EmpData.csv

ID,Name,Department,JoinDate,Salary

1,Ian,Quality Assurance,2021,28113

2,Beatrice,Tech Support,2021,35330

3,Vladimir,Human Resources,2020,51445

4,Whitney,IT,2020,23818

5,Leslie,Customer Service,2021,59882

6,Bernard,IT,2021,50330

7,Mary,Customer Service,2021,26558

8,Jerome,RnD,2021,45333

9,Joshua,IT,2021,59538

10,Keane,Human Resources,2022,46500

11,Velma,Human Resources,2022,19816

12,Rogan,Tech Support,2022,27554

13,Aurelia,RnD,2021,20762

14,Merrill,Quality Assurance,2021,59660

15,Blaine,Tech Support,2022,28768

$ vim EmpData.csv

$ hive

hive> show databases;

hive> show databases;

OK

default

Time taken: 0.355 seconds, Fetched: 1 row(s)

hive> create database office;

hive> use office;

hive> CREATE TABLE employee

> (id INT, name STRING, dept STRING, yoj INT, salary INT)

> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> TBLPROPERTIES ("skip.header.line.count"="1");

hive> DESCRIBE employee;

hive> LOAD DATA LOCAL INPATH

> '/EmpData.csv'

> INTO TABLE employee;

hive> SELECT \* FROM employee;

hive> SELECT COUNT(\*) FROM employee;

# Column operations

hive> SELECT id, name FROM employee;

# Row operation

hive> SELECT \* FROM employee WHERE salary > 30000;

# Column operation that need aggregation

hive> SELECT count(\*) FROM employee;

#Exports to HDFS directory

hive> INSERT OVERWRITE DIRECTORY '/user/root/output/export'

> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> SELECT \* FROM employee;

hive> dfs -ls /user/root/output/export

#Exports to LOCAL directory

hive> INSERT OVERWRITE LOCAL DIRECTORY '/export'

> ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> SELECT \* FROM emp.employee;

$ cat output.csv