Hive:

```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in jar:file:/usr/jars/hive-common-1.1.0-
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show databases;
OK
default
demo1
emp
Time taken: 0.638 seconds, Fetched: 3 row(s)
```

```
hive> drop database emp;
OK
Time taken: 0.112 seconds
hive> create database emp;
OK
Time taken: 0.077 seconds
hive> create table emp.employees(EmpId int,Age int,sales_percent float,No_of_pro
ject int,increment float)
    > row format delimited
    > fields terminated by ',' ;
OK
Time taken: 0.273 seconds
hive> describe emp.employees;
OK
empid                   int
age                     int
sales_percent           float
no_of_project           int
increment               float
Time taken: 0.112 seconds, Fetched: 5 row(s)
```

```
hive> create table if not exists emp.copy_table like emp.employees;
OK
Time taken: 0.125 seconds
```

```
hive> load data local inpath '/home/cloudera/Downloads/sales_data_sample.csv' in
to table emp.employees;
Loading data to table emp.employees
Table emp.employees stats: [numFiles=1, totalSize=371]
OK
Time taken: 0.57 seconds
hive> select * from emp.employees;
OK
10107    30      95.7     2       2871.0
10121    34      81.35    5       2765.9
10134    41      94.74    2       3884.34
10145    45      83.26    6       3746.7
10159    49      100.0    14      5205.27
10168    36      96.66    1       3479.76
10180    29      86.13    9       2497.77
10188    48      100.0    1       5512.32
10201    22      98.57    2       2168.54
10211    41      100.0    14      4708.44
10223    37      100.0    1       3965.66
10237    23      100.0    7       2333.12
10251    28      100.0    2       3188.64
10263    34      100.0    2       3676.76


hive> select * from emp.employees2;
OK
10107    30       95.7     2       2871.0
10121    34       81.35    5       2765.9
10134    41       94.74    2       3884.34
10145    45       83.26    6       3746.7
10159    49       100.0    14      5205.27
10168    36       96.66    1       3479.76
10180    29       86.13    9       2497.77
10188    48       100.0    1       5512.32
10201    22       98.57    2       2168.54
10211    41       100.0    14      4708.44
10223    37       100.0    1       3965.66
10237    23       100.0    7       2333.12
10251    28       100.0    2       3188.64
10263    34       100.0    2       3676.76
10275    45       92.83    1       4177.35
Time taken: 0.088 seconds, Fetched: 15 row(s)
hive> Alter table emp.employees rename to emp.employees1;
OK
Time taken: 0.116 seconds
```

```
hive> use emp;
OK
Time taken: 0.011 seconds
hive> show tables;
OK
copy_table
employees1
employees2
Time taken: 0.011 seconds, Fetched: 3 row(s)
```

```
hive> select employees1.empid,employees1.empid from employees1,employees2;
Warning: Map Join MAPJOIN[7][bigTable=employees2] in task 'Stage-3:MAPRED' is a cross product
Query ID = cloudera_20221108021515_c2be0502-19ff-4712-b602-cb37da41bc32
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20221108021515_c2be0502-19ff-4712-b602-cb37da41bc32.log
2022-11-08 02:15:47     Starting to launch local task to process map join;     maximum memory = 49807360
2022-11-08 02:15:48     Dump the side-table for tag: 0 with group count: 1 into file: file:/tmp/cloudera/35025124-0251-
42d3-9bce-da8388c5b69a/hive_2022-11-08_02-15-43_644_3264878058206245337-1/-local-10003/HashTable-Stage-3/MapJoin-mapfil
e00--.hashtable
2022-11-08 02:15:48     Uploaded 1 File to: file:/tmp/cloudera/35025124-0251-42d3-9bce-da8388c5b69a/hive_2022-11-08_02-
15-43_644_3264878058206245337-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashtable (392 bytes)
2022-11-08 02:15:48     End of local task; Time Taken: 1.124 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_1667898921734_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667898921734_0
001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1667898921734_0001
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-11-08 02:15:58,541 Stage-3 map = 0%,  reduce = 0%
2022-11-08 02:16:04,856 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.82 sec
MapReduce Total cumulative CPU time: 1 seconds 820 msec
```

```
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.82 sec   HDFS Read: 6434 HDFS Write: 2700 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 820 msec
OK
10107    10107
10121    10121
10134    10134
10145    10145
10159    10159
10168    10168
10180    10180
10188    10188
10201    10201
10211    10211
10223    10223
10237    10237
10251    10251
10263    10263
10275    10275
10107    10107
10121    10121
10134    10134
10145    10145
```

```
10251     10251
10263     10263
10275     10275
10107     10107
10121     10121
10134     10134
10145     10145
10159     10159
10168     10168
10180     10180
10188     10188
10201     10201
10211     10211
10223     10223
10237     10237
10251     10251
10263     10263
10275     10275
Time taken: 22.333 seconds, Fetched: 225 row(s)
```

Apache pig :


```
Job DAG:
job_local1501966295_0002


2022-11-14 23:27:49,502 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2022-11-14 23:27:49,503 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-14 23:27:49,503 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-14 23:27:49,504 [main] WARN  org.apache.pig.data.SchemaTupleBackend - Sc
hemaTupleBackend has already been initialized
2022-11-14 23:27:49,551 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2022-11-14 23:27:49,551 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(1,utkarsh,gupta,,pune)
(2,parth,kabadi,,mumbai)
(3,harsh,pande,,amravati)
(4,aryan,bhusari,,pune)
grunt> describe MyCSVData
MyCSVData: {id: int,firstname: chararray,lastname: chararray,phone: int,city: ch
ararray}
grunt>
```



```
Output(s):
Successfully stored records in: "file:/tmp/temp-1306075918/tmp-1499326309"

Job DAG:
job_local1501966295_0002


2022-11-14 23:27:49,502 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2022-11-14 23:27:49,503 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-14 23:27:49,503 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-14 23:27:49,504 [main] WARN  org.apache.pig.data.SchemaTupleBackend - Sc
hemaTupleBackend has already been initialized
2022-11-14 23:27:49,551 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2022-11-14 23:27:49,551 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(1,utkarsh,gupta,,pune)
(2,parth,kabadi,,mumbai)
(3,harsh,pande,,amravati)
(4,aryan,bhusari,,pune)
grunt>
```

```
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-14 23:17:12,973 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-14 23:17:12,981 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-14 23:17:13,059 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-14 23:17:13,061 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-14 23:17:13,183 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-14 23:17:13,185 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-14 23:17:13,281 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-14 23:17:13,282 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
grunt> MyCSVData = LOAD '/home/cloudera/Downloads/sample_1.csv' using PigStorage
(',') AS (id:int,firstname:chararray,lastname:chararray,phone:int,city:chararray
);
grunt>
```



```
1,utkarsh,gupta,9860762595,pune
2,parth,kabadi,8778903465,mumbai
3,harsh,pande,7889057481,amravati
4,aryan,bhusari,9809813345,pune
```

cloudera@quickstart:~   _ □ ✕

File   Edit   View   Search   Terminal   Help

```
oncatenation threshold: 100 optimistic? false
2022-11-14 23:31:14,583 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
 - MR plan size before optimization: 1
2022-11-14 23:31:14,583 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
 - MR plan size after optimization: 1
2022-11-14 23:31:14,584 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the
 job
2022-11-14 23:31:14,584 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler
- mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-11-14 23:31:14,596 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been ini
tialized
2022-11-14 23:31:14,599 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Al
iases being processed per job phase (AliasName[line,offset]): M: MyCSVData[2,12] C:  R:
2022-11-14 23:31:14,599 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File c
oncatenation threshold: 100 optimistic? false
2022-11-14 23:31:14,600 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
 - MR plan size before optimization: 1
2022-11-14 23:31:14,600 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
 - MR plan size after optimization: 1
2022-11-14 23:31:14,601 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the
 job
2022-11-14 23:31:14,601 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler
- mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-11-14 23:31:14,663 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been ini
tialized
2022-11-14 23:31:14,666 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Al
iases being processed per job phase (AliasName[line,offset]): M: MyCSVData[2,12] C:  R:
2022-11-14 23:31:14,666 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File c
oncatenation threshold: 100 optimistic? false
2022-11-14 23:31:14,667 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
 - MR plan size before optimization: 1
2022-11-14 23:31:14,667 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer
 - MR plan size after optimization: 1
2022-11-14 23:31:14,667 [main] INFO  org.apache.pig.tools.pigstats.ScriptState - Pig script settings are added to the
 job
2022-11-14 23:31:14,667 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler
- mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-11-14 23:31:14,683 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been ini
tialized
2022-11-14 23:31:14,685 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Al
iases being processed per job phase (AliasName[line,offset]): M: MyCSVData[2,12] C:  R:
-------------------------------------------------------------------------------------------------
| MyCSVData    | id:int   | firstname:chararray  | lastname:chararray  | phone:int   | city:chararray  |
-------------------------------------------------------------------------------------------------
|              | 4        | aryan                | bhusari             | 9809813345  | pune            |
-------------------------------------------------------------------------------------------------

grunt>
```

```
    |   |
    |   |---Project[bytearray][4] - scope-51
    |
    |---MyCSVData: Load(/home/cloudera/Downloads/sample_1-2.csv:PigStorage(','))
 - scope-38

2022-11-14 23:29:36,125 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? fal
se
2022-11-14 23:29:36,128 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-11-14 23:29:36,128 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
#--------------------------------------------------
# Map Reduce Plan
#--------------------------------------------------
MapReduce node scope-56
Map Plan
MyCSVData: Store(fakefile:org.apache.pig.builtin.PigStorage) - scope-55
|
|---MyCSVData: New For Each(false,false,false,false,false)[bag] - scope-54
    |   |
    |   Cast[int] - scope-40
    |   |
    |   |---Project[bytearray][0] - scope-39
    |   |
    |   Cast[chararray] - scope-43
    |   |
    |   |---Project[bytearray][1] - scope-42
    |   |
    |   Cast[chararray] - scope-46
    |   |
    |   |---Project[bytearray][2] - scope-45
    |   |
    |   Cast[int] - scope-49
    |   |
    |   |---Project[bytearray][3] - scope-48
    |   |
    |   Cast[chararray] - scope-52
    |   |
    |   |---Project[bytearray][4] - scope-51
    |
    |---MyCSVData: Load(/home/cloudera/Downloads/sample_1-2.csv:PigStorage(','))
 - scope-38--------
Global sort: false
----------------

grunt> █
```

```
cloudera@quickstart:~                               _  □  ✕

File  Edit  View  Search  Terminal  Help

Output(s):
Successfully stored records in: "file:/tmp/temp1223820747/tmp-1156799919"

Job DAG:
job_local1016883113_0003


2022-11-14 23:35:17,997 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.mapReduceLayer.MapReduceLauncher - Success!
2022-11-14 23:35:17,998 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - fs.default.name is deprecated. Instead, use fs.defaultFS
2022-11-14 23:35:17,998 [main] INFO  org.apache.hadoop.conf.Configuration.deprec
ation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-11-14 23:35:17,998 [main] WARN  org.apache.pig.data.SchemaTupleBackend - Sc
hemaTupleBackend has already been initialized
2022-11-14 23:35:18,004 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileI
nputFormat - Total input paths to process : 1
2022-11-14 23:35:18,004 [main] INFO  org.apache.pig.backend.hadoop.executionengi
ne.util.MapRedUtil - Total input paths to process : 1
(pune,{(4,aryan,bhusari,,pune),(1,utkarsh,gupta,,pune)})
(mumbai,{(2,parth,kabadi,,mumbai)})
(amravati,{(3,harsh,pande,,amravati)})
grunt> █
```