

## Analysis the Supermarket Sales data using HQL

### BDMA – EXERCISE 6

**SURUTHI S**

**225229141**

```
.  
hive> create database if not exists db;  
OK  
Time taken: 0.444 seconds  
hive> show databases;  
OK  
db  
default  
Time taken: 0.227 seconds, Fetched: 2 row(s)  
hive> █
```

```
hive> describe database db;  
OK  
db          hdfs://quickstart.cloudera:8020/user/hive/warehouse  
db    cloudera    USER  
Time taken: 0.024 seconds, Fetched: 1 row(s)  
hive> █
```

```
hive> drop database db cascade;  
OK  
Time taken: 0.203 seconds  
hive> █
```

```
hive> create table emp(id INT,name STRING, dept STRING, yoj INT, sal INT)
      > ROW FORMAT DELIMITED
      > FIELDS TERMINATED BY ','
      > tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.918 seconds
```

```
hive> show tables;
OK
emp
Time taken: 0.031 seconds, Fetched: 1 row(s)
hive> █
```

```
hive> describe emp;
OK
id                int
name              string
dept             string
yoj              int
sal              int
Time taken: 0.209 seconds, Fetched: 5 row(s)
hive> █
```

```
[cloudera@quickstart ~]$ cat emp_data.csv
id,name,dept,yoj,sal
100,alex,DS,2002,2000000
101,tim,CS,2003,100000
103,martin,DS,2004,100020
104,sarah,CA,2010,102010
105,zenin,CS,2010,32021
105,carlin,CA,2013,242302

[cloudera@quickstart ~]$
```

```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/emp_data.csv' INTO TABLE
> emp;
Loading data to table default.emp
Table default.emp stats: [numFiles=1, totalSize=171]
OK
Time taken: 0.829 seconds
hive>
```

```
hive> select * from emp;
OK
100      alex      DS      2002      2000000
101      tim       CS      2003      100000
103      martin    DS      2004      100020
104      sarah     CA      2010      102010
105      zenin     CS      2010      32021
105      carlin    CA      2013      242302
NULL     NULL      NULL     NULL      NULL
Time taken: 0.26 seconds, Fetched: 7 row(s)
hive>
```

## QUERY 1

```
hive> SELECT AVG(sal) FROM emp GROUP BY DEPT;  
Query ID = cloudera_20230913095959_9f11cad5-5051-4bdf-adff-98d513b48610  
Total jobs = 1  
Launching Job 1 out of 1  
Number of reduce tasks not specified. Estimated from input data size: 1
```

```
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.68 s  
: 8528 HDFS Write: 30 SUCCESS  
Total MapReduce CPU Time Spent: 5 seconds 680 msec  
OK  
NULL  
172156.0  
66010.5  
1050010.0  
Time taken: 38.223 seconds, Fetched: 4 row(s)  
hive> █
```

## QUERY 2

```
hive>  
hive> SELECT dept, AVG(sal) FROM emp GROUP BY DEPT;  
Query ID = cloudera_20230913100101_8b39dd0c-ca6d-487e-bbe1  
Total jobs = 1  
Launching Job 1 out of 1
```

```
CA      172156.0  
CS      66010.5  
DS      1050010.0  
Time taken: 32.933 seconds, Fetched: 4 row(s)  
hive> █
```

## QUERY 3

```
hive> CREATE TABLE customer(id INT, name STRING, age INT, add STRING, salary FLOAT)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.449 seconds
hive> CREATE TABLE order(id INT, date DATE, cid INT, amt FLOAT)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ','
> tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.191 seconds
hive>
>
> show tables;
OK
customer
emp
order
Time taken: 0.197 seconds, Fetched: 3 row(s)
hive> █
```

```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/customer_data.csv'
> INTO TABLE customer;
Loading data to table default.customer
Table default.customer stats: [numFiles=1, totalSize=243]
OK
Time taken: 0.649 seconds
..
```

```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/order_data.csv'
> INTO TABLE order;
Loading data to table default.order
Table default.order stats: [numFiles=1, totalSize=178]
OK
Time taken: 0.673 seconds
hives>
```

```
>
> select * from customer;
OK
100      mike      23      london  20000.0
101      zenon     24      chennai 10000.0
103      sarah     25      bangalore      50000.0
104      franklin  39      sydney   34000.0
105      maddy     24      newyork 32000.0
106      jack      53      los angeles    21000.0
107      penny     24      delhi   43000.0
108      sheldon  37      texas    39000.0
Time taken: 0.227 seconds, Fetched: 8 row(s)
hive> █
```

```
hive>
> select * from order;
OK
200      2023-09-01      108      1000.0
201      2023-08-31      101      500.0
202      2023-08-06      100      2000.0
203      2023-09-02      107      250.0
204      2023-09-07      105      45.0
205      2023-08-15      106      321.0
206      2023-09-10      104      233.0
Time taken: 0.223 seconds, Fetched: 7 row(s)
hive> █
```

```
hive> SELECT c.id,c.name,c.add,o.amt  
> from customer c JOIN order o  
> on (c.id = o.cid);
```

Total MapReduce CPU Time Spent: 17 seconds 220 msec

OK

100	mike	london	2000.0	
101	zenon	chennai	500.0	
104	franklin		sydney	233.0
105	maddy	newyork	45.0	
106	jack	los angeles		321.0
107	penny	delhi	250.0	
108	sheldon	texas	1000.0	

Time taken: 44.303 seconds, Fetched: 7 row(s)

hive>