DSBDA MOCK - Hive Inner Join on Employees and Department
Roll no - 33322

[cloudera@quickstart Desktop]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create table employeed(emp_id int, emp_name string, dept_id int) row format delimited fields terminated by ','
      > ;
OK
Time taken: 5.686 seconds
hive> create table employees(emp_id int, emp_name string, dept_id int) row format delimited fields terminated by ',' stored as TEXTFILE;
OK
Time taken: 0.323 seconds
hive> create Display all 461 possibilities? (y or n)^C[cloudera@quickstart Desktop]$ ^C
[cloudera@quickstart Desktop]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create table departments(dept_name string, dept_id int) row format delimited fields terminated by ',';
OK
Time taken: 1.991 seconds
hive> create table departments(dept_name string, dept_id int) row format delimited fields terminated by ',' stored as TEXTFILE;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask.
AlreadyExistsException(message:Table departments already exists)
hive> DROP TABLE departments;
OK
Time taken: 1.259 seconds
hive>
      > create table departments(dept_name string, dept_id int) row format delimited fields terminated by ',' stored as TEXTFILE;
OK
Time taken: 0.144 seconds
hive> load data inpath 'HIVEJOIN/employees.txt' into table employees;
Loading data to table default.employees
chgrp: changing ownership of
'hdfs://quickstart.cloudera:8020/user/hive/warehouse/employees/employees.txt': User does not belong to hive
Table default.employees stats: [numFiles=1, totalSize=43]
OK
Time taken: 1.483 seconds
hive> load data inpath 'HIVEJOIN/departments.txt' into table departments;
Loading data to table default.departments

chgrp: changing ownership of
'hdfs://quickstart.cloudera:8020/user/hive/warehouse/departments/departments.txt': User
does not belong to hive
Table default.departments stats: [numFiles=1, totalSize=15]
OK
Time taken: 0.555 seconds
hive> hive> load data inpath 'HIVEJOIN/departments.txt' into table departments;
NoViableAltException(26@[])
        at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1020)
        at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:199)
        at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:166)
        at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:393)
        at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:307)
        at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1110)
        at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1158)
        at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1047)
        at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1037)
        at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:207)
        at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:159)
        at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:370)
        at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:756)
        at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:675)
        at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:615)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
        at
sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:606)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:0 cannot recognize input near 'hive' '>' 'load'
hive> Loading data to table default.departments
        > chgrp: changing ownership of
'hdfs://quickstart.cloudera:8020/user/hive/warehouse/departments/departments.txt': User
does not belong to hive
        > Table default.departments stats: [numFiles=1, totalSize=15]
        > OK
        > Time taken: 0.555 seconds
        > hive>
        > [cloudera@quickstart Desktop]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> select * from employees;
OK
1   john   2
2   jacob   2
3   tim   1

```
4       tom     3
5       liz     1
Time taken: 3.118 seconds, Fetched: 5 row(s)
hive> select * from departments
        > ;
OK
hr      1
it      2
qs      3
Time taken: 0.151 seconds, Fetched: 3 row(s)
hive> select employees.emp_id, employees.emp_name, departments.dept_name from
employees join departments on employees.dept_id = departments.dept_id;
Query ID = cloudera_20250325232727_32129e88-1c69-48cc-b40d-97716630d68f
Total jobs = 1
Execution log at:
/tmp/cloudera/cloudera_20250325232727_32129e88-1c69-48cc-b40d-97716630d68f.log
2025-03-25 11:27:48  Starting to launch local task to process map join;     maximum memory
= 1013645312
2025-03-25 11:27:52  Dump the side-table for tag: 1 with group count: 3 into file:
file:/tmp/cloudera/2df66564-4086-4dbd-8395-5711558cbed5/hive_2025-03-25_23-27-33_91
8_9042112746110761413-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
2025-03-25 11:27:52  Uploaded 1 File to:
file:/tmp/cloudera/2df66564-4086-4dbd-8395-5711558cbed5/hive_2025-03-25_23-27-33_91
8_9042112746110761413-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
(326 bytes)
2025-03-25 11:27:52  End of local task; Time Taken: 4.325 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1742969079981_0001, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1742969079981_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1742969079981_0001
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-03-25 23:28:23,562 Stage-3 map = 0%,  reduce = 0%
2025-03-25 23:28:42,382 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.45 sec
MapReduce Total cumulative CPU time: 1 seconds 450 msec
Ended Job = job_1742969079981_0001
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.45 sec   HDFS Read: 6098 HDFS Write: 48
SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 450 msec
OK
1       john    it
2       jacob   it
3       tim     hr
4       tom     qs
5       liz     hr
```

Time taken: 69.673 seconds, Fetched: 5 row(s)
hive>


```
[cloudera@quickstart Desktop]$ hdfs dfs -put employees.txt HIVE
[cloudera@quickstart Desktop]$ hdfs dfs -put departments.txt HIVE
put: `HIVE': File exists
[cloudera@quickstart Desktop]$ hdfs dfs -put employees.txt HIVE/
put: `HIVE': File exists
[cloudera@quickstart Desktop]$ hdfs dfs -put employees.txt HIVEJOIN/
put: `HIVEJOIN/': No such file or directory
[cloudera@quickstart Desktop]$ hdfs dfs -mkdir HIVEJOIN
[cloudera@quickstart Desktop]$ hdfs dfs -put employees.txt HIVEJOIN/
[cloudera@quickstart Desktop]$ hdfs dfs -put departments.txt HIVEJOIN/
[cloudera@quickstart Desktop]$
```


Employees.txt
1,john,2
2,jacob,2
3,tim,1
4,tom,3
5,liz,1

Departments.txt
hr,1
it,2
qs,3