

BIG DATA TECHNOLOGIES LAB

(ACADEMIC YEAR : 2017-2018)

I SEMESTER

ASSIGNMENT 5

TOPIC: APACHE HIVE - DDL

ASHISH CHANDRAKANT DUSANE

M. TECH. (ACDS)

COMPUTER ENGG. DEPARTMENT

{ PRN : 170101261004 }

~: Apache Hive – DDL :~

1. Install Hive on the system using the distribution provided.
2. Create two databases with the names 'retail_db' and 'bank_db' having comments and located at /user/hive/mywarehouse/.

```
Time taken: 0.032 seconds
hive> create database retail_db;
OK
Time taken: 0.529 seconds
hive> show databases;
OK
default
retail_db
Time taken: 0.06 seconds, Fetched: 2 row(s)
hive> create database bank_db;
OK
Time taken: 0.074 seconds
hive> show databases;
OK
bank_db
default
retail_db
Time taken: 0.014 seconds, Fetched: 3 row(s)
```

3. Create a hive managed table 'employee' in 'bank_db' database with the following

Field Name	Data Type
EID	INT
Name	STRING
Salary	FLOAT
Designation	STRING

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
Time taken: 0.016 seconds
hive> CREATE TABLE employee(EID int,Name string,Salary float,Designation string);
OK
Time taken: 0.669 seconds
hive> desc employee;
OK
eid                int
name               string
salary            float
designation        string
Time taken: 0.257 seconds, Fetched: 4 row(s)
hive> |
```

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

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> CREATE EXTERNAL TABLE new_employee(EID int,Name string,Salary float,Designation string);
OK
Time taken: 0.522 seconds
hive> show tables;
OK
employee
new_employee
Time taken: 0.056 seconds, Fetched: 2 row(s)
hive> desc new_employee;
OK
eid                int
name               string
salary            float
designation        string
Time taken: 0.071 seconds, Fetched: 4 row(s)
hive>
```

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

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> show tables;
OK
employee
new_employee
Time taken: 0.056 seconds, Fetched: 2 row(s)
hive> desc new_employee;
OK
eid                int
name               string
salary            float
designation        string
Time taken: 0.071 seconds, Fetched: 4 row(s)
hive> ALTER TABLE new_employee RENAME TO all_employee;
OK
Time taken: 0.127 seconds
hive> show tables;
OK
all_employee
employee
Time taken: 0.016 seconds, Fetched: 2 row(s)
hive>
```

6. Apply the following changes to 'employee' and 'all_employee' tables:

Field Name	Convert From Datatype	New Field Name	New Datatype
EID	INT	Emp_ID	INT
Name	STRING	Emp_Name	STRING
Salary	FLOAT	Salary	DOUBLE
Designation	STRING	Designation	VARCHAR(30)

Employee :~

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> ALTER TABLE employee CHANGE EID Emp_ID int;
OK
Time taken: 0.308 seconds
hive> ALTER TABLE employee CHANGE Name Emp_Name string;
OK
Time taken: 0.16 seconds
hive> ALTER TABLE employee CHANGE Salary Salary double;
OK
Time taken: 0.115 seconds
hive> ALTER TABLE employee CHANGE Designation Designation varchar(30);
OK
Time taken: 0.127 seconds
hive> desc employee;
OK
emp_id          int
emp_name        string
salary          double
designation      varchar(30)
Time taken: 0.083 seconds, Fetched: 4 row(s)
hive>
```

all_employee :~

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> ALTER TABLE all_employee CHANGE EID Emp_ID int;
OK
Time taken: 0.108 seconds
hive> ALTER TABLE all_employee CHANGE Name Emp_Name string;
OK
Time taken: 0.092 seconds
hive> ALTER TABLE all_employee CHANGE Salary Salary double;
OK
Time taken: 0.107 seconds
hive> ALTER TABLE all_employee CHANGE Designation Designation varchar(30);
OK
Time taken: 0.112 seconds
hive> desc all_employee;
OK
emp_id          int
emp_name        string
salary          double
designation      varchar(30)
Time taken: 0.046 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

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> ALTER TABLE employee ADD COLUMNS (Dt_of_Joining date, Phone int);
OK
Time taken: 0.101 seconds
hive> ALTER TABLE all_employee ADD COLUMNS (Dt_of_Joining date, Phone int);
OK
Time taken: 0.128 seconds
hive> desc employee;
OK
emp_id          int
emp_name        string
salary          double
designation      varchar(30)
dt_of_joining   date
phone           int
Time taken: 0.041 seconds, Fetched: 6 row(s)
hive> desc all_employee;
OK
emp_id          int
emp_name        string
salary          double
designation      varchar(30)
dt_of_joining   date
phone           int
Time taken: 0.077 seconds, Fetched: 6 row(s)
hive>
```

8. Drop column 'Salary' from the above table.

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, 2000010.

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> CREATE TABLE mobile(Cust_Name string, Mobile_Type varchar(20),Company varchar(20),Price int);
OK
Time taken: 0.118 seconds
hive> desc mobile;
OK
cust_name       string
mobile_type     varchar(20)
company         varchar(20)
price           int
Time taken: 0.035 seconds, Fetched: 4 row(s)
hive> show databases;
```

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

```
hive> insert into mobile(cust_name, mobile_type, company, price) VALUES ("Ajay",  
"Lumia 1020","Nokia",10000);  
Query ID = student_20171222104955_723c9a9a-9c77-4722-995f-bf8cc17165e1  
Total jobs = 3  
Launching Job 1 out of 3  
Number of reduce tasks is set to 0 since there's no reduce operator
```

```
hive> insert into mobile(cust_name, mobile_type, company, price) VALUES ("Shiva",  
"iphone6","Apple",34000);  
Query ID = student_20171222105050_569e1a93-22a1-4573-a10f-b003265a682f  
Total jobs = 3  
Launching Job 1 out of 3  
Number of reduce tasks is set to 0 since there's no reduce operator
```

```
hive> insert into mobile(cust_name, mobile_type, company, price) VALUES ("Srejee  
th","Galaxy 4","Samsung",2000010);  
Query ID = student_20171222105151_d461d5e9-7d01-4ac0-ae00-74321432930f  
Total jobs = 3  
Launching Job 1 out of 3  
Number of reduce tasks is set to 0 since there's no reduce operator
```

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin  
hive> select * from mobile;  
OK  
Ajay      Lumia 1020      Nokia    10000  
Shiva     iphone6 Apple     34000  
Srejeeth      Galaxy 4      Samsung 2000010  
Time taken: 0.174 seconds, Fetched: 3 row(s)  
hive>
```

11. Apply 'describe' DDL command for the databases and tables created.

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin  
hive> show databases;  
OK  
bank_db  
default  
retail_db  
Time taken: 0.019 seconds, Fetched: 3 row(s)  
hive> show tables;  
OK  
all_employee  
employee  
mobile  
values__tmp__table__1  
values__tmp__table__2  
values__tmp__table__3  
Time taken: 0.021 seconds, Fetched: 6 row(s)  
hive> desc mobile;  
OK  
cust_name      string  
mobile_type    varchar(20)  
company        varchar(20)  
price          int  
Time taken: 0.031 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 departments(department_id int,department_name varchar(45))  
row format delimited fields terminated by '|';  
OK  
Time taken: 0.104 seconds
```

```
hive> create TABLE categories(category_id int,category_department_id int,category_name varchar(45)) row format delimited fields terminated by '|';  
OK  
Time taken: 0.067 seconds
```

```
hive> create TABLE products(product_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 '|';  
OK  
Time taken: 0.073 seconds
```

```
hive> create TABLE order_items(order_item_id int,order_item_order_id int,order_item_product_id int,order_item_quantity tinyint,order_item_subtotal float,order_item_product_price float) row format delimited fields terminated by '|';  
OK  
Time taken: 0.154 seconds
```

```
hive> create TABLE orders(order_id int,order_date date,order_customer_id int,order_status varchar(45)) row format delimited fields terminated by '|';  
OK  
Time taken: 0.073 seconds
```

```
hive> create TABLE customers(customer_id int,customer_fname varchar(45),customer_lname varchar(45),customer_email varchar(45),customer_password varchar(45),customer_street varchar(255),customer_city varchar(45),customer_state varchar(45),customer_zipcode varchar(45)) row format delimited fields terminated by '|';  
OK  
Time taken: 0.095 seconds  
hive> █
```


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

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> DROP TABLE orders;
OK
Time taken: 3.618 seconds
hive> DROP TABLE order_items;
OK
Time taken: 0.159 seconds
hive> show tables;
OK
categories
customers
departments
products
Time taken: 0.026 seconds, Fetched: 4 row(s)
hive>
```

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

```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> create external TABLE orders(order_id int,order_date date,order_customer
_id int,order_status varchar(45)) row format delimited fields terminated by '|'
';
OK
Time taken: 0.128 seconds
hive> create external TABLE order_items(order_item_id int,order_item_order_id
int,order_item_product_id int,order_item_quantity tinyint,order_item_subtotal
float,order_item_product_price float) row format delimited fields terminated b
y '|';
OK
Time taken: 0.067 seconds
hive> show tables;
OK
categories
customers
departments
order_items
orders
products
Time taken: 0.01 seconds, Fetched: 6 row(s)
hive>
```


15. Drop database 'bank_db' along with all the tables in it in one command.



```
student@student-OptiPlex-3020: ~/Documents/apache-hive-1.2.1-bin
hive> drop database bank_db cascade;
OK
Time taken: 1.001 seconds
hive> show database;
```