

1. Install Hive on the system using the distribution provided.

Initialize schema:-

schematool -dbType derby -initSchema

>hive

```
Activities Terminal
osboxes@osboxes: ~$ start-dfs.sh
Starting namenodes on [localhost]
Starting secondary namenodes [osboxes]
osboxes@osboxes: ~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
osboxes@osboxes: ~$ jps
3344 NodeManager
2480 NameNode
3493 Jps
3001 ResourceManager
2810 SecondaryNameNode
2557 DataNode
osboxes@osboxes: ~$ schematool -dbType derby -initSchema
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/osboxes/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Metastore connection URL: jdbc:derby:databaseName=metastore_db;create=true
Metastore Connection Driver : org.apache.derby.jdbc.EmbeddedDriver
Metastore Connection User: APP
Starting metastore schema initialization to 3.1.0
Initialization script hive-schema-3.1.0.derby.sql

Error: FUNCTION 'NUCLEUS_ASCII' already exists. (state=X0Y68,code=30000)
org.apache.hadoop.hive.metastore.HiveMetaException: Schema initialization FAILED! Metastore state would be inconsistent !!
Underlying cause: java.io.IOException : Schema script failed, errorcode 2
Use --verbose for detailed stacktrace.
*** schemaTool failed ***

osboxes@osboxes: ~$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/osboxes/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = e4ccb47-45c7-4ae4-9d0a-6772cf4a8626

Logging initialized using configuration in jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 2e9d388-8274-43f7-8e89-aa4ca3a6ec0c
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

2. Create two databases with the names ‘retail_db’ and ‘bank_db’ having comments and located at /user/hive/mywarehouse/.

Create database retail_db

Create database bnk_db

```
Activities Terminal
osboxes@osboxes: ~$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/osboxes/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = 6780588d-fc89-4ab3-9b70-417739883f04

Logging initialized using configuration in jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 39df10e-5d3c-4c77-b2b5-5e14834e18fe
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> drop database retail_db
> ;
OK
Time taken: 0.488 seconds
hive> create database retail_db
> comment 'database for assignments'
> location '/user/hive/mywarehouse';
OK
Time taken: 0.882 seconds
hive> create database bank_db
> comment 'database for assignments'
> location '/user/hive/mywarehouse';
OK
Time taken: 0.018 seconds
hive> show databases
> ;
OK
bank_db
default
retail_db
Time taken: 0.115 seconds, Fetched: 3 row(s)
hive>
```

3. Create a hive managed table ‘employee’ in ‘bank_db’ database with the following structure:

Create table if not exists bank_db.employee (eid int, name string, salary float, designation string)

Comment 'table for employees inside bank_db'
Row format delimited
Fields terminated by ','
Lines terminated by '\n'
Stored as textfile;

```

osboxes@osboxes:~$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/osboxes/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = 5589add8-401f-41d9-8247-a2f68908b41d
Logging initialized using configuration in jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = bc1ce79c-f6af-4a83-a67a-a7f0571ab55d
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> create table if not exists bank_db.employee ( eid int, Name string, Salary float, Designation string)
> comment 'table for employees inside bank_db'
> row format delimited
> fields terminated by ','
> lines terminated by '\n'
> stored as textfile;
OK
Time taken: 0.814 seconds
hive>

```

4. Create external table 'new_employee' with the same structure at HDFS location /public/bank_db

Create external table if not exists new_employee (eid int, name string, salary float, designation string)

Comment 'table '
Row format delimited
Fields terminated by ','
Lines terminated by '\n'
Location '/user/public/bank_db'

```

osboxes@osboxes:~$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/osboxes/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = a0d347df-f93a-42ad-897a-9cf2309edf41
Logging initialized using configuration in jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 78b8833e-1d1b-42e8-86a9-086c28f62d51
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> create external table if not exists new_employee ( eid int, Name string, Salary float, Designation string)
> comment 'table '
> row format delimited
> fields terminated by ','
> lines terminated by '\n'
> location '/user/public/bank_db';
OK
Time taken: 0.872 seconds
hive>

```

5. Rename table 'new_employee' to 'all_employee'.

Alter table new_employee rename to all_employee;

```
Activities Terminal Jan 11 13:05
osboxes@osboxes: -
osboxes@osboxes: -
Hive Session ID = ad4347df-f93a-43ad-897a-8cf23698df44
Logging initialized using configuration in jar:file:/home/osboxes/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 7fb9833c-1d1b-43e8-8e0f-69062bf67d51
Hive-on-HR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> create external table if not exists new_employee ( Eid int, Name string, Salary float, Designation string)
> comment 'table for new_employees inside hdfs local location'
> row format delimited
> fields terminated by ','
> lines terminated by '\n'
> location '/user/public/bank_db';
OK
Time taken: 0.572 seconds
hive> alter table new_employee to all_employee;
NovTableAltException(3128[])
at org.apache.hadoop.hive.ql.parse.HiveParser.alterTableStatementSuffix(HiveParser.java:7971)
at org.apache.hadoop.hive.ql.parse.HiveParser.alterStatement(HiveParser.java:7447)
at org.apache.hadoop.hive.ql.parse.HiveParser.ddlStatement(HiveParser.java:4337)
at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:2494)
at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1420)
at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:220)
at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:74)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:67)
at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:616)
at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1826)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:1773)
at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:1768)
at org.apache.hadoop.hive.ql.rexec.ReExecDriver.compileAndRespond(ReExecDriver.java:126)
at org.apache.hadoop.hive.ql.rexec.ReExecDriver.run(ReExecDriver.java:214)
at org.apache.hadoop.hive.ql.rexec.ReExecDriver.run(ReExecDriver.java:239)
at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:188)
at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:402)
at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:683)
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.main(RunJar.java:236)
FAILED: ParseException line 1:25 cannot recognize input near 'to' 'all_employee' '<EOF>' in alter table statement
hive> alter table new_employee rename to all_employee;
OK
Time taken: 0.23 seconds
hive>
```

**6. Apply the following changes to ‘employee’ and ‘new_employee’ tables:
Describe bank_db.employee;
Alter table all_employee change designation designation varchar(30);
Alter table bank_db.employee change name emp_name string;
Alter table bank_db.employee change salary salary double;
Alter table bank_db.employee change designation designation varchar(30);**

```
Activities Terminal Jan 11 13:16
osboxes@osboxes: -
osboxes@osboxes: -
OK
Time taken: 0.088 seconds
hive> alter table all_employee change designation designation varchar(30);
OK
Time taken: 0.076 seconds
hive> describe all_employee;
OK
emp_id          int
emp_name        string
salary         double
designation      varchar(30)
Time taken: 0.049 seconds, Fetched: 4 row(s)
hive> describe bank_db.employee;
OK
eid            int
name           string
salary         float
designation     string
Time taken: 0.032 seconds, Fetched: 4 row(s)
hive> alter table bank_db.employee change eid emp_id int;
OK
Time taken: 0.081 seconds
hive> alter table bank_db.employee change name emp_name string;
OK
Time taken: 0.095 seconds
hive> alter table bank_db.employee change salary salary double;
OK
Time taken: 0.07 seconds
hive> alter table bank_db.employee change designation designation varchar(30);
OK
Time taken: 0.075 seconds
hive> describe bank_db.employee;
OK
emp_id          int
emp_name        string
salary         double
designation      varchar(30)
Time taken: 0.032 seconds, Fetched: 4 row(s)
hive> describe all_employee;
OK
emp_id          int
emp_name        string
salary         double
designation      varchar(30)
Time taken: 0.028 seconds, Fetched: 4 row(s)
hive>
```

**7. Add two column to the above table:
Column Name: Dt_of_Joining**

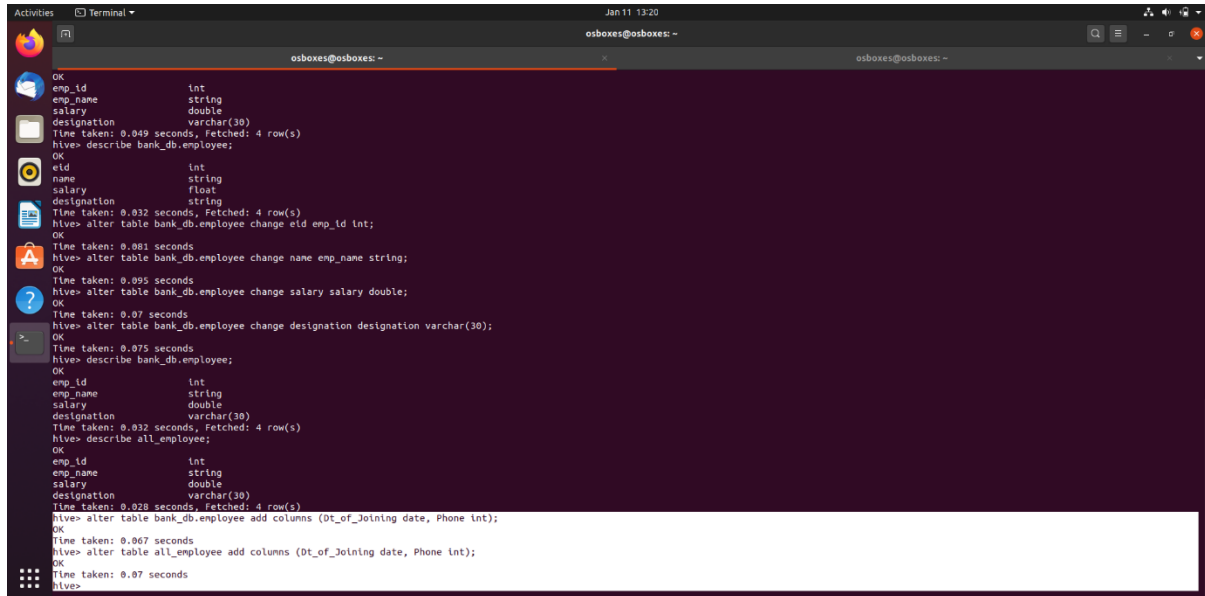
Data Type: DATE

Column Name: Phone

Data Type: INT

Alter table bank_db.employee add columns (dt_of_joining date, phone int);

Alter table all_employee add cloumns (dt_of_joining date, phone int);



```
osboxes@osboxes: -
OK
emp_id      int
emp_name    string
salary      double
designation  varchar(30)
Time taken: 0.049 seconds, Fetched: 4 row(s)
hive> describe bank_db.employee;
OK
eid          int
name         string
salary       float
designation   string
Time taken: 0.032 seconds, Fetched: 4 row(s)
hive> alter table bank_db.employee change eid emp_id int;
OK
Time taken: 0.081 seconds
hive> alter table bank_db.employee change name emp_name string;
OK
Time taken: 0.095 seconds
hive> alter table bank_db.employee change salary salary double;
OK
Time taken: 0.07 seconds
hive> alter table bank_db.employee change designation designation varchar(30);
OK
Time taken: 0.075 seconds
hive> describe bank_db.employee;
OK
emp_id      int
emp_name    string
salary      double
designation  varchar(30)
Time taken: 0.032 seconds, Fetched: 4 row(s)
hive> describe all_employee;
OK
emp_id      int
emp_name    string
salary      double
designation  varchar(30)
Time taken: 0.028 seconds, Fetched: 4 row(s)
hive> alter table bank_db.employee add columns (Dt_of_Joining date, Phone int);
OK
Time taken: 0.067 seconds
hive> alter table all_employee add columns (Dt_of_Joining date, Phone int);
OK
Time taken: 0.07 seconds
hive>
```

9. Create table in hive that can hold the following records. Choose appropriate field names and types.

Ajay, Lumia 1020, Nokia, 10000

Shiva, iphone6, Apple, 34000

Srejeeth, Galaxy 4, Samsung, 20000

Create table if not exists (name string. Model string, brand string, price int)

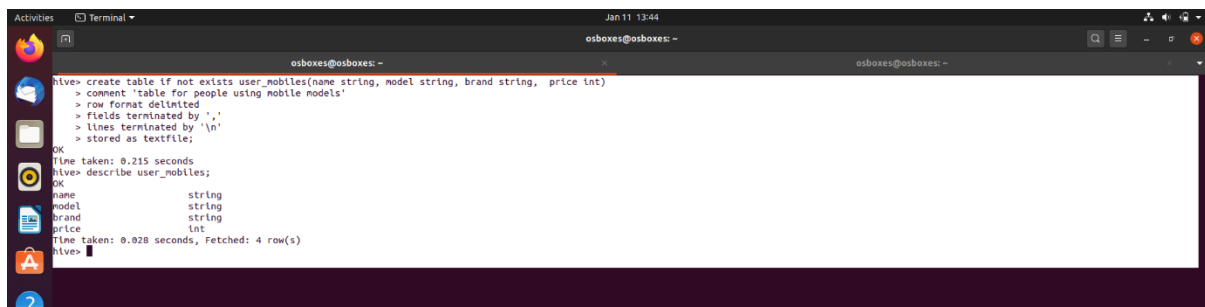
Comment 'table'

Row format delimited

Fields terminated by ','

Lines terminated by '\n'

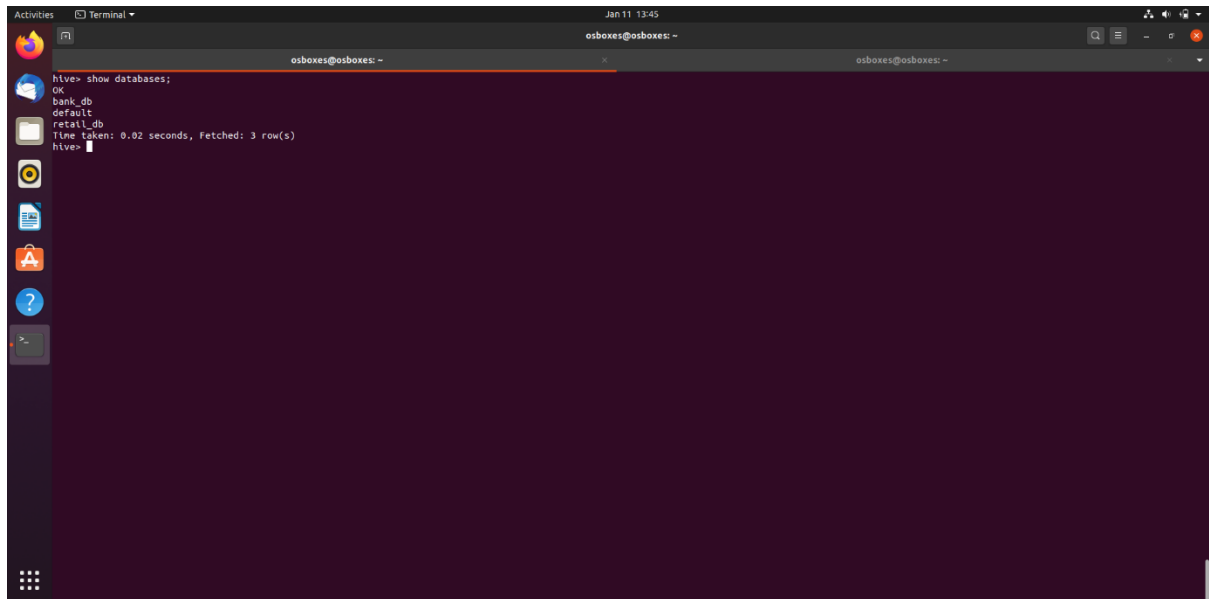
Stored as textfile;



```
osboxes@osboxes: -
hive> create table if not exists user_mobiles(name string, model string, brand string, price int)
> comment 'table for people using mobile models'
> row format delimited
> fields terminated by ','
> lines terminated by '\n'
> stored as textfile;
OK
Time taken: 0.215 seconds
hive> describe user_mobiles;
OK
name        string
model       string
brand       string
price       int
Time taken: 0.028 seconds, Fetched: 4 row(s)
hive>
```

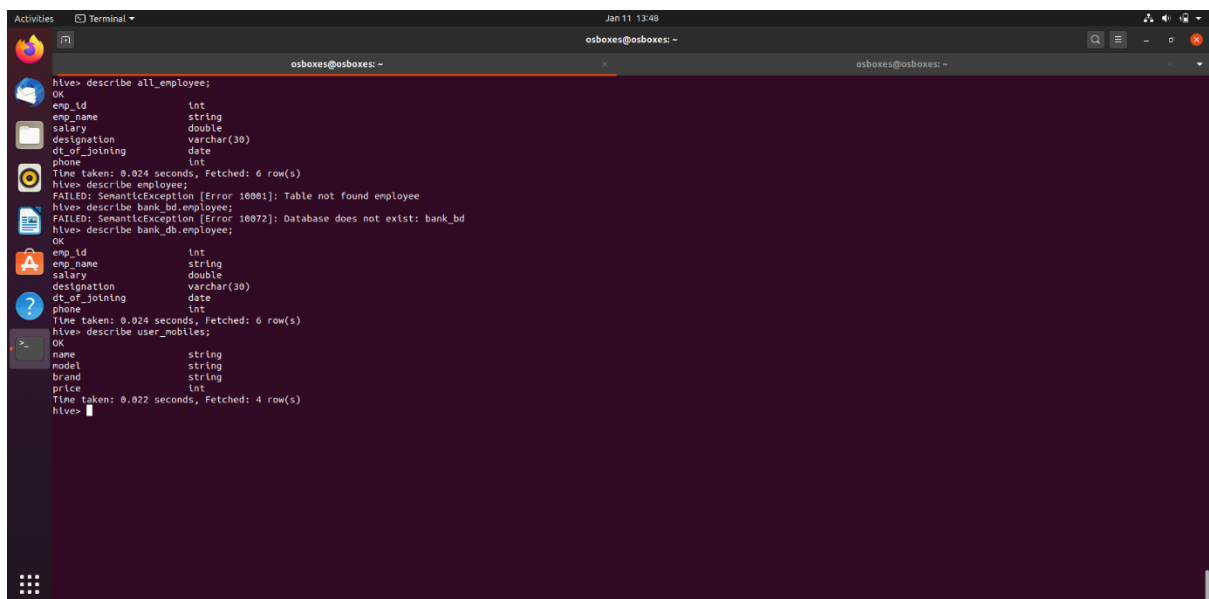
10. Apply 'show' DDL command for the databases and tables created.

Show databases;



```
osboxes@osboxes: -  
hive> show databases;  
OK  
bank_db  
default  
retail_db  
Time taken: 0.02 seconds, Fetched: 3 row(s)  
hive>
```

11. Apply ‘describe’ DDL command for the databases and tables created.
Describe all_employees;



```
osboxes@osboxes: -  
hive> describe all_employee;  
OK  
emp_id      int  
emp_name    string  
salary      double  
designation  varchar(30)  
dt_of_joining date  
phone       int  
Time taken: 0.024 seconds, Fetched: 6 row(s)  
hive> describe employee;  
FAILED: SemanticException [Error 10001]: Table not found employee  
hive> describe bank_db.employee;  
FAILED: SemanticException [Error 10072]: Database does not exist: bank_db  
hive> describe bank_db.employee;  
OK  
emp_id      int  
emp_name    string  
salary      double  
designation  varchar(30)  
dt_of_joining date  
phone       int  
Time taken: 0.024 seconds, Fetched: 6 row(s)  
hive> describe user_mobiles;  
OK  
name        string  
model       string  
brand       string  
price       int  
Time taken: 0.022 seconds, Fetched: 4 row(s)  
hive>
```

12. Create all 6 tables listed in figure in hive as managed tables, delimiter is "|", in database 'retail_db'.

```
hive> create table if not exists categories(category_id int, category_department_id int, category_name varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
Time taken: 0.22 seconds
hive> create table if not exists products1(products_id int, product_category_id int, product_name varchar(45), product_description varchar(255), product_price float, product_image varchar(255))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
Time taken: 0.188 seconds
hive> create table if not exists order_items(order_item_id int, order_item_order_id int, order_item_product_id int, order_item_quantity int, order_item_subtotal float, order_item_product_price float)
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
Time taken: 0.158 seconds
hive> create table if not exists orders(order_id int, order_date datetime, order_customer_id int, order_status varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
```

```
hive> create table if not exists departments1(department_id int, department_name varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
Time taken: 0.105 seconds
```

```
hive> create table if not exists orders(order_id int, order_date datetime, order_customer_id int, order_status varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
FAILED: SemanticException [Error 10099]: DATETIME type isn't supported yet. Please use DATE or TIMESTAMP instead
hive> create table if not exists orders(order_id int, order_date date, order_customer_id int, order_status varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
Time taken: 0.116 seconds
hive> create table if not exists customers(customer_id int, customer_fname varchar(45), customer_lname varchar(45), customer_email varchar(45), customer_password varchar(45), customer_street varchar(45), customer_city varchar(45), customer_state varchar(45), customer_zipcode varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
Time taken: 0.124 seconds
```

```
hive> describe order_items;
OK
order_item_id          int
order_item_order_id    int
order_item_product_id  int
order_item_quantity    int
order_item_subtotal    float
order_item_product_price float
Time taken: 0.107 seconds, Fetched: 6 row(s)
hive> describe orders;
OK
order_id               int
order_date             date
order_customer_id      int
order_status           varchar(45)
Time taken: 0.108 seconds, Fetched: 4 row(s)
hive> describe customers;
OK
customer_id           int
customer_fname        varchar(45)
customer_lname        varchar(45)
customer_email        varchar(45)
customer_password     varchar(45)
customer_street       varchar(45)
customer_city         varchar(45)
customer_state        varchar(45)
customer_zipcode      varchar(45)
Time taken: 0.109 seconds, Fetched: 9 row(s)
hive>
```

```
hive> describe departments1;
OK
department_id          int
department_name        varchar(45)
Time taken: 0.124 seconds, Fetched: 2 row(s)
hive> describe categories;
OK
category_id            int
category_department_id int
category_name          varchar(45)
Time taken: 0.107 seconds, Fetched: 3 row(s)
hive> describe products;
FAILED: SemanticException [Error 10001]: Table not found products
hive> describe products1;
OK
products_id            int
product_category_id    int
product_name           varchar(45)
product_description    varchar(255)
product_price          float
product_image          varchar(255)
Time taken: 0.071 seconds, Fetched: 6 row(s)
```

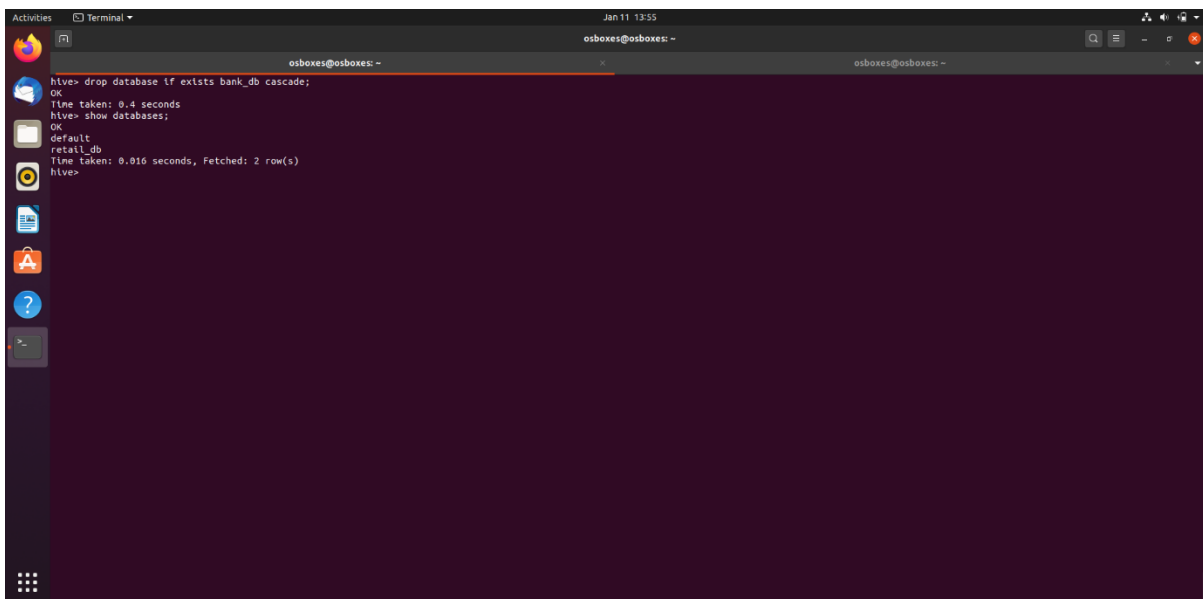
13. Drop tables 'order_items' and 'orders'.

```
hive> Drop table order_items;
OK
Time taken: 2.329 seconds
hive> Drop table orders;
OK
Time taken: 0.309 seconds
hive> █
```

14. Create the above two tables again as external tables at a location different from rest of the tables.

```
hive> create external table if not exists customers(customer_id int, customer_f
name varchar(45), customer_lname varchar(45), customer_email varchar(45), custo
mer_password varchar(45), customer_street varchar(45), customer_city varchar(45
), customer_state varchar(45), customer_zipcode varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
Time taken: 0.071 seconds
hive> create external table if not exists orders(order_id int, order_date date,
order_customer_id int, order_status varchar(45))
> row format delimited
> fields terminated by '|'
> lines terminated by '\n';
OK
```

**15. Drop database 'bank_db' along with all the tables in it in one command.
Drop database if exists bank_db cascade;**



The screenshot shows a terminal window with the following commands and output:

```
osboxes@osboxes: ~
hive> drop database if exists bank_db cascade;
OK
Time taken: 0.4 seconds
hive> show databases;
OK
default
retail_db
Time taken: 0.016 seconds, Fetched: 2 row(s)
hive>
```