

19162121031

SMIT R PATEL

SEM 5

BDA

PRACTICAL 11

AIM-To understand the working of Basic Hive commands.

Exercise:

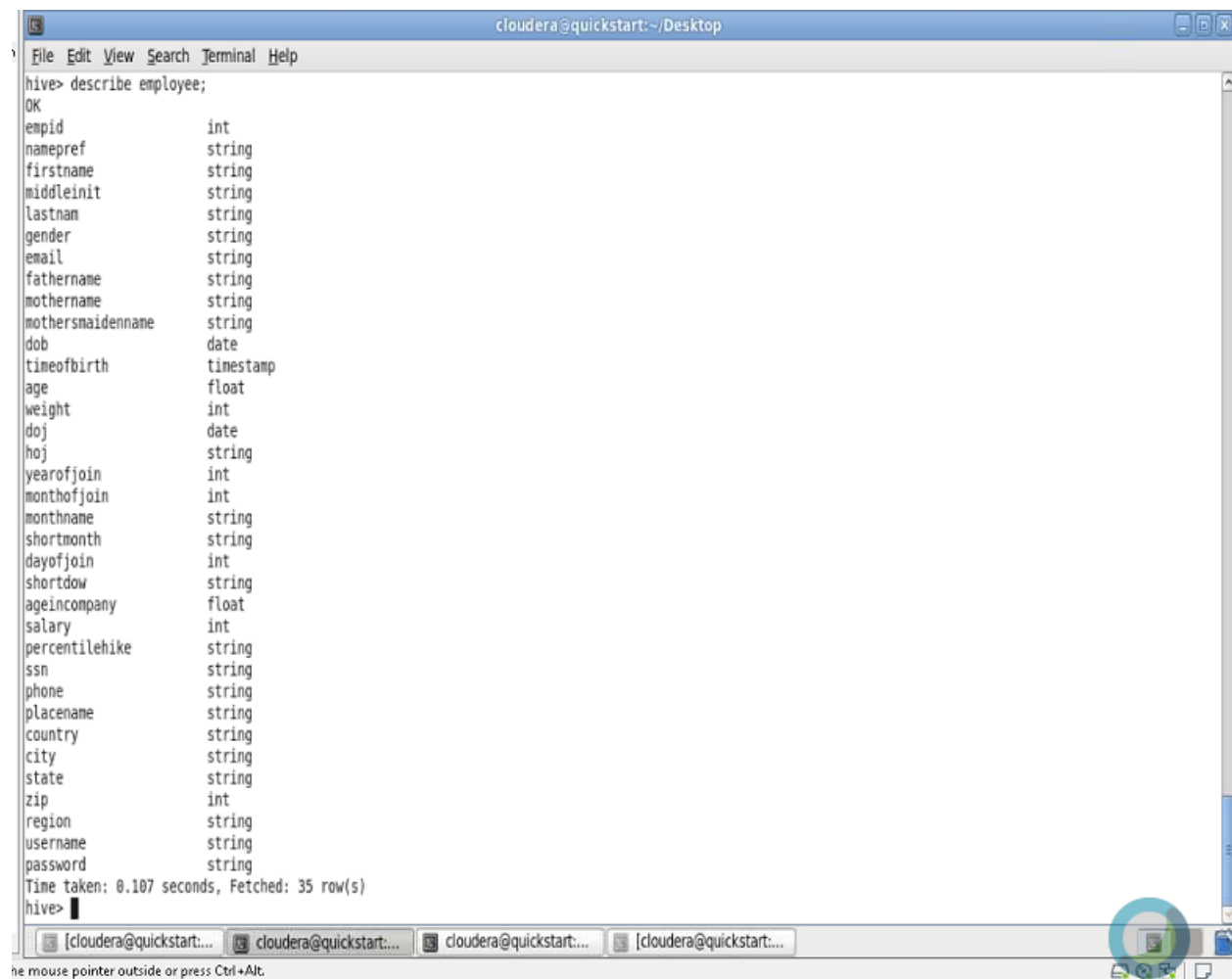
You work for the HR department in a Multi-National Company. You notice that Excel is taking a long time to open the file, hence you are suggested by a Co-worker to use Hive in Hadoop since you are okay with working in SQL like queries. You need to access some information for your employees. Do the following:

Tasks:

```
[cloudera@quickstart Desktop]$ hadoop fs -ls p11
Found 3 items
-rw-r--r-- 1 cloudera cloudera 2775303 2021-09-16 00:30 p11/10000-Records.csv
-rw-r--r-- 1 cloudera cloudera 1247263 2021-09-14 00:15 p11/10000-Sales-Records.csv
-rw-r--r-- 1 cloudera cloudera 1247263 2021-09-14 00:17 p11/10000_Sale_Records.csv
[cloudera@quickstart Desktop]$
```

1. Create a table "employee" and move your employee records into it.

```
hive> create table employee
> (empid INT, namepref STRING, firstname STRING, middleinit STRING, lastnam STRING, gender STRING, email STRING, fathernam STRING, mothername STRING, MothersMaidenName STRING, DOB DATE, timeOFBIRTH TIMESTAMP, AGE FLOAT, weight INT, DOJ DATE, HOJ STRING, YEAROFJOIN INT, MONTHOFJOIN INT, MONTHNAME STRING, SHORTMONTH STRING, DAYOFJOIN INT, SHORTDOW STRING, AGEINCOMPANY FLOAT, SALARY INT, PERCENTILEHIKE STRING, SSN STRING, phone STRING, PLACENAME STRING, COUNTRY STRING, City STRING, State STRING, Zip INT, Region STRING, UserName STRING, Password STRING )
> row format delimited fields terminated by ',';
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table employee already exists)
hive>
```



```
cloudera@quickstart: ~/Desktop
File Edit View Search Terminal Help
hive> describe employee;
OK
empid          int
namepref       string
firstname      string
middleinit     string
lastnam        string
gender         string
email          string
fathernam      string
mothername     string
MothersMaidenName string
DOB            date
timeOFBIRTH    timestamp
AGE            float
weight         int
DOJ            date
HOJ            string
YEAROFJOIN     int
MONTHOFJOIN    int
MONTHNAME      string
SHORTMONTH     string
DAYOFJOIN      int
SHORTDOW       string
AGEINCOMPANY   float
salary         int
percentilehike string
ssn            string
phone          string
placename      string
country        string
city           string
state          string
zip            int
region         string
username       string
password       string
Time taken: 0.107 seconds, Fetched: 35 row(s)
hive>
```

2. Display employee information where salary>25000.

```
hive> select firstname, lastnam , gender, salary from employee where salary>25000 limit 15;
OK
Time taken: 0.082 seconds
hive> █
```

3. Create another table “orders” and move your order data into it.

```
File Edit View Search Terminal Help
hive> create table order
> (Region STRING, Country STRING, ItemType STRING, SalesChannel STRING, OrderPriority STRING, ORDERDATE DATE, ORDERID INT, SHIPDATE DATE, UNITSSOLD INT,
UNITPRICE FLOAT, UNITCOST FLOAT , TOTALREVENUE FLOAT , TOTALCOST FLOAT, TOTALPROFIT FLOAT)
> row format delimited fields terminated by ',' ;
OK
Time taken: 0.435 seconds
```

```
hive> describe order;
OK
region                string
country               string
itemtype              string
saleschannel          string
orderpriority         string
orderdate             date
orderid               int
shipdate              date
unitssold             int
unitprice             float
unitcost              float
totalrevenue          float
totalcost             float
totalprofit           float
Time taken: 0.105 seconds, Fetched: 14 row(s)
hive> █
```

Use mouse pointer outside of press Ctrl+Alt

```
File Edit View Search Terminal Help
hive> LOAD DATA LOCAL INPATH '10000_Sale_Records.csv' INTO TABLE order;
Loading data to table default.order
Table default.order stats: [numFiles=1, totalSize=1247263]
OK
Time taken: 0.612 seconds
hive>
> select * from order where 10;
OK
Failed with exception java.io.IOException:java.lang.ClassCastException: java
.lang.Integer cannot be cast to java.lang.Boolean
Time taken: 0.443 seconds
hive> select * from order limit 10;
OK
Region Country Item Type Sales Channel Order Priority NULL NULL
NULL NULL NULL NULL NULL NULL
Sub-Saharan Africa Chad Office Supplies Online L NULL 2924
94523 NULL 4484 651.21 524.96 2920025.8 2353920.8 5661
05.0
Europe Latvia Beverages Online C NULL 361825549 NULL
1075 47.45 31.79 51008.75 34174.25 16834.5
Middle East and North Africa Pakistan Vegetables Offline C N
ULL 141515767 NULL 6515 154.06 90.93 1003700.9 5924
08.94 411291.94
Sub-Saharan Africa Democratic Republic of the Congo Household 0
nline C NULL 500364005 NULL 7683 668.27 502.54 5134
318.5 3861014.8 1273303.6
Europe Czech Republic Beverages Online C NULL 127481591 N
ULL 3491 47.45 31.79 165647.95 110978.89 54669.06
Sub-Saharan Africa South Africa Beverages Offline H NULL
482292354 NULL 9880 47.45 31.79 468806.0 314085.2 1
54720.8
Asia Laos Vegetables Online L NULL 844532620 NULL
4825 154.06 90.93 743339.5 438737.25 304602.25
Asia China Baby Food Online C NULL 564251220 NULL
3330 255.20 159.42 850082.4 530060.6 319213.8
Sub-Saharan Africa Eritrea Meat Online L NULL 411809480 N
ULL 2431 421.89 364.69 1025614.56 886561.4 139053.2
Time taken: 0.097 seconds, Fetched: 10 row(s)
```

4. Try joining the two tables together.

→ We can not join two tables together because

- No column same in both table.
- No find primary key in both table.
- No find unique column in both table.

So , here we can not join two tables together.