```
[cloudera@quickstart ~]$ hdfs dfs -mkdir -p /user/hive/data

[cloudera@quickstart ~]$ hdfs dfs -put employees.csv /user/hive/data/

[cloudera@quickstart ~]$ hdfs dfs -put departments.csv /user/hive/data/

put: `departments.csv': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/data/
Found 1 items
-rw-r--r--   1 cloudera hive         82 2025-09-22 23:06
/user/hive/data/employees.csv
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
[cloudera@quickstart ~]$ hdfs dfs -put departments.csv /user/hive/data/
[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/data/
Found 2 items
-rw-r--r--   1 cloudera hive         24 2025-09-22 23:07
/user/hive/data/departments.csv
-rw-r--r--   1 cloudera hive         82 2025-09-22 23:06
/user/hive/data/employees.csv
[cloudera@quickstart ~]$ 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> LOAD DATA INPATH '/user/hive/data/employees.csv' INTO TABLE
employees;
Loading data to table default.employees
Table default.employees stats: [numFiles=1, totalSize=82]
OK
Time taken: 2.754 seconds
hive> LOAD DATA INPATH '/user/hive/data/departments.csv' INTO TABLE
departments;
Loading data to table default.departments
Table default.departments stats: [numFiles=1, totalSize=24]
OK
Time taken: 0.56 seconds
hive> SELECT * FROM employees;
OK
1      Alice 10     60000
2      Bob   20     50000
3      Carol 10     70000
```

```
4       David 30      40000
5       Eva   20      55000
NULL    NULL  NULL    NULL
Time taken: 1.443 seconds, Fetched: 6 row(s)
hive> SELECT * FROM departments;
OK
10      HR
20      IT
30      Finance
NULL    NULL
Time taken: 0.091 seconds, Fetched: 4 row(s)
hive> SELECT name, salary
    > FROM employees;
OK
Alice 60000
Bob   50000
Carol 70000
David 40000
Eva   55000
NULL  NULL
Time taken: 0.204 seconds, Fetched: 6 row(s)
hive> SELECT *
    > FROM employees
    > WHERE salary > 55000;
OK
1       Alice 10      60000
3       Carol 10      70000
Time taken: 0.437 seconds, Fetched: 2 row(s)
hive> SELECT e.emp_id, e.name, e.salary, d.dept_name
    > FROM employees e
    > JOIN departments d
    > ON e.dept_id = d.dept_id;
Query ID = cloudera_20250922230909_3953ade6-a08e-4571-8c79-4deaa30cac68
Total jobs = 1
Execution log at:
/tmp/cloudera/cloudera_20250922230909_3953ade6-a08e-4571-8c79-4deaa30cac68.
log
2025-09-22 11:09:36     Starting to launch local task to process map join;
        maximum memory = 932184064
2025-09-22 11:09:46     Dump the side-table for tag: 1 with group count: 3
into file:
file:/tmp/cloudera/050338da-91ae-45ac-b485-afd778d623d6/hive_2025-09-22_23-
09-21_798_3529207987371396255-1/-local-10003/HashTable-Stage-3/MapJoin-mapf
```

ile01--.hashtable
2025-09-22 11:09:46   Uploaded 1 File to:
file:/tmp/cloudera/050338da-91ae-45ac-b485-afd778d623d6/hive_2025-09-22_23-09-21_798_3529207987371396255-1/-local-10003/HashTable-Stage-3/MapJoin-mapf
ile01--.hashtable (331 bytes)
2025-09-22 11:09:46   End of local task; Time Taken: 9.932 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_1758605977300_0001, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1758605977300_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1758605977300_0001
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-09-22 23:10:30,361 Stage-3 map = 0%, reduce = 0%
2025-09-22 23:11:07,393 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 6.39 sec
MapReduce Total cumulative CPU time: 6 seconds 390 msec
Ended Job = job_1758605977300_0001
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 6.39 sec   HDFS Read: 6291 HDFS Write: 86 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 390 msec
OK
1   Alice 60000 HR
2   Bob   50000 IT
3   Carol 70000 HR
4   David 40000 Finance
5   Eva   55000 IT
Time taken: 108.938 seconds, Fetched: 5 row(s)

hive> SELECT dept_id, AVG(salary) AS avg_salary
    > FROM employees

```
    > GROUP BY dept_id;
Query ID = cloudera_20250922231616_732b3a84-e12c-4745-8a36-99a93bacdf4b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 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>
Starting Job = job_1758605977300_0002, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1758605977300_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758605977300_0002
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2025-09-22 23:16:50,823 Stage-1 map = 0%,  reduce = 0%
2025-09-22 23:17:45,397 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
6.94 sec
2025-09-22 23:18:07,418 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU
11.48 sec
MapReduce Total cumulative CPU time: 11 seconds 480 msec
Ended Job = job_1758605977300_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 11.48 sec   HDFS Read:
7374 HDFS Write: 39 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 480 msec
OK
NULL   NULL
10     65000.0
20     52500.0
30     40000.0
Time taken: 108.6 seconds, Fetched: 4 row(s)
hive> SELECT *
    > FROM employees
    > ORDER BY salary DESC;
Query ID = cloudera_20250922231818_a3db25b8-e34c-4102-8c9c-8455be94d0e2
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>
Starting Job = job_1758605977300_0003, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1758605977300_0003/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758605977300_0003
Hadoop job information for Stage-1: number of mappers: 1; number of
reducers: 1
2025-09-22 23:18:34,083 Stage-1 map = 0%,  reduce = 0%
2025-09-22 23:18:55,834 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU
4.91 sec
2025-09-22 23:19:17,392 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU
8.74 sec
MapReduce Total cumulative CPU time: 8 seconds 740 msec
Ended Job = job_1758605977300_0003
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 8.74 sec   HDFS Read:
6337 HDFS Write: 93 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 740 msec
OK
3      Carol 10     70000
1      Alice 10     60000
5      Eva   20     55000
2      Bob   20     50000
4      David 30     40000
NULL   NULL  NULL   NULL
Time taken: 63.074 seconds, Fetched: 6 row(s)
hive>
```