ACADGILD

Session 08: Advanced Hive

Assignment 1

1. Problem Statement

- Get a list of employees who receive a salary less than 100, compared to their immediate employee with higher salary in the same unit

```
hive> create table emp_data(id int,name string,salary int,dept string) row forma
t delimited fields terminated by '\t' stored as textfile;
OK
Time taken: 0.284 seconds
```

```
hive> create view emp_view as select name,salary,lag(salary,1,0) over (partition by dept order by salary desc) from emp_data;

OK
Time taken: 0.125 seconds
hive> select * from emp_view;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
tion engine (i.e. spark, tez) or using Hive 1.X releases.
Query ID = acadgild_20170505144236_2adf136d-cbbf-460a-bf36-e3e4f343e87a
```

```
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.0 sec   HDFS Read: 10092 HDFS Write: 240 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 0 msec
OK
Yadav   300     0
Sumit   200     300
Amit    100     200
Sunil   500     0
Mahoor  200     500
Kranti  100     200
Time taken: 20.504 seconds, Fetched: 6 row(s)
```

```
hive> describe emp_view;
OK
name                    string
salary                  int
lag_window_0            int
Time taken: 0.09 seconds, Fetched: 3 row(s)
hive> select * from emp_view where lag_window_0-salary>100;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
Ended Job = job_1493957425975_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.05 sec   HDFS Read: 10335 HDFS Write: 114 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 50 msec
OK
Mahoor  200     500
Time taken: 22.604 seconds, Fetched: 1 row(s)
```

- List of all employees who draw higher salary than the average salary of that department

```
hive> create view avg as select name,salary,avg(salary) over (partition by dept) from emp_data;
OK
Time taken: 0.198 seconds
hive> describe avg;
OK
name                    string
salary                  int
avg_window_0            double
Time taken: 0.072 seconds, Fetched: 3 row(s)
hive> select * from avg where salary>avg_window_0;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execu
```

```
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.57 sec   HDFS Read: 9990 HDFS Write: 155 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 570 msec
OK
Yadav   300     200.0
Sunil   500     266.6666666666667
Time taken: 20.493 seconds, Fetched: 2 row(s)
```