**DE-HIVE ASSIGNMENT 1**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**--Create table and upload data from file to this table (one column)**

create external table dyn\_p\_temp(data1 STRING);

load data inpath '/user/maria\_dev/Assignment1/apat63\_99.txt'

overwrite into table dyn\_p\_temp;



\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

--**Create other table and extract data (partially) from previous table (with different columns)**

create external table dyn\_p\_temp1

(PatientID INT, GYEAR STRING, COUNTRY STRING)

tblproperties("skip.header.line.count"="1");

insert overwrite table dyn\_p\_temp1

SELECT

regexp\_extract(data1, '^(?:([^,]\*),?){1}', 1) PatientId,

regexp\_extract(data1, '^(?:([^,]\*),?){2}', 1) GYEAR,

regexp\_extract(data1, '^(?:([^,]\*),?){5}', 1) COUNTRY

from dyn\_p\_temp;



\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**-- Partitioning dynamic (in two parts).**

SET hive.exec.dynamic.partition=true;

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

CREATE TABLE apat\_dyn

(PatientID INT, COUNTRY STRING)

PARTITIONED BY (GYEAR STRING)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

ESCAPED BY "" LINES TERMINATED BY '\n'

STORED AS textfile

LOCATION "/user/maria\_dev/Assignment1";

SET hive.exec.dynamic.partition=true;

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

INSERT OVERWRITE TABLE apat\_dyn

PARTITION (GYEAR)

SELECT PatientId, COUNTRY, GYEAR

FROM dyn\_p\_temp1;



\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**-- Find out number of patents granted in year 1963.**

SELECT gyear as year, count(PatientID) as NumberPatient

from apat\_dyn

where gyear = '1963'

group by gyear;



\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**-- Find out number of patents granted in each country in year 1999.**

SELECT country, count(PatientID) as NumberPatient

from apat\_dyn

where gyear = '1999'

group by country;

