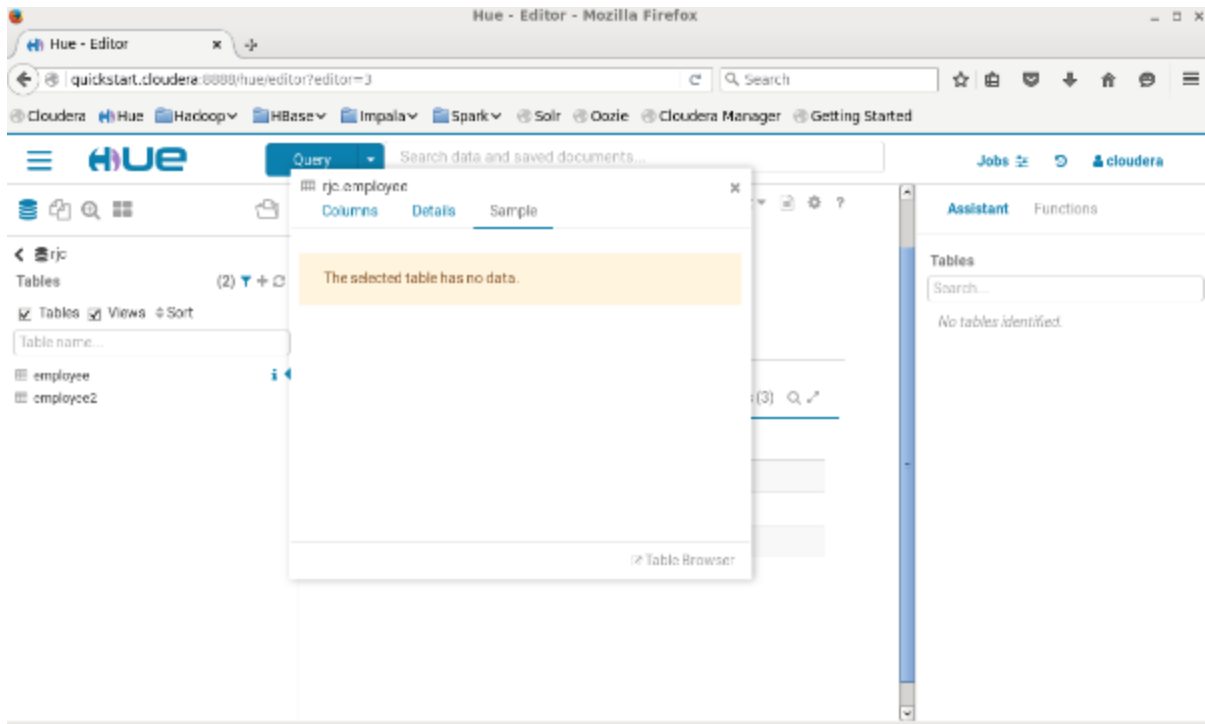
Kaushal Phutane
roll no 19

practical no 7



Kaushal Phutane
roll no 19

practical no 7



Create employee3

Kaushal Phutane
roll no 19

practical no 7

```
Time taken: 0.242 seconds, Fetched: 31 row(s)
hive> create external table employee3(ID int,name string,salary float,age int)
> row format delimited
> fields terminated by ','
> location '/user/cloudera/vj';
OK
Time taken: 0.176 seconds
hive>
```

Describe employee3

```
Time taken: 0.176 seconds
hive> describe employee3;
OK
id                int
name              string
salary           float
age              int
Time taken: 0.104 seconds, Fetched: 4 row(s)
hive>
```

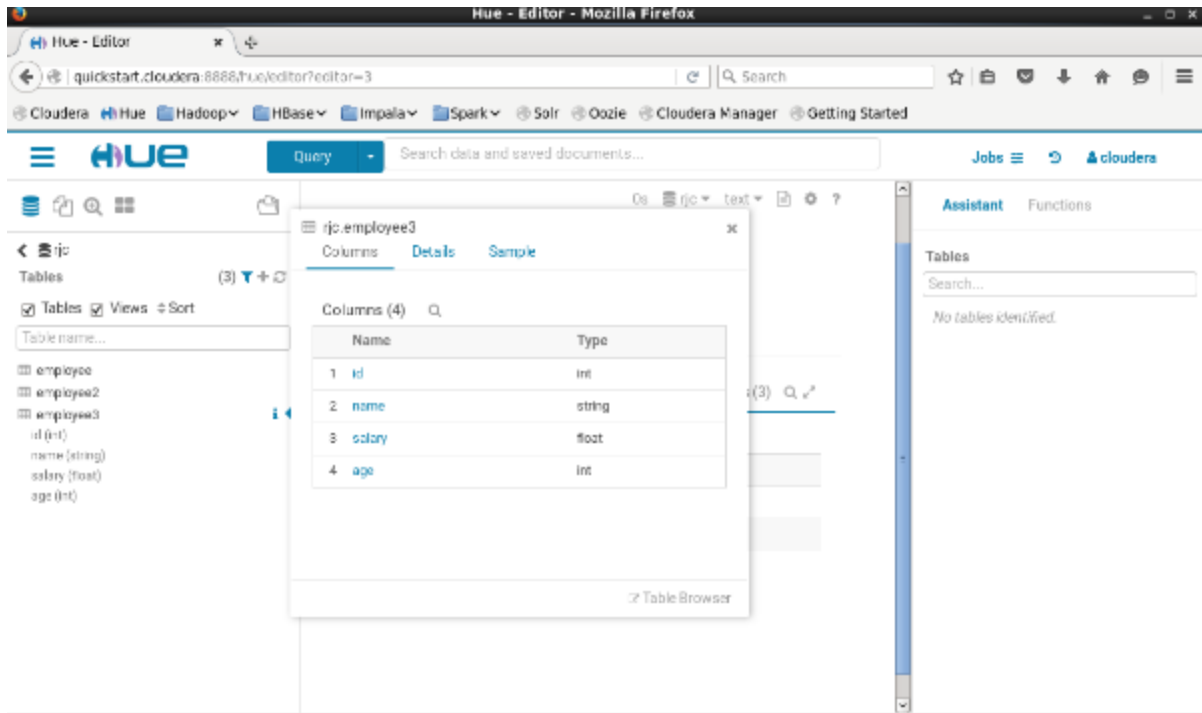
Now refresh browser ,employee3 is also generated in rjc database.

The screenshot shows the Hue web interface in a Mozilla Firefox browser. The URL is `quickstart.cloudera:8888/hue/editor?editor=3`. The interface includes a top navigation bar with various tools like Cloudera, Hue, Hadoop, HBase, Impala, Spark, Sair, Oozie, and Cloudera Manager. The main area displays a query editor with the command `show databases;` and its results. The results are shown in a table with the header `database_name` and three rows: `1 default`, `2 rjc`, and `3 rjc_joinx`. On the left, a sidebar shows a tree view of databases, including `employee`, `employee2`, and `employee3`. On the right, there is an 'Assistant' panel with a 'Tables' section that currently shows 'No tables identified.'

database_name
1 default
2 rjc
3 rjc_joinx

Kaushal Phutane
roll no 19

practical no 7



Listing out all tables

```
oys      1111
Time taken: 0.104 seconds, Fetched: 4 row(s)
hive> show tables;
OK
employee
employee2
employee3
Time taken: 0.044 seconds, Fetched: 3 row(s)
hive> █
```

Alter table

```
Time taken: 0.044 seconds, Fetched: 3 row(s)
hive> alter table employee3 RENAME to emptable;
OK
Time taken: 0.219 seconds
hive> █
```

Show tables after altering

Kaushal Phutane
roll no 19

practical no 7

```
Time taken: 0.219 seconds
hive> show tables;
OK
employee
employee2
emptable
Time taken: 0.024 seconds, Fetched: 3 row(s)
hive> █
```

describe emptable;

Alter table emptable add columns (surname string);

describe emptable;

```
Time taken: 0.024 seconds, Fetched: 3 row(s)
hive> describe emptable;
OK
id                int
name              string
salary            float
age              int
Time taken: 0.105 seconds, Fetched: 4 row(s)

hive> Alter table emptable add columns (surname string);
OK
Time taken: 0.173 seconds
hive> describe emptable;
OK
id                int
name              string
salary            float
age              int
surname           string
Time taken: 0.155 seconds, Fetched: 5 row(s)
hive> █
```

Alter table emptable change name first_name string;

describe emptable;

```
hive> Alter table emptable change name first_name string;
OK
Time taken: 0.203 seconds
hive> describe emptable;
OK
id                int
first_name        string
salary            float
age               int
surname           string
Time taken: 0.109 seconds, Fetched: 5 row(s)
hive> █
```

Loading the data in the table

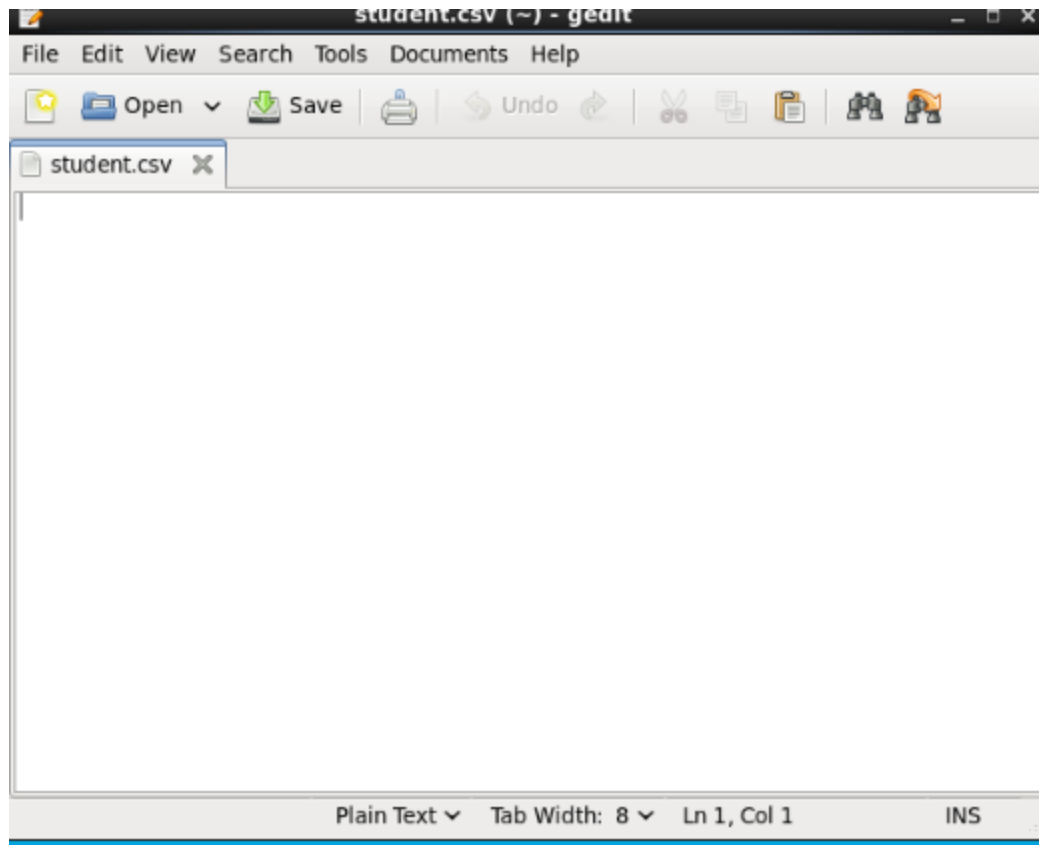
```
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ ls
cloudera-manager  Downloads          lib               Templates
cm_api.py         eclipse            Music            Videos
departments.java  enterprise-deployment.json  parcels          workspace
Desktop           express-deployment.json  Pictures
Documents         kerberos           Public
```

gedit command

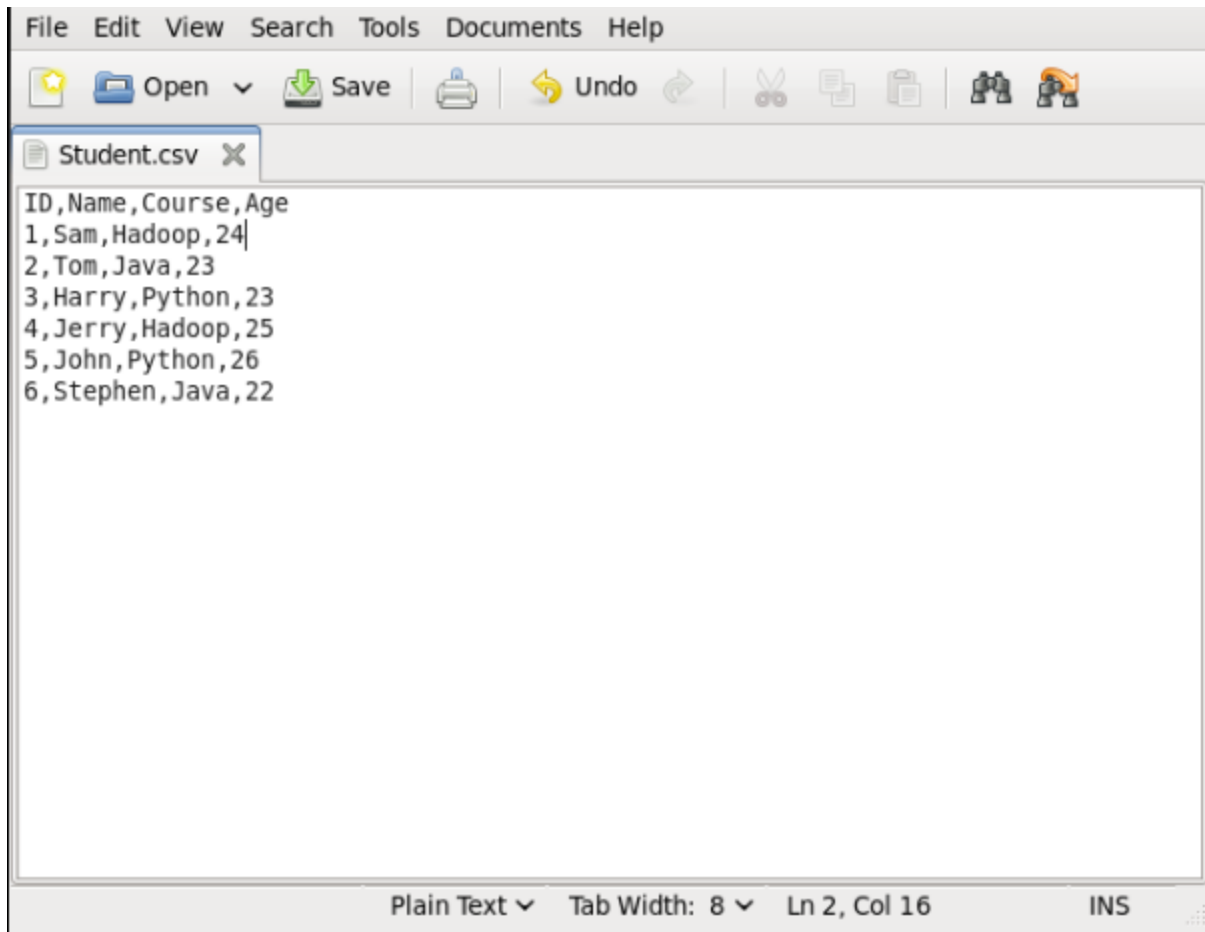
```
[cloudera@quickstart Desktop]$ gedit Student.csv
█
```

Kaushal Phutane
roll no 19

practical no 7



Inset data in student.csv file



create database rjcstudent;
show databases;

```
Time taken: 0.109 seconds, Fetched: 5 row(s)
hive> create database rjcstudent;
OK
Time taken: 0.088 seconds
hive> show databases;
OK
default
rjc
rjc_joins
rjcstudent
Time taken: 0.026 seconds, Fetched: 4 row(s)
hive> █
```

Use rjcstudent;

```
Time taken: 0.026 seconds, Fetched: 4 row(s)
hive> use rjcstudent;
OK
Time taken: 0.03 seconds
hive> █
```

Use rjcstudent;

Create table student

```
hive> use rjcstudent;
OK
Time taken: 0.03 seconds
hive> create table student (ID int,Name string,Age int)
> partitioned by(Course string)
> row format delimited
> fields terminated by ',';
OK
Time taken: 0.093 seconds
hive>
```

Describe Student

```
Time taken: 0.093 seconds
hive> describe student;
OK
id                int
name              string
age              int
course            string

# Partition Information
# col_name        data_type        comment

course            string
Time taken: 0.464 seconds, Fetched: 9 row(s)
hive>
```

```
hive> load data local inpath '/home/cloudera/Desktop' into table student
> partition(Course="Hadoop")
> ;
Loading data to table rjcstudent.student partition (course=Hadoop)
Partition rjcstudent.student{course=Hadoop} stats: [numFiles=10, numRows=0, totalSize=26158, rawDataSize=0]
OK
Time taken: 1.436 seconds
hive>
```

Select * from Student

```
NULL    NULL    NULL    Hadoop
4500    NULL    NULL    Hadoop
NULL    Fetched: 1 row(s)    NULL    Hadoop
NULL    NULL    NULL    Hadoop
NULL    Name    NULL    Hadoop
1       Sam     NULL    Hadoop
2       tom     NULL    Hadoop
3       harry   NULL    Hadoop
4       jerry   NULL    Hadoop
5       john    NULL    Hadoop
6       stephen NULL    Hadoop
Time taken: 0.092 seconds, Fetched: 507 row(s)
hive>
```


Kaushal Phutane
roll no 19

practical no 7

```
Loading data to table rjcstudent.student partition (course=Java)
Partition rjcstudent.student{course=Java} stats: [numFiles=1, numRows=0, totalSize=121, rawDataSize=0]
OK
Time taken: 1.276 seconds
hive>
```

Again Create Table Student

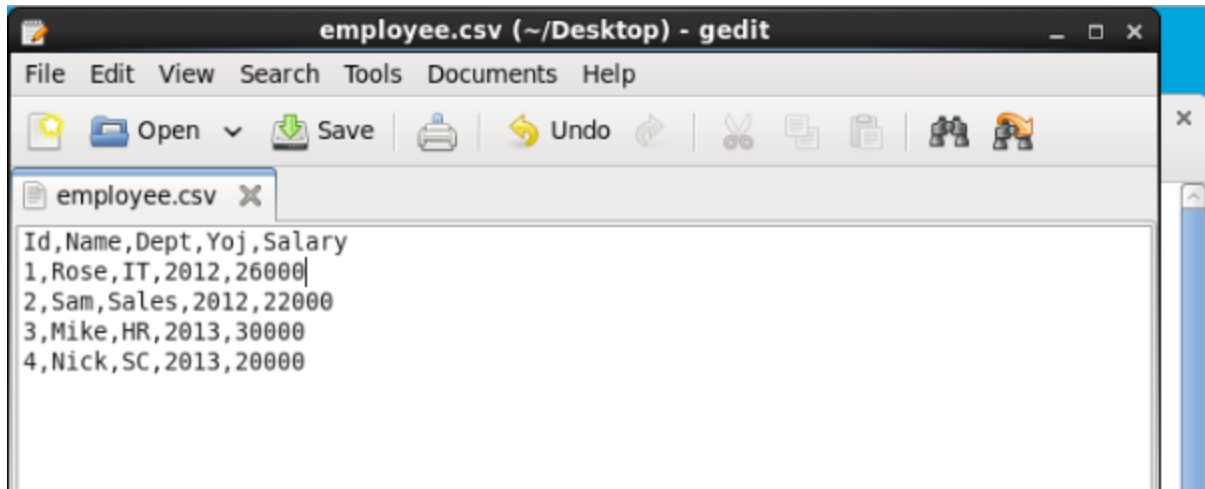
```
hive> create table student(ID int,Name String,Course string,Age int)
> row format delimited
> fields terminated by ','
> tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.217 seconds
hive>
```

Select * from student

```
hive> select * from student
> ;
OK
1      Sam      Hadoop  24
2      Tom      Java    23
3      Harry    Python  23
4      Jerry    Hadoop  NULL
5      John     Python  26
6      Stephen  Java    22
Time taken: 0.19 seconds, Fetched: 6 row(s)
```

Gedit employee.csv

```
cloudera@quickstart:~/Desktop
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ cd Desktop
[cloudera@quickstart Desktop]$ gedit employee.csv
```



Create database hiveeq1

```
hive> create database hiveeq1;  
OK  
Time taken: 0.15 seconds  
hive> █
```

```
hive> create table employee(ID int,Name string,Department string,Yoj int,Salary float)  
> row format delimited  
> fields terminated by ','  
> tblproperties("skip.header.linr.count"="1");  
OK  
Time taken: 0.831 seconds  
hive> █
```

Describe employee

```
Time taken: 0.831 seconds  
hive> describe employee;  
OK  
id                int  
name              string  
department        string  
yoj               int  
salary            float  
Time taken: 0.414 seconds, Fetched: 5 row(s)  
hive> █
```

Select * from employee

```
hive> load data local inpath '/home/cloudera/Desktop/employee.csv' into table employee;
Loading data to table rjcstudent.employee
Table rjcstudent.employee stats: [numFiles=1, totalSize=110]
OK
Time taken: 0.734 seconds
hive> select * from employee;
OK
NULL    Name    Dept    NULL    NULL
1       Rose    IT       2012    26000.0
2       Sam     Sales    2012    22000.0
3       Mike    HR       2013    30000.0
4       Nick    SC       2013    20000.0
Time taken: 0.143 seconds, Fetched: 5 row(s)
```

select * from employee where salary >=25000;

```
Time taken: 0.227 seconds, Fetched: 5 row(s)
hive> select * from employee where salary>=25000;
OK
1       Rose    It       2012    26000.0
3       Mike    Hr       2013    30000.0
Time taken: 0.456 seconds, Fetched: 2 row(s)
hive>
```

select * from employee where salary <=25000;

```
Time taken: 0.456 seconds, Fetched: 2 row(s)
hive> select * from employee where salary<=25000;
OK
2       Sam     Sales    2012    22000.0
4       Nick    Sc       2013    20000.0
Time taken: 0.351 seconds, Fetched: 2 row(s)
hive>
```

Aggregating

41.Arithmetic operations:

select ID, name, salary + 5000 from employee;

```
Time taken: 0.351 seconds, Fetched: 2 row(s)
hive> select ID,name,salary+5000 from employee;
OK
NULL    Name    NULL
1       Rose    31000.0
2       Sam     27000.0
3       Mike    35000.0
4       Nick    25000.0
Time taken: 0.295 seconds, Fetched: 5 row(s)
hive>
```

select max(salary) from employee;

Kaushal Phutane
roll no 19

practical no 7

```
hive> select max(salary)from employee;
Query ID = cloudera_20220321063636_14d659b3-6489-419f-8b62-cde90056bc2f
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_1646550420897_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 06:37:10,875 Stage-1 map = 0%, reduce = 0%
2022-03-21 06:37:31,219 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.12 sec
2022-03-21 06:37:45,854 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.29 sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_1646550420897_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.29 sec HDFS Read: 7930 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 290 msec
OK
```

```
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_1646550420897_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 06:37:10,875 Stage-1 map = 0%, reduce = 0%
2022-03-21 06:37:31,219 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.12 sec
2022-03-21 06:37:45,854 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.29 sec
MapReduce Total cumulative CPU time: 4 seconds 290 msec
Ended Job = job_1646550420897_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.29 sec HDFS Read: 7930 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 290 msec
OK
30000.0
Time taken: 68.323 seconds, Fetched: 1 row(s)
hive>
```

select min(salary) from employee;

```
Time taken: 68.323 seconds, Fetched: 1 row(s)
hive> select min(salary)from employee;
Query ID = cloudera_20220321064242_a9d4da75-3de1-41d2-bbfa-c9d76e102b9c
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_1646550420897_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0019
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 06:42:40,367 Stage-1 map = 0%, reduce = 0%
2022-03-21 06:42:51,562 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.57 sec
2022-03-21 06:43:03,892 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.54 sec
MapReduce Total cumulative CPU time: 3 seconds 540 msec
Ended Job = job_1646550420897_0019
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.54 sec HDFS Read: 8016 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 540 msec
```

Select ID, name, sqrt(salary) from employee;

```
20000.0
Time taken: 37.051 seconds, Fetched: 1 row(s)
hive> select ID,lower(name)from employee;
OK
NULL      name
1         rose
2         sam
3         mike
4         nick
Time taken: 0.08 seconds, Fetched: 5 row(s)
hive> █
```

```
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_1646550420897_0019, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0019/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0019
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 06:42:40,367 Stage-1 map = 0%, reduce = 0%
2022-03-21 06:42:51,562 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.57 sec
2022-03-21 06:43:03,892 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.54 sec
MapReduce Total cumulative CPU time: 3 seconds 540 msec
Ended Job = job_1646550420897_0019
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.54 sec HDFS Read: 8016 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 540 msec
OK
20000.0
Time taken: 37.051 seconds, Fetched: 1 row(s)
hive>
```

Select ID, name, sqrt(salary) from employee;

```
hive> select ID,name,sqrt(salary)from employee;
OK
NULL      Name      NULL
1         Rose      161.24515496597098
2         Sam       148.32396974191326
3         Mike      173.20508075688772
4         Nick       141.4213562373095
Time taken: 0.334 seconds, Fetched: 5 row(s)
```

select ID, upper(name) from employee;

```
hive> select ID,upper(name)from employee;
OK
NULL      NAME
1         ROSE
2         SAM
3         MIKE
4         NICK
Time taken: 0.206 seconds, Fetched: 5 row(s)
```

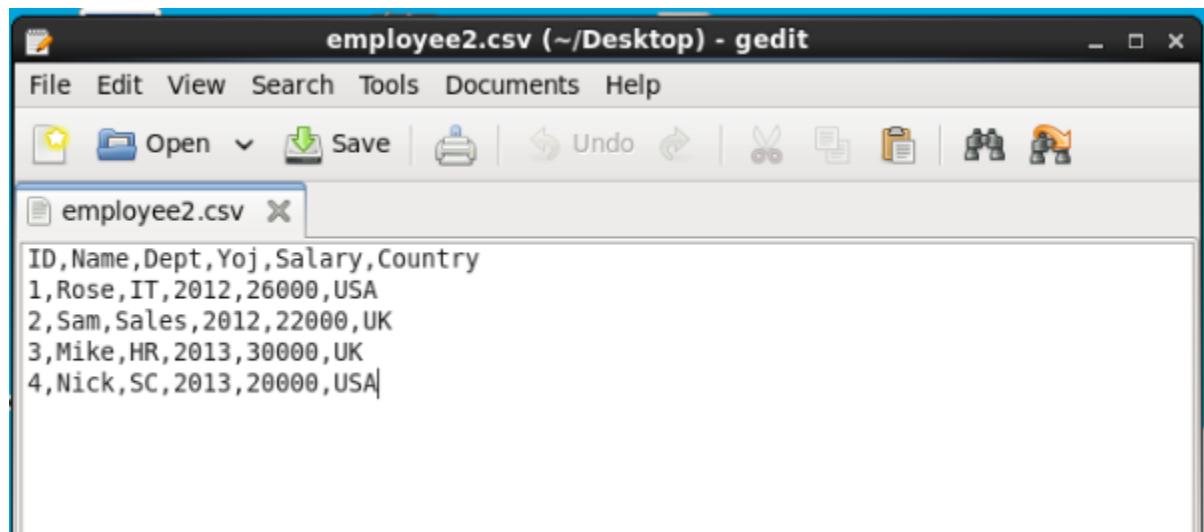
select ID, lower(name) from employee;

Kaushal Phutane
roll no 19

practical no 7

```
Time taken: 0.206 seconds, Fetched: 5 row(s)
hive> select ID,lower(name)from employee;
OK
NULL    name
1       rose
2       sam
3       mike
4       nick
Time taken: 0.165 seconds, Fetched: 5 row(s)
hive>
```

Creating a new table as empgroup.



```
Time taken: 0.206 seconds
hive> create table empgroup(ID int,Name string,Dept string,yoj int,salary float,Country st
ring)
> row format delimited
> fields terminated by ','
> tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.206 seconds
```

select * from empgroup;

```
Time taken: 0.055 seconds
hive> load data local inpath '/home/cloudera/Desktop/employee2.csv' into table empgroup;
Loading data to table default.empgroup
Table default.empgroup stats: [numFiles=1, totalSize=132]
OK
Time taken: 0.379 seconds
hive> select * from empgroup;
OK
1      Rose    It      2012    26000.0 USA
2      Sam     Sales   2012    22000.0 UK
3      Mike    Hr      2013    38000.0 UK
4      Nick    Sc      2013    20000.0 USA
Time taken: 0.067 seconds, Fetched: 4 row(s)
hive>
```

Groupby clause

select country, sum(salary) from empgroup group by country;

```
4      Nick    Sc      2013    20000.0 USA
Time taken: 0.067 seconds, Fetched: 4 row(s)
hive> select country, sum(salary) from empgroup group by country;
Query ID = cloudera_20220321070202_dc7f91cc-5208-4aa7-85fd-48e63ae6459f
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_1646550420897_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 07:02:16,122 Stage-1 map = 0%, reduce = 0%
2022-03-21 07:02:28,232 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.59 sec
2022-03-21 07:02:39,452 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.52 sec
MapReduce Total cumulative CPU time: 3 seconds 520 msec
Ended Job = job_1646550420897_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.52 sec HDFS Read: 8597 HDFS Write: 23 SUCCESS
```



```
File Edit View Search Terminal Help
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_1646550420897_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 07:02:16,122 Stage-1 map = 0%, reduce = 0%
2022-03-21 07:02:28,232 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.59 sec
2022-03-21 07:02:39,452 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.52 sec
MapReduce Total cumulative CPU time: 3 seconds 520 msec
Ended Job = job_1646550420897_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.52 sec HDFS Read: 8597 HDFS Write: 23 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 520 msec
OK
UK      52000.0
USA     46000.0
Time taken: 38.43 seconds, Fetched: 2 row(s)
hive>
```

Groupby clause along with the having clause

select country, sum(salary) from empgroup group by country having sum(salary)>50000;

```
File Edit View Search Terminal Help
hive> select country,sum(salary)from empgroup group by country having sum(salary)>50000;
Query ID = cloudera_20220321070505_1d5211be-3f8f-446b-8275-a706bca4b949
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_1646550420897_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0021
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 07:05:57,625 Stage-1 map = 0%, reduce = 0%
2022-03-21 07:06:08,860 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.58 sec
2022-03-21 07:06:22,012 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.97 sec
MapReduce Total cumulative CPU time: 3 seconds 970 msec
Ended Job = job_1646550420897_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.97 sec HDFS Read: 9099 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 970 msec
OK
```

```
File Edit View Search Terminal Help
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_1646550420897_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0021
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 07:05:57,625 Stage-1 map = 0%, reduce = 0%
2022-03-21 07:06:08,860 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.58 sec
2022-03-21 07:06:22,012 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.97 sec
MapReduce Total cumulative CPU time: 3 seconds 970 msec
Ended Job = job_1646550420897_0021
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.97 sec HDFS Read: 9099 HDFS Write: 11 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 970 msec
OK
UK      52000.0
Time taken: 37.793 seconds, Fetched: 1 row(s)
hive>
```

Sorting : Order by

Select * from empgroup order by salary desc;

```
UK      52000.0
Time taken: 37.793 seconds, Fetched: 1 row(s)
hive> select * from empgroup order by salary desc;
Query ID = cloudera_20220321070808_419ecd0d-b165-441c-a777-ba56e105e830
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_1646550420897_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0022
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 07:08:26,859 Stage-1 map = 0%, reduce = 0%
2022-03-21 07:08:38,059 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.68 sec
2022-03-21 07:08:51,276 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.67 sec
MapReduce Total cumulative CPU time: 3 seconds 670 msec
Ended Job = job_1646550420897_0022
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.67 sec HDFS Read: 8585 HDFS Write: 108 SUCCESS
```

Kaushal Phutane
roll no 19

practical no 7

```
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_1646550420897_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0022
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 07:08:26,859 Stage-1 map = 0%, reduce = 0%
2022-03-21 07:08:38,059 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.68 sec
2022-03-21 07:08:51,276 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.67 sec
MapReduce Total cumulative CPU time: 3 seconds 670 msec
Ended Job = job_1646550420897_0022
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.67 sec HDFS Read: 8585 HDFS Write: 108 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 670 msec
OK
3      Mike    Hr      2013    30000.0 UK
1      Rose    It      2012    26000.0 USA
2      Sam     Sales   2012    22000.0 UK
4      Nick    Sc      2013    20000.0 USA
Time taken: 38.503 seconds, Fetched: 4 row(s)
hive> █
```

Select * from empgroup sort by salary desc;

```
4      Nick    Sc      2013    20000.0 USA
Time taken: 38.503 seconds, Fetched: 4 row(s)
hive> select * from empgroup sort by salary desc;
Query ID = cloudera_20220321071010_71dac3d5-6a60-4cc2-b350-6bf39b3a8f42
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_1646550420897_0023, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1646550420897_0023/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0023
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-21 07:10:44,595 Stage-1 map = 0%, reduce = 0%
2022-03-21 07:10:54,567 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.48 sec
2022-03-21 07:11:08,670 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.46 sec
MapReduce Total cumulative CPU time: 3 seconds 460 msec
Ended Job = job_1646550420897_0023
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.46 sec HDFS Read: 8585 HDFS Write: 108 SUCCESS
```

```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
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_1646550420897_0023, Tracking URL = http://quickstart.cloudera:8088/proxy/applicat  
ion_1646550420897_0023/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1646550420897_0023  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2022-03-21 07:10:44,595 Stage-1 map = 0%, reduce = 0%  
2022-03-21 07:10:54,567 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.48 sec  
2022-03-21 07:11:08,670 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.46 sec  
MapReduce Total cumulative CPU time: 3 seconds 460 msec  
Ended Job = job_1646550420897_0023  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.46 sec HDFS Read: 8585 HDFS Write: 108 SUCCE  
S  
Total MapReduce CPU Time Spent: 3 seconds 460 msec  
OK  
3 Mike Hr 2013 30000.0 UK  
1 Rose It 2012 26000.0 USA  
2 Sam Sales 2012 22000.0 UK  
4 Nick Sc 2013 20000.0 USA  
Time taken: 37.319 seconds, Fetched: 4 row(s)  
hive>
```