Practice File

**1) DDL Commands**

Create Schema - create schema employee\_data;

show schema- show schemas;

go to schems - use employee\_data;

Drop schema - Drop schema employee\_data;

Show tables - show tables;

**Create Table - Managed – Table**

create table employee\_uk

(

first\_name String,

last\_name String,

address String,

city String,

county String,

post String,

phone1 String,

phone2 String,

email String,

Country String

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n';

Drop table employee\_uk;

all alter table queries;

**3) Load data into table from a local file**

LOAD DATA LOCAL INPATH 'Employee\_data/uk-500/uk-500.csv' OVERWRITE INTO TABLE employee\_uk;

Create External table for US,AU,CA and all1500 employee

create external table employee\_us

(

first\_name String,

last\_name String,

address String,

city String,

state String,

post String,

phone1 String,

phone2 String,

email String,

Country String

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

LOCATION '/user/cloudera/Employee\_data/us-500';

create external table employee\_au

(

first\_name String,

last\_name String,

address String,

city String,

state String,

post String,

phone1 String,

phone2 String,

email String,

Country String

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

LOCATION '/user/cloudera/Employee\_data/au-500';

create external table employee\_ca

(

first\_name String,

last\_name String,

address String,

city String,

state String,

post String,

phone1 String,

phone2 String,

email String,

Country String

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

LOCATION '/user/cloudera/Employee\_data/ca-500';

create external table employee\_all

(

first\_name String,

last\_name String,

address String,

city String,

state String,

post String,

phone1 String,

phone2 String,

email String,

Country String

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n'

LOCATION '/user/cloudera/Employee\_data/AllEmp-1500';

Create Partition table

create table employee\_us\_static\_partition

(

first\_name String,

last\_name String,

address String,

city String,

post String,

phone1 String,

phone2 String,

email String

)

Partitioned by (Country String,State String)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n';

Static Partiton

insert overwrite table employee\_us\_static\_partition

partition (Country='US',State='AK')

select first\_name,last\_name,address,city,post,phone1,phone2,email

from employee\_all where country='US' and state='AK';

Create Partition table

create table employee\_all\_dynamic\_partition

(

first\_name String,

last\_name String,

address String,

city String,

post String,

phone1 String,

phone2 String,

email String

)

Partitioned by (Country String,State String)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n';

Dynamic Partition

set hive.exec.dynamic.partition.mode=nonstrict;

insert overwrite table employee\_all\_dynamic\_partition

partition (Country,State)

select first\_name,last\_name,address,city,post,phone1,phone2,email,country,state

from employee\_all ;

select \* from employee\_all\_dynamic\_partition where country='US' and state='WY';

Static and Dynamic Partition

create table employee\_us\_both\_partition

(

first\_name String,

last\_name String,

address String,

city String,

post String,

phone1 String,

phone2 String,

email String

)

Partitioned by (Country String,State String)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n';

insert overwrite table employee\_us\_both\_partition

partition (Country='US',State)

select first\_name,last\_name,address,city,post,phone1,phone2,email,state

from employee\_all where country='US';

Bucketing

create table employee\_all\_bucketed

(

first\_name String,

last\_name String,

address String,

city String,

state String,

post String,

phone1 String,

phone2 String,

email String

)

Partitioned by (Country String)

Clustered by (state) into 4 BUCKETS

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

LINES TERMINATED BY '\n';

set hive.enforce.bucketing = true;

set hive.exec.dynamic.partition.mode=nonstrict;

INSERT OVERWRITE TABLE employee\_all\_bucketed PARTITION (country)

select first\_name,last\_name,address,city,state,post,phone1,phone2,email,country

from employee\_all;

UDF

1) Write this java code in eclipse

package hiveUDF;

import org.apache.hadoop.hive.ql.exec.UDF;

import org.apache.hadoop.io.Text;

public class LowerToUpper extends UDF{

public Text evaluate(String input) {

if(input == null) return null;

return new Text(input.toUpperCase());

}

}

2) Create JAR file

3) add jar hiveUDF.jar

4) CREATE TEMPORARY FUNCTION MyLowerToUpper as 'hiveUDF.LowerToUpper';

5) select MyLowerToUpper(first\_name),MyLowerToUpper(first\_name) from employee\_ca limit 10;