BDA LAB 9 - SAIFUR RAHMAN (1BM17CS086)

Q7. Write Queries in Hive to do the following.

1. Create an external table named with the following attributes → Empl_ID ->Emp_Name → Designation → Salary

CREATE DATABASE IF NOT EXISTS lab9 COMMENT 'employee program' WITH DBPROPERTIES

('creator'=SAIF);

SHOW DATABASES;

DESCRIBE DATABASE lab9;

USE lab9;

CREATE EXTERNAL TABLE IF NOT EXISTS Employee(EmpID INT,EmpName

STRING,Designation STRING,Salary FLOAT) ROW FORMAT DELIMITED FIELDS TERMINATED

BY '\t';

2. Load data into table from a given file

LOAD DATA LOCAL INPATH '/home/ayush/Desktop/employeeInput.txt' OVERWRITE INTO

TABLE Employee;

SELECT * FROM Employee;

3. Create a view to Generate a query to retrieve the employee details who earn

a salary of more than Rs 30000.

CREATE VIEW emp_30000 AS SELECT * FROM Employee WHERE Salary>30000;

SELECT * FROM emp_30000;

4. Alter the table to add a column Dept_Id and Generate a query to retrieve the

employee details in order by using Dept_Id

ALTER TABLE Employee ADD COLUMNS(DeptID INT);

LOAD DATA LOCAL INPATH '/home/saif/Desktop/employeeInputAltered.txt'

OVERWRITE INTO TABLE Employee;

SELECT * FROM Employee;

SELECT * FROM Employee ORDER BY DeptID;

5. Generate a query to retrieve the number of employees in each department whose

salary is greater than 30000

SELECT DeptID,count(*) FROM Employee WHERE Salary>=30000 GROUP BY DeptID;

6. Create another table Department with attributes → Dept_Id - >Dept_name

- **>Emp_Id**

CREATE EXTERNAL TABLE IF NOT EXISTS Department(DeptId INT,DeptName STRING) ROW

FORMAT DELIMITED FIELDS TERMINATED BY '\t';

LOAD DATA LOCAL INPATH '/home/saif/Desktop/DepartmentInput.txt' OVERWRITE

INTO TABLE Department;

SELECT * FROM Department;

7.Display the cumulative details of each employee along with department details

SELECT a.EmpID,a.EmpName,a.Designation,a.Salary,b.DeptName FROM Employee a

JOIN Department b ON a.DeptID=b.DeptId;

```
hive> SELECT * FROM Employee ORDER BY DeptID;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different exec
ution engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = hduser_20201219031430_f37e5e18-d55b-4d50-8836-b3d86a9ca242
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2020-12-19 03:14:36,260 Stage-1 map = 0%,  reduce = 0%
2020-12-19 03:14:45,727 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local426704584_0001
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 1944 HDFS Write: 972 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
2       Kelly   Sr. Manager     60000.0 11
1       John    Manager 50000.0 11
7       Rohan   HR manager      25000.0 12
8       Joana   Software Engineer       23000.0 13
4       Kedrik  Software Engineer       30000.0 13
3       Harry   Software Engineer       45000.0 13
6       Ubern   Test Engineer   20000.0 15
5       Tom     Test Engineer   20000.0 15
Time taken: 15.476 seconds, Fetched: 8 row(s)
```

```
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
2020-12-19 03:18:13,778 Stage-3 map = 100%,  reduce = 0%
Ended Job = job_local1327814845_0003
MapReduce Jobs Launched:
Stage-Stage-3:  HDFS Read: 1542 HDFS Write: 546 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1       John    Manager 50000.0  Business Management
2       Kelly   Sr. Manager     60000.0  Business Management
3       Harry   Software Engineer       45000.0 Development
4       Kedrik  Software Engineer       30000.0 Development
5       Tom     Test Engineer   20000.0 Testing
6       Ubern   Test Engineer   20000.0 Testing
7       Rohan   HR manager      25000.0 HR
8       Joana   Software Engineer       23000.0 Development
Time taken: 51.043 seconds, Fetched: 8 row(s)
```

```
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/usr/local/apache-hive-2.1.0-bin/lib/hive-common-2.1.0.jar!/hive-log4j2.
properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution eng
ine (i.e. spark, tez) or using Hive 1.X releases.
hive> CREATE DATABASE IF NOT EXISTS lab9 COMMENT 'employee program' WITH DBPROPERTIES ('creator'='CHANDANA');
OK
Time taken: 3.886 seconds
hive> SHOW DATABASES;
OK
default
lab9
Time taken: 1.768 seconds, Fetched: 2 row(s)
hive> DESCRIBE DATABASE lab9;
OK
lab9    hive employee program   hdfs://localhost:54310/user/hive/warehouse/lab9.db      hduser  USER
Time taken: 0.104 seconds, Fetched: 1 row(s)
hive> USE lab9;
OK
Time taken: 0.08 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS Employee(EmpID INT,EmpName STRING,Designation STRING,Salary FLOAT) ROW FORMAT DELI
MITED FIELDS TERMINATED BY '\t';
OK
Time taken: 1.98 seconds
hive> LOAD DATA LOCAL INPATH '/home/chandana/Desktop/employeeInput.txt' OVERWRITE INTO TABLE Employee;
Loading data to table lab9.employee
OK
Time taken: 5.339 seconds
hive>
```