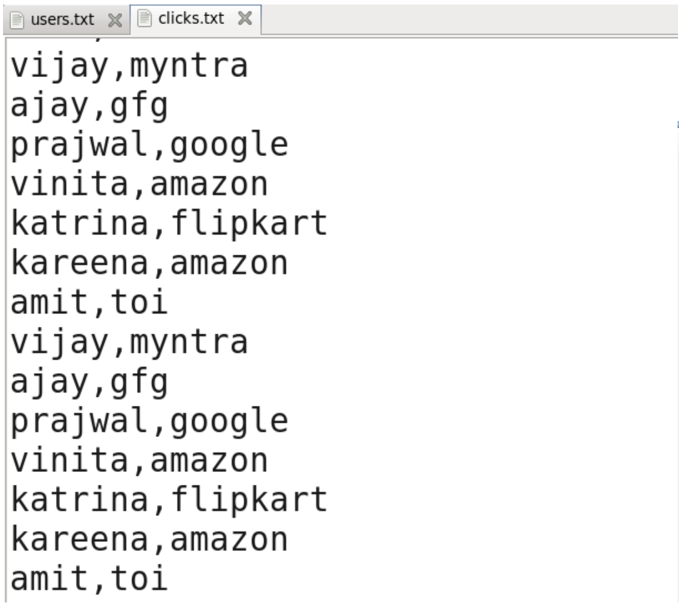
MUMBAI EDUCATIONAL TRUST

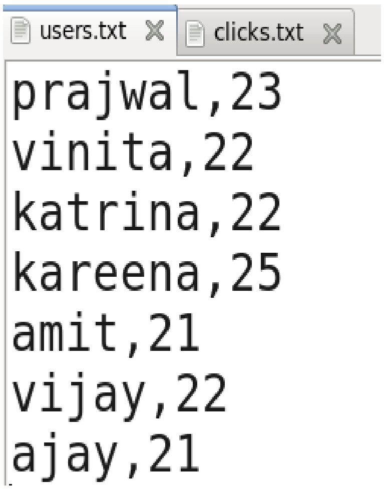
MET INSTITUTE OF COMPUTER SCIENCE

Program Number	
Roll Number	1546
Title of program	Pig 2, hive joins
Program	

create .txt files

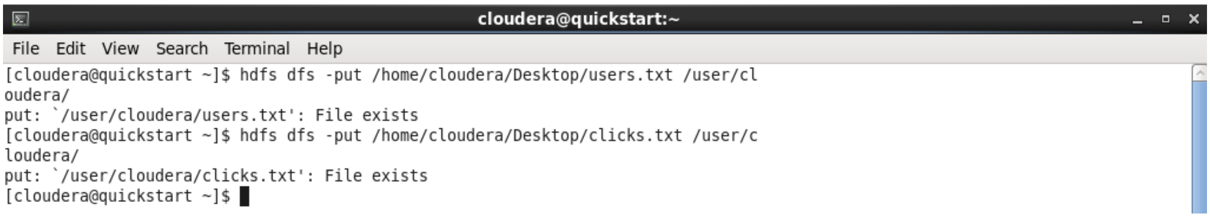


```
vijay,myntra
ajay,gfg
prajwal,google
vinita,amazon
katrina,flipkart
kareena,amazon
amit,toi
vijay,myntra
ajay,gfg
prajwal,google
vinita,amazon
katrina,flipkart
kareena,amazon
amit,toi
```



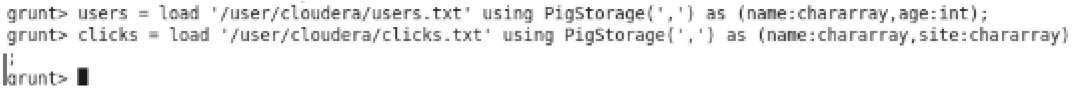
```
prajwal,23
vinita,22
katrina,22
kareena,25
amit,21
vijay,22
ajay,21
```

put files to cloud

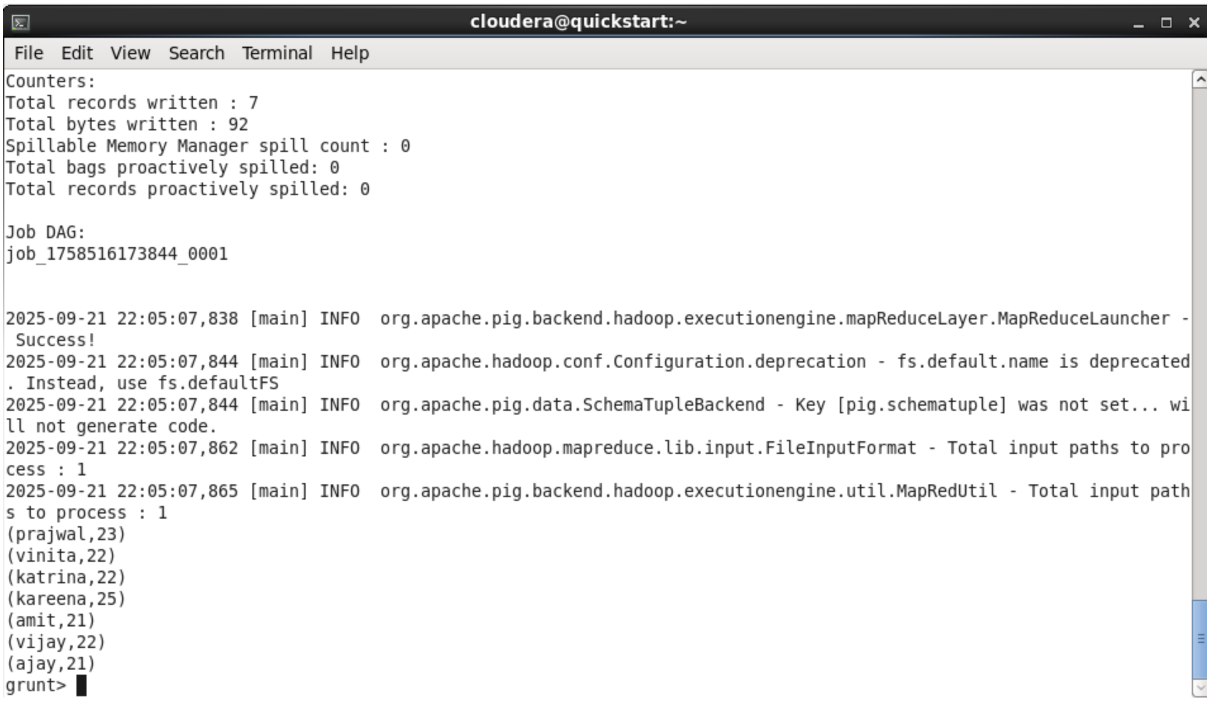


```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/users.txt /user/cl
oudera/
put: `/user/cloudera/users.txt': File exists
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/clicks.txt /user/c
loudera/
put: `/user/cloudera/clicks.txt': File exists
[cloudera@quickstart ~]$ 
```

create tables and load data

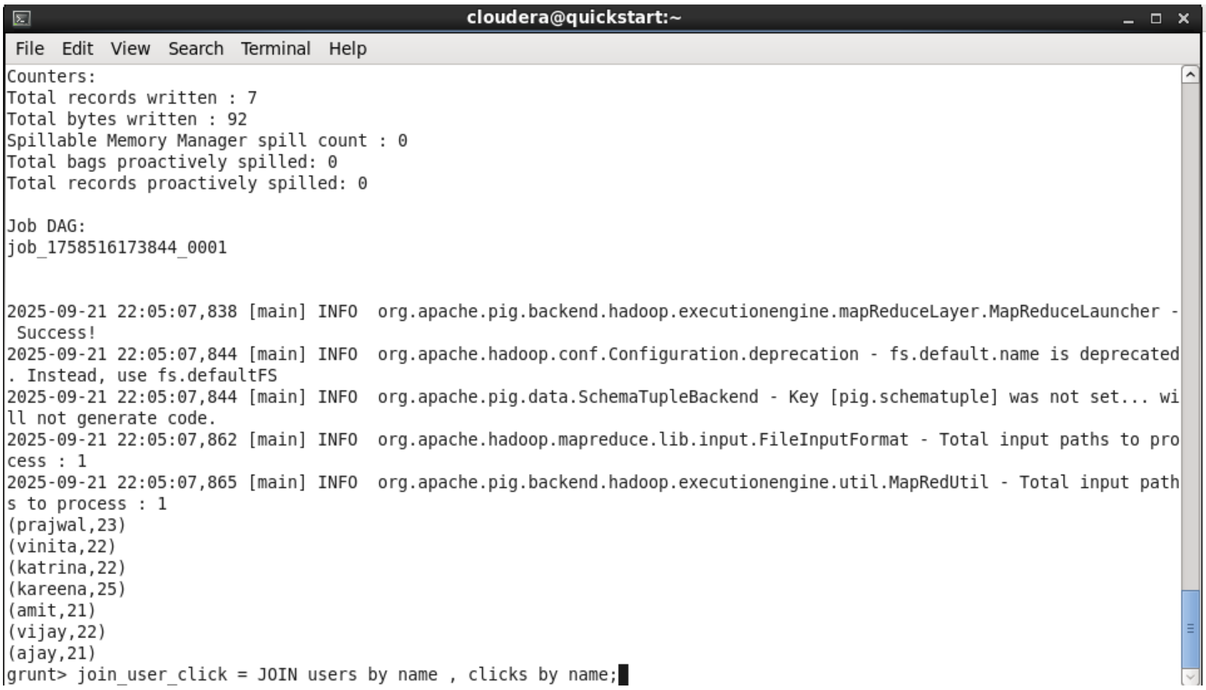


```
grunt> users = load '/user/cloudera/users.txt' using PigStorage(',') as (name:chararray,age:int);
grunt> clicks = load '/user/cloudera/clicks.txt' using PigStorage(',') as (name:chararray,site:chararray)
;
grunt> 
```

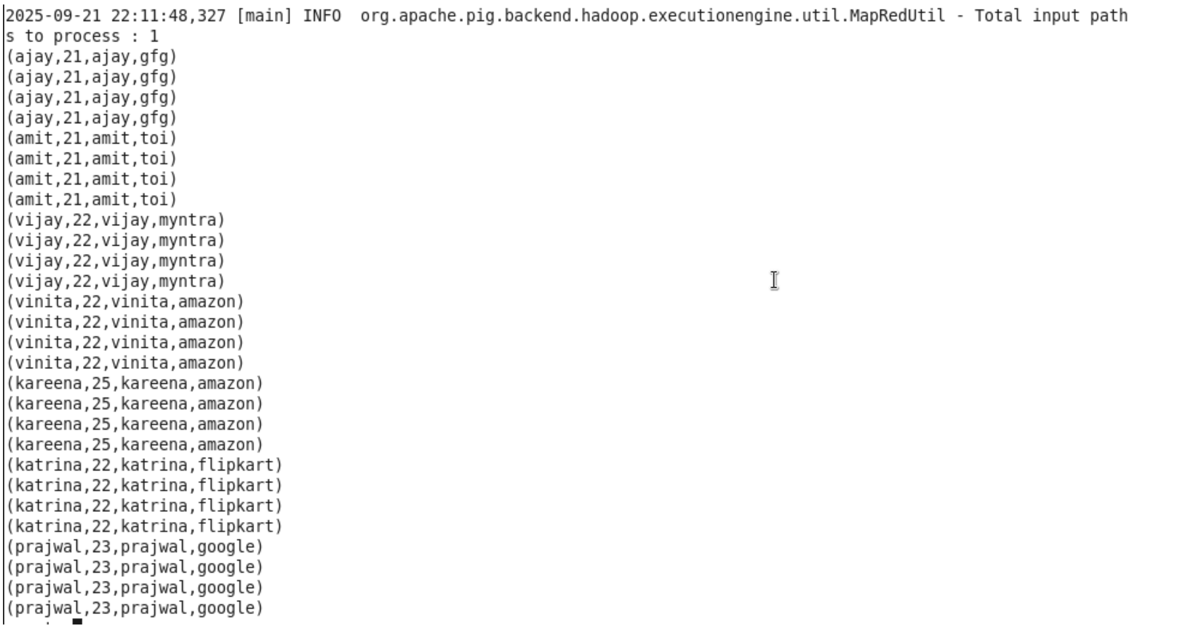


```
Counters:
Total records written : 7
Total bytes written : 92
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1758516173844_0001


2025-09-21 22:05:07,838 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
 Success!
2025-09-21 22:05:07,844 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated
. Instead, use fs.defaultFS
2025-09-21 22:05:07,844 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... wi
ll not generate code.
2025-09-21 22:05:07,862 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to pro
cess : 1
2025-09-21 22:05:07,865 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input path
s to process : 1
(prajwal,23)
(vinita,22)
(katrina,22)
(kareena,25)
(amit,21)
(vijay,22)
(ajay,21)
grunt> 
```

natural join

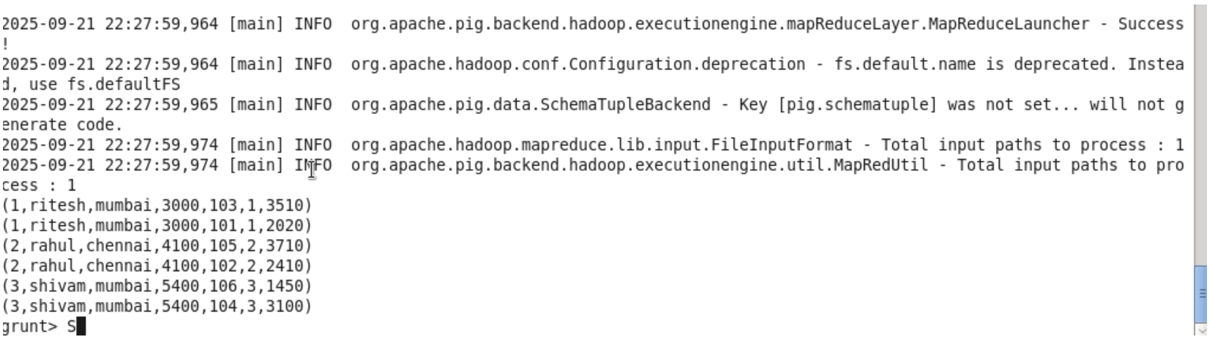


```
Counters:
Total records written : 7
Total bytes written : 92
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_1758516173844_0001


2025-09-21 22:05:07,838 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher -
 Success!
2025-09-21 22:05:07,844 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated
. Instead, use fs.defaultFS
2025-09-21 22:05:07,844 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... wi
ll not generate code.
2025-09-21 22:05:07,862 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to pro
cess : 1
2025-09-21 22:05:07,865 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input path
s to process : 1
(prajwal,23)
(vinita,22)
(katrina,22)
(kareena,25)
(amit,21)
(vijay,22)
(ajay,21)
grunt> join_user_click = JOIN users by name , clicks by name;
```

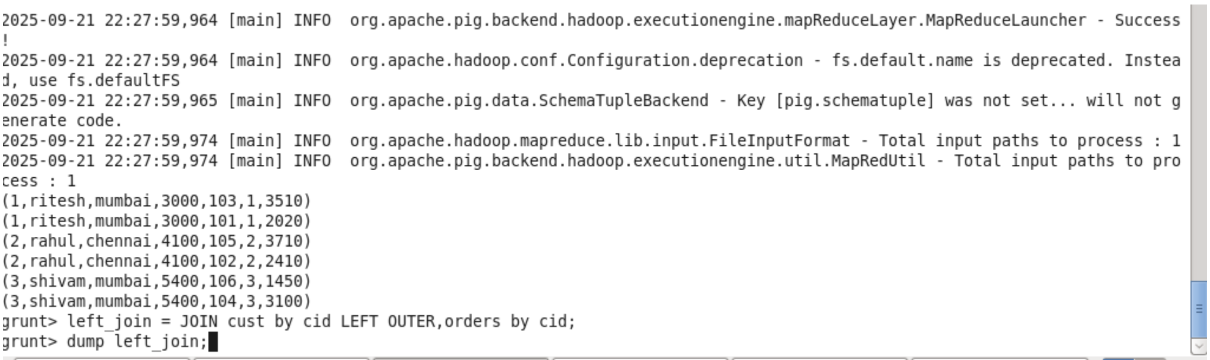


```
2025-09-21 22:11:48,327 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input path
s to process : 1
(ajay,21,ajay,gfg)
(ajay,21,ajay,gfg)
(ajay,21,ajay,gfg)
(ajay,21,ajay,gfg)
(amit,21,amit,toi)
(amit,21,amit,toi)
(amit,21,amit,toi)
(amit,21,amit,toi)
(vijay,22,vijay,myntra)
(vijay,22,vijay,myntra)
(vijay,22,vijay,myntra)
(vijay,22,vijay,myntra)
(vinita,22,vinita,amazon)
(vinita,22,vinita,amazon)
(vinita,22,vinita,amazon)
(vinita,22,vinita,amazon)
(kareena,25,kareena,amazon)
(kareena,25,kareena,amazon)
(kareena,25,kareena,amazon)
(kareena,25,kareena,amazon)
(katrina,22,katrina,flipkart)
(katrina,22,katrina,flipkart)
(katrina,22,katrina,flipkart)
(katrina,22,katrina,flipkart)
(prajwal,23,prajwal,google)
(prajwal,23,prajwal,google)
(prajwal,23,prajwal,google)
(prajwal,23,prajwal,google)
```

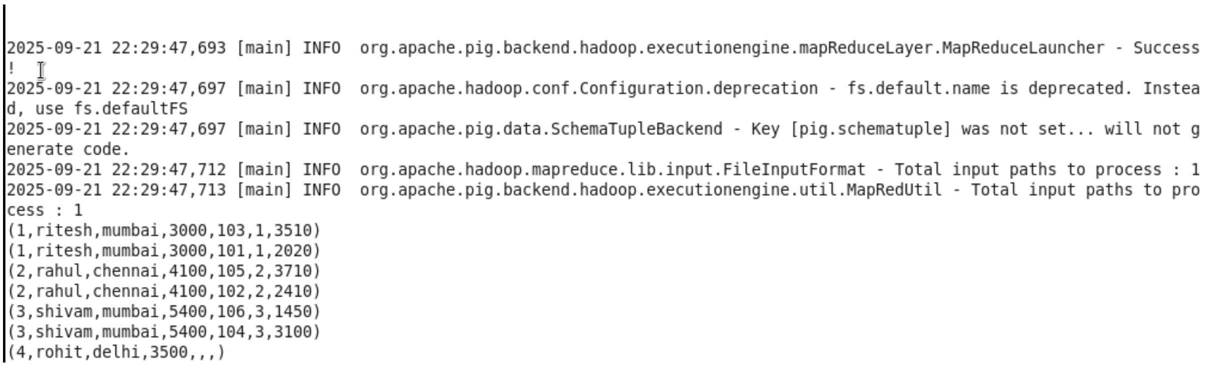
Left join //for left =>left outer for right =>right //



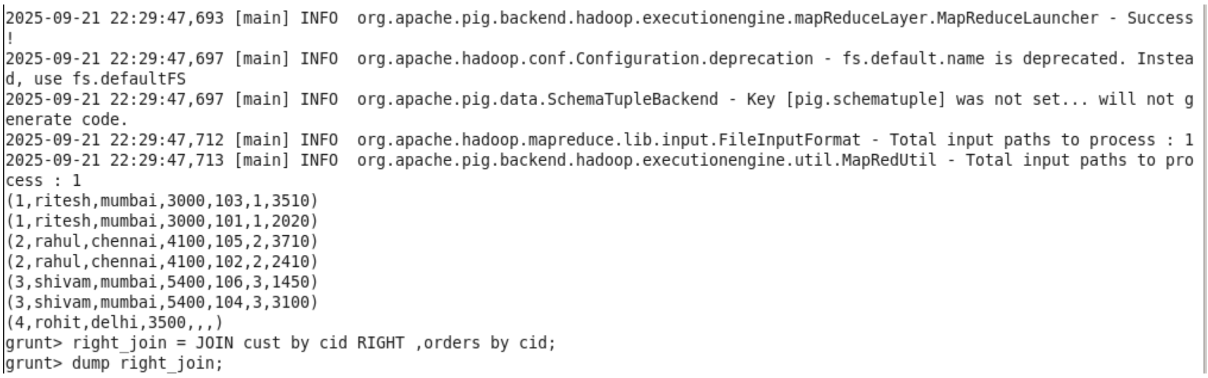
```
2025-09-21 22:27:59,964 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:27:59,964 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:27:59,965 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:27:59,974 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:27:59,974 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
grunt> S
```



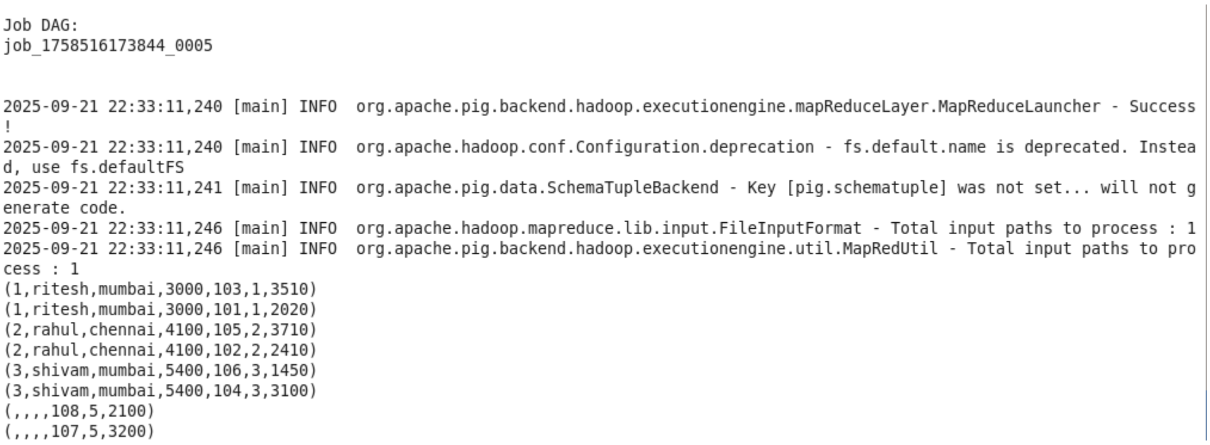
```
2025-09-21 22:27:59,964 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:27:59,964 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:27:59,965 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:27:59,974 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:27:59,974 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
grunt> left_join = JOIN cust by cid LEFT OUTER,orders by cid;
grunt> dump left_join;
```



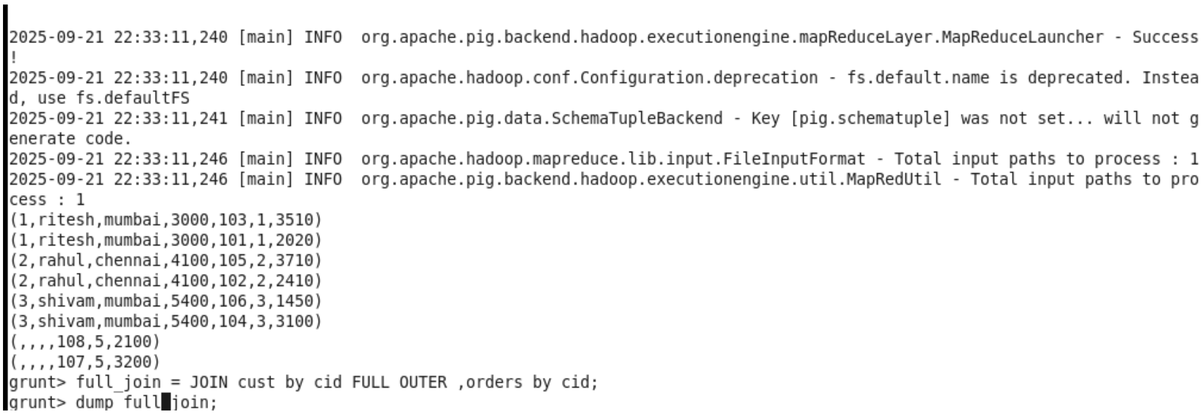
```
2025-09-21 22:29:47,693 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:29:47,697 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:29:47,697 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:29:47,712 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:29:47,713 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
(4,rohit,delhi,3500,,,)
```



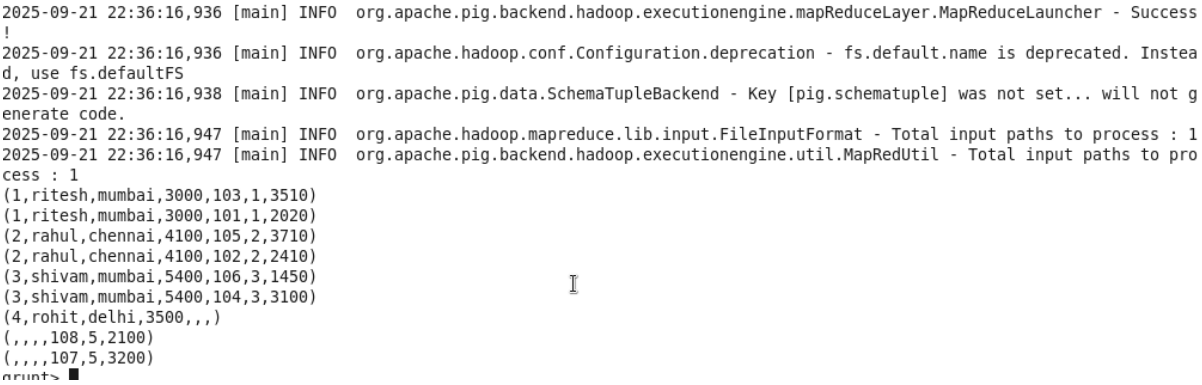
```
2025-09-21 22:29:47,693 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:29:47,697 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:29:47,697 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:29:47,712 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:29:47,713 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
(4,rohit,delhi,3500,,,)
grunt> right_join = JOIN cust by cid RIGHT ,orders by cid;
grunt> dump right_join;
```



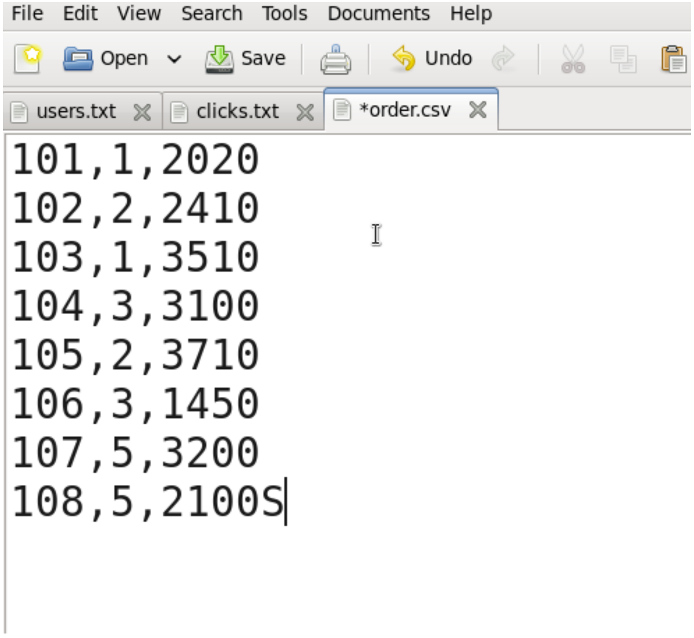
```
Job DAG:
job_1758516173844_0005


2025-09-21 22:33:11,240 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:33:11,240 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:33:11,241 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:33:11,246 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:33:11,246 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
(,,,,108,5,2100)
(,,,,107,5,3200)
```



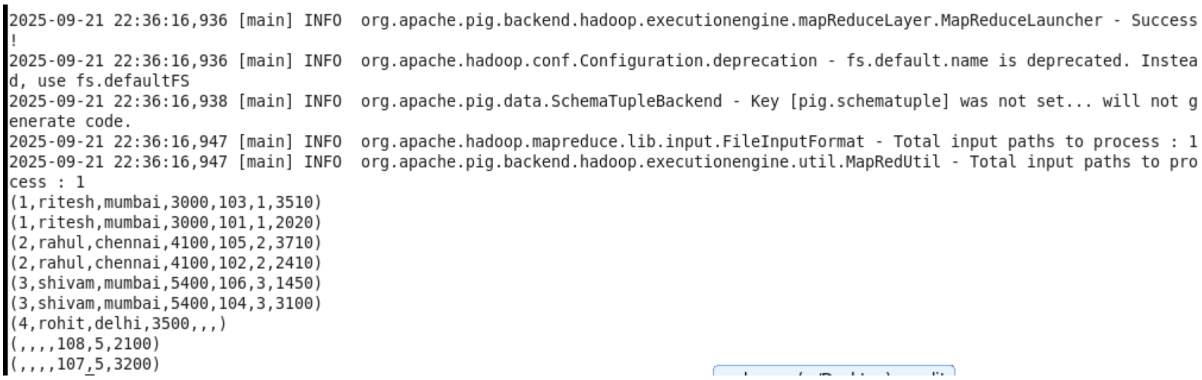
```
2025-09-21 22:33:11,240 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:33:11,240 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:33:11,241 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:33:11,246 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:33:11,246 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
(,,,,108,5,2100)
(,,,,107,5,3200)
grunt> full_join = JOIN cust by cid FULL OUTER ,orders by cid;
grunt> dump full_join;
```



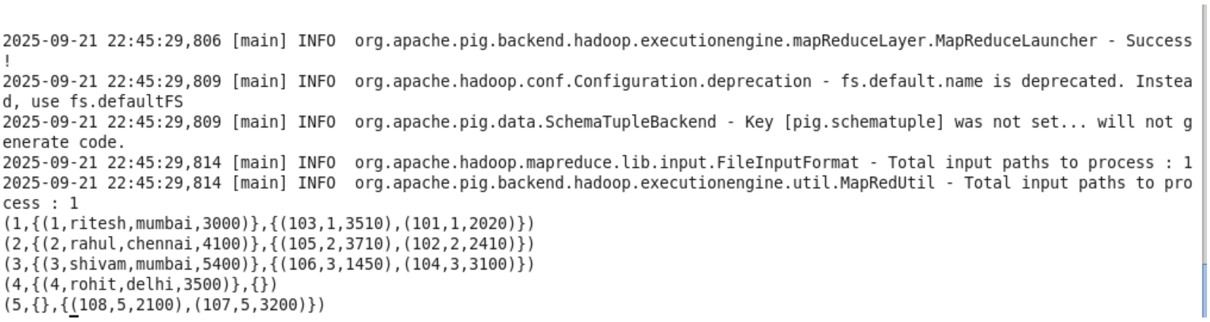
```
2025-09-21 22:36:16,936 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:36:16,936 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:36:16,938 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:36:16,947 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:36:16,947 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
(4,rohit,delhi,3500,,,)
(,,,,108,5,2100)
(,,,,107,5,3200)
grunt>
```



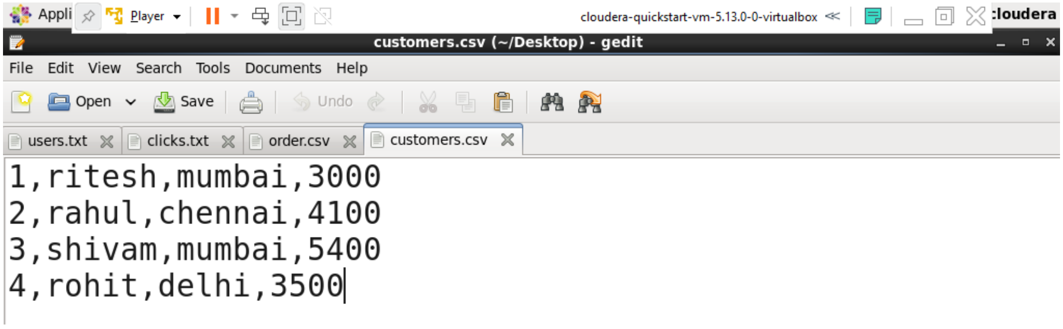
```
101,1,2020
102,2,2410
103,1,3510
104,3,3100
105,2,3710
106,3,1450
107,5,3200
108,5,2100S
```



```
2025-09-21 22:36:16,936 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:36:16,936 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:36:16,938 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:36:16,947 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:36:16,947 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,ritesh,mumbai,3000,103,1,3510)
(1,ritesh,mumbai,3000,101,1,2020)
(2,rahul,chennai,4100,105,2,3710)
(2,rahul,chennai,4100,102,2,2410)
(3,shivam,mumbai,5400,106,3,1450)
(3,shivam,mumbai,5400,104,3,3100)
(4,rohit,delhi,3500,,,)
(,,,,108,5,2100)
(,,,,107,5,3200)
```

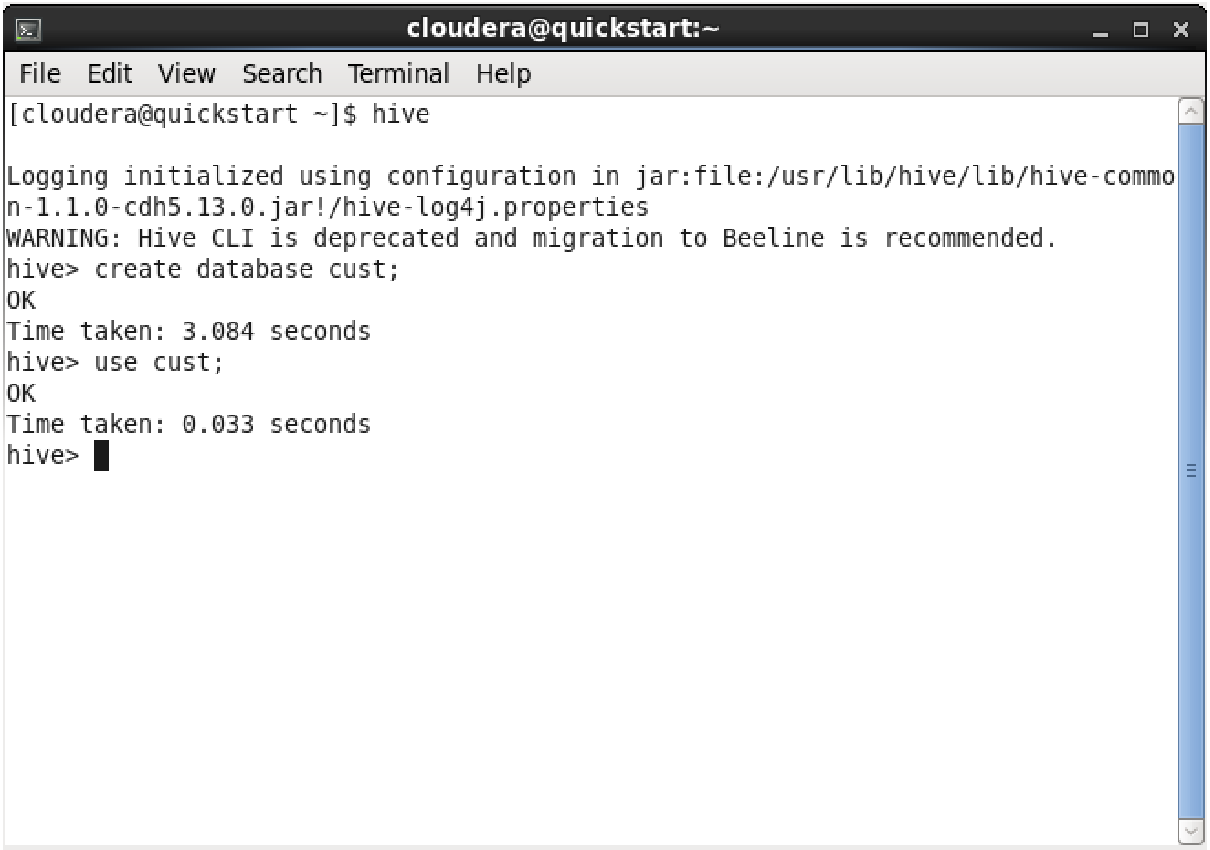


```
2025-09-21 22:45:29,806 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2025-09-21 22:45:29,809 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2025-09-21 22:45:29,809 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not g
enerate code.
2025-09-21 22:45:29,814 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2025-09-21 22:45:29,814 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(1,{(1,ritesh,mumbai,3000)},{(103,1,3510),(101,1,2020)})
(2,{(2,rahul,chennai,4100)},{(105,2,3710),(102,2,2410)})
(3,{(3,shivam,mumbai,5400)},{(106,3,1450),(104,3,3100)})
(4,{(4,rohit,delhi,3500)},{})
(5,{},{(108,5,2100),(107,5,3200)})
```

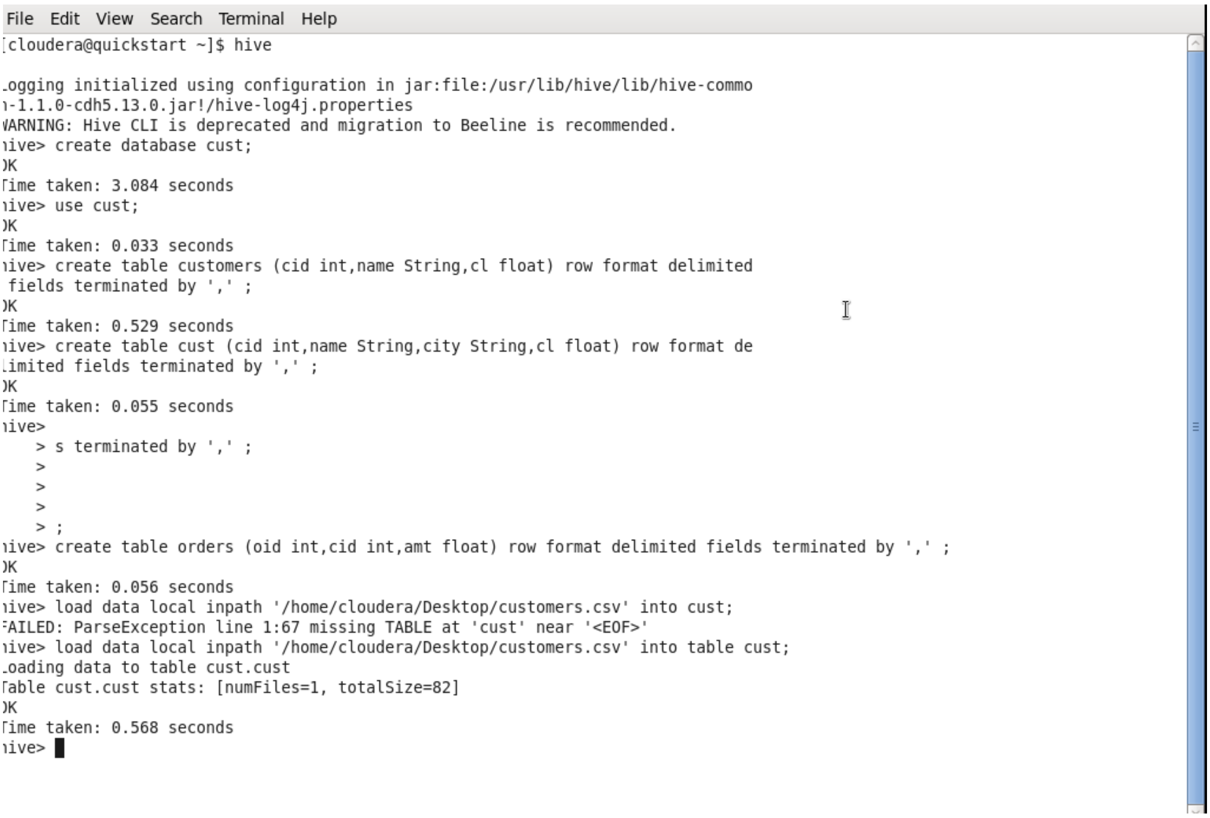


Appli
Player
cloudera-quickstart-vm-5.13.0-0-virtualbox
cloudera
customers.csv (~/Desktop) - gedit
File Edit View Search Tools Documents Help
Open
Save
Undo
users.txt
clicks.txt
order.csv
customers.csv

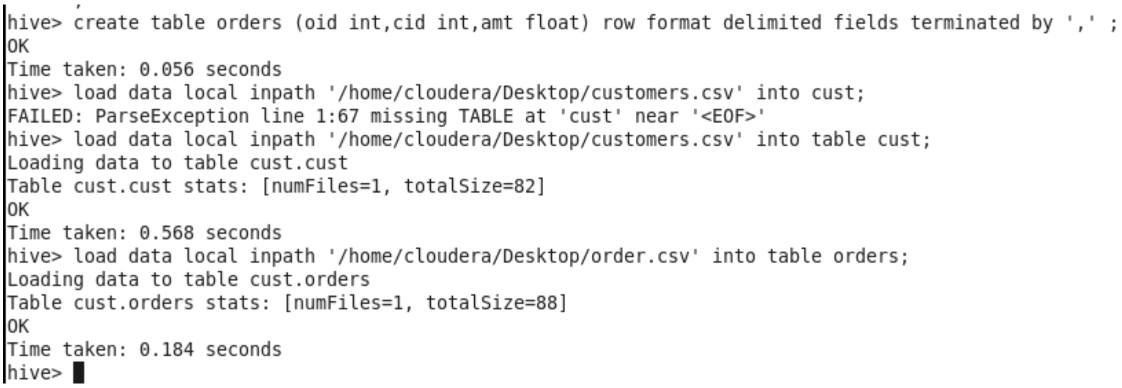
```
1,ritesh,mumbai,3000
2,rahul,chennai,4100
3,shivam,mumbai,5400
4,rohit,delhi,3500
```



```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create database cust;
OK
Time taken: 3.084 seconds
hive> use cust;
OK
Time taken: 0.033 seconds
hive> 
```

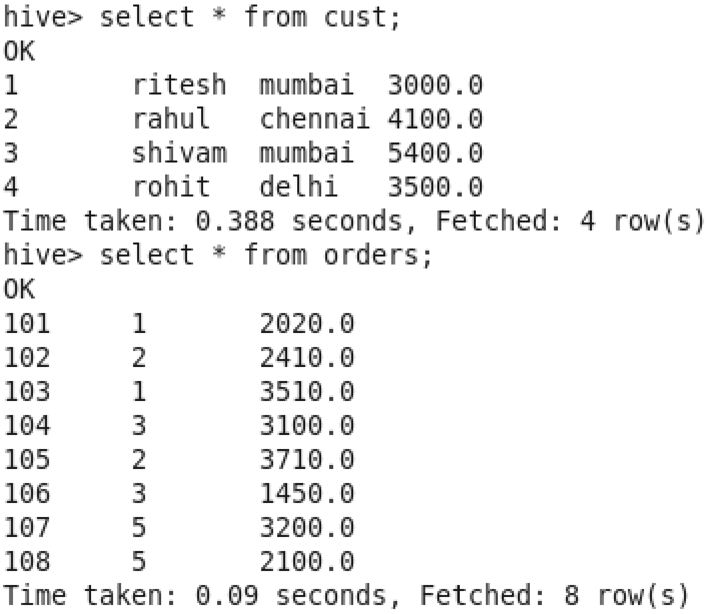


```
[cloudera@quickstart ~]$ hive

ogging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
ARNING: Hive CLI is deprecated and migration to Beeline is recommended.
ive> create database cust;
K
ime taken: 3.084 seconds
ive> use cust;
K
ime taken: 0.033 seconds
ive> create table customers (cid int,name String,cl float) row format delimited
 fields terminated by ',' ;
K
ime taken: 0.529 seconds
ive> create table cust (cid int,name String,city String,cl float) row format de
imited fields terminated by ',' ;
K
ime taken: 0.055 seconds
ive>
    > s terminated by ',' ;
    >
    >
    >
    > ;
ive> create table orders (oid int,cid int,amt float) row format delimited fields terminated by ',' ;
K
ime taken: 0.056 seconds
ive> load data local inpath '/home/cloudera/Desktop/customers.csv' into cust;
AILED: ParseException line 1:67 missing TABLE at 'cust' near '<EOF>'
ive> load data local inpath '/home/cloudera/Desktop/customers.csv' into table cust;
oading data to table cust.cust
able cust.cust stats: [numFiles=1, totalSize=82]
K
ime taken: 0.568 seconds
ive> █
```

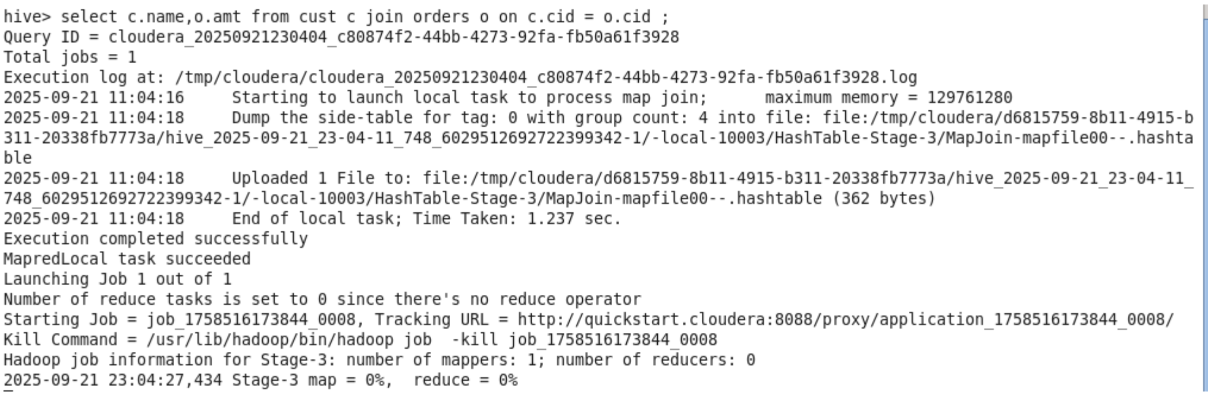


```
hive> create table orders (oid int,cid int,amt float) row format delimited fields terminated by ',' ;
OK
Time taken: 0.056 seconds
hive> load data local inpath '/home/cloudera/Desktop/customers.csv' into cust;
FAILED: ParseException line 1:67 missing TABLE at 'cust' near '<EOF>'
hive> load data local inpath '/home/cloudera/Desktop/customers.csv' into table cust;
Loading data to table cust.cust
Table cust.cust stats: [numFiles=1, totalSize=82]
OK
Time taken: 0.568 seconds
```

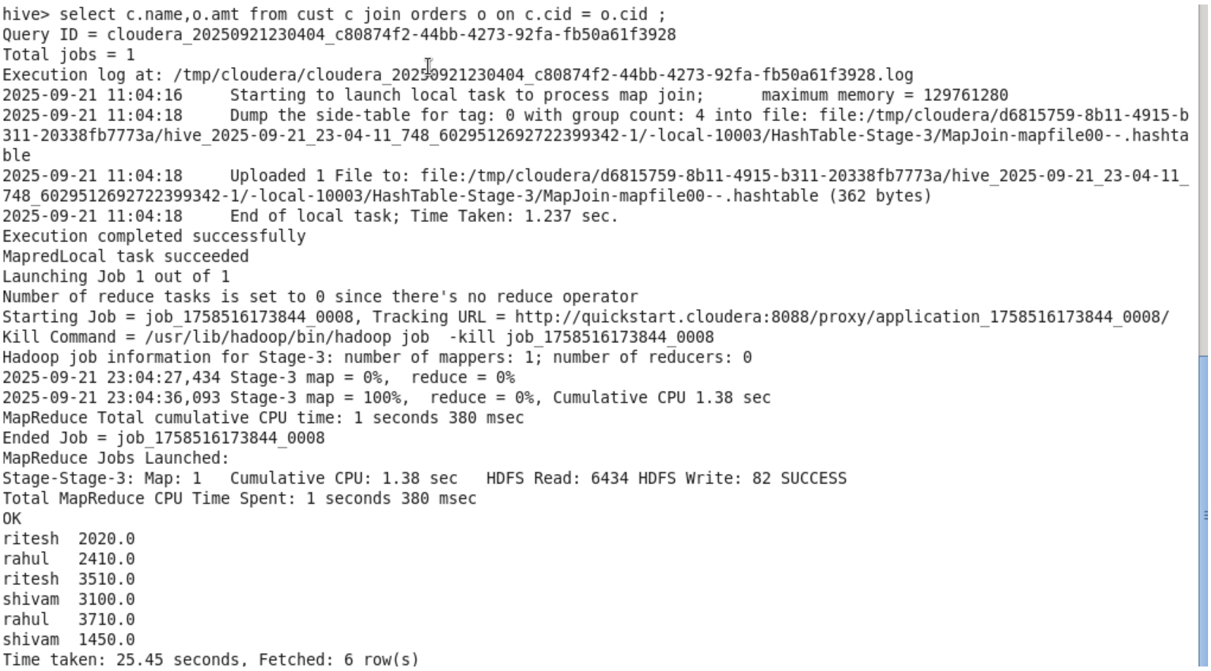
```
hive> load data local inpath '/home/cloudera/Desktop/order.csv' into table orders;
Loading data to table cust.orders
Table cust.orders stats: [numFiles=1, totalSize=88]
OK
Time taken: 0.184 seconds
hive> █
```



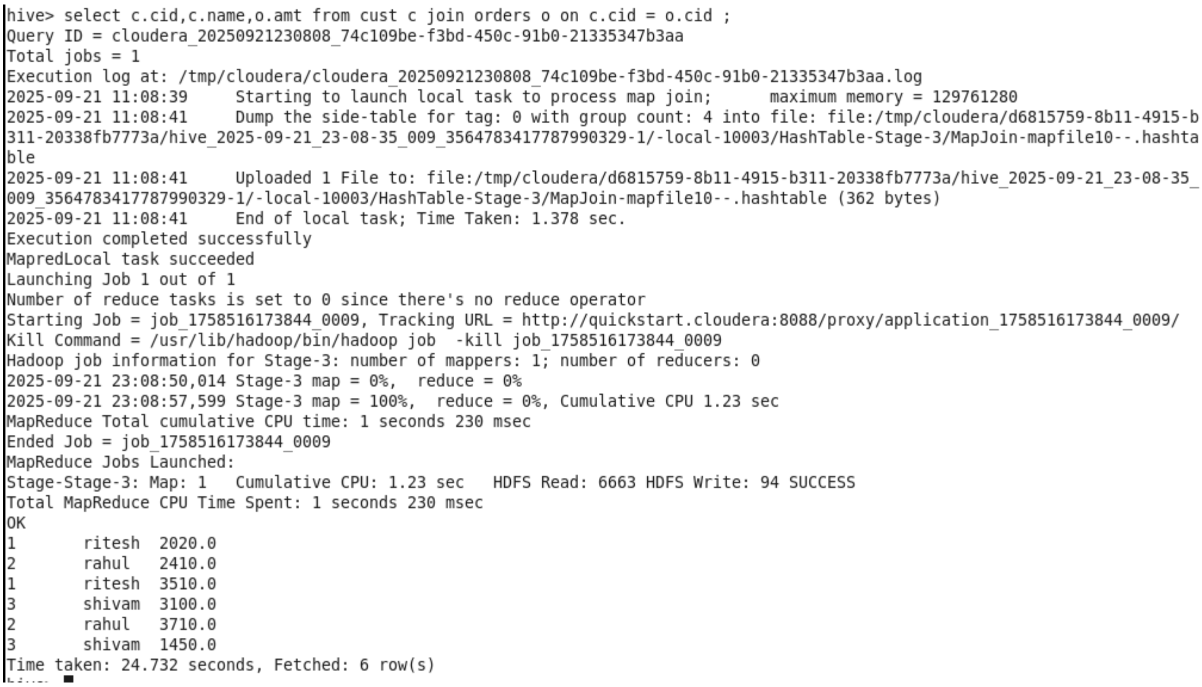
```
hive> select * from cust;
OK
1       ritesh  mumbai  3000.0
2       rahul   chennai 4100.0
3       shivam  mumbai  5400.0
4       rohit   delhi   3500.0
Time taken: 0.388 seconds, Fetched: 4 row(s)
hive> select * from orders;
OK
101     1       2020.0
102     2       2410.0
103     1       3510.0
104     3       3100.0
105     2       3710.0
106     3       1450.0
107     5       3200.0
108     5       2100.0
Time taken: 0.09 seconds, Fetched: 8 row(s)
```



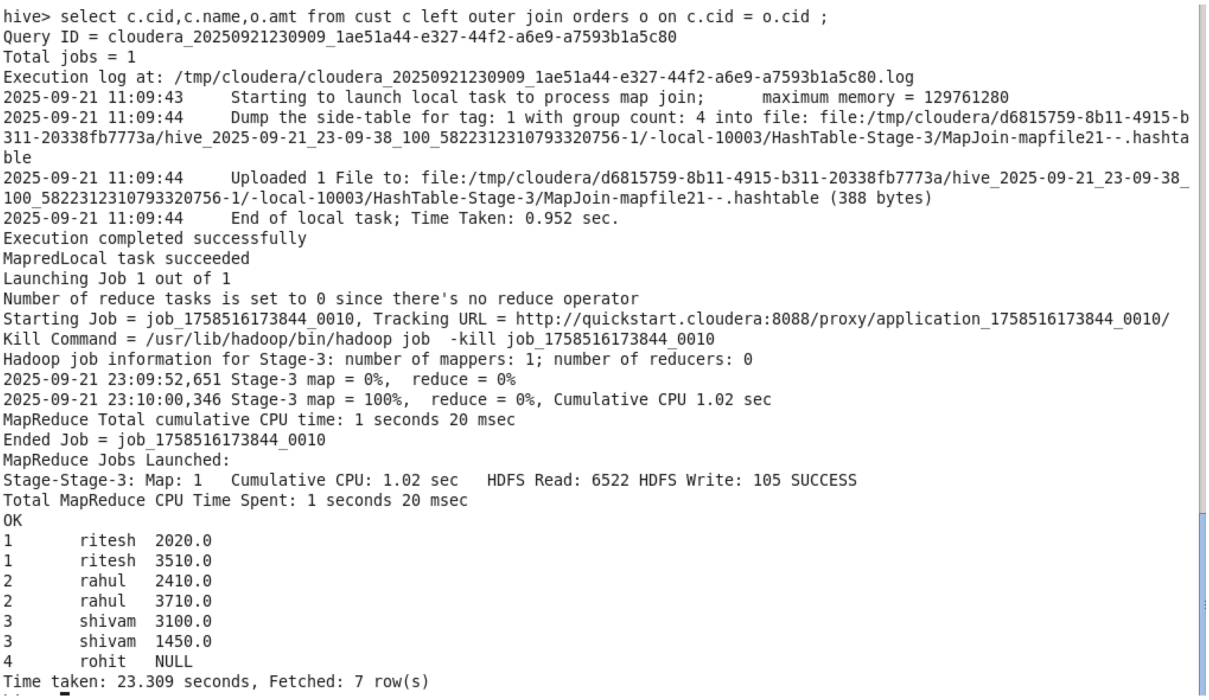
```
hive> select c.name,o.amt from cust c join orders o on c.cid = o.cid ;
Query ID = cloudera_20250921230404_c80874f2-44bb-4273-92fa-fb50a61f3928
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250921230404_c80874f2-44bb-4273-92fa-fb50a61f3928.log
2025-09-21 11:04:16     Starting to launch local task to process map join;      maximum memory = 129761280
2025-09-21 11:04:18     Dump the side-table for tag: 0 with group count: 4 into file: file:/tmp/cloudera/d6815759-8b11-4915-b
311-20338fb7773a/hive_2025-09-21_23-04-11_748_6029512692722399342-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashta
ble
2025-09-21 11:04:18     Uploaded 1 File to: file:/tmp/cloudera/d6815759-8b11-4915-b311-20338fb7773a/hive_2025-09-21_23-04-11_
748_6029512692722399342-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashtable (362 bytes)
2025-09-21 11:04:18     End of local task; Time Taken: 1.237 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_1758516173844_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1758516173844_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758516173844_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-09-21 23:04:27,434 Stage-3 map = 0%,  reduce = 0%
```



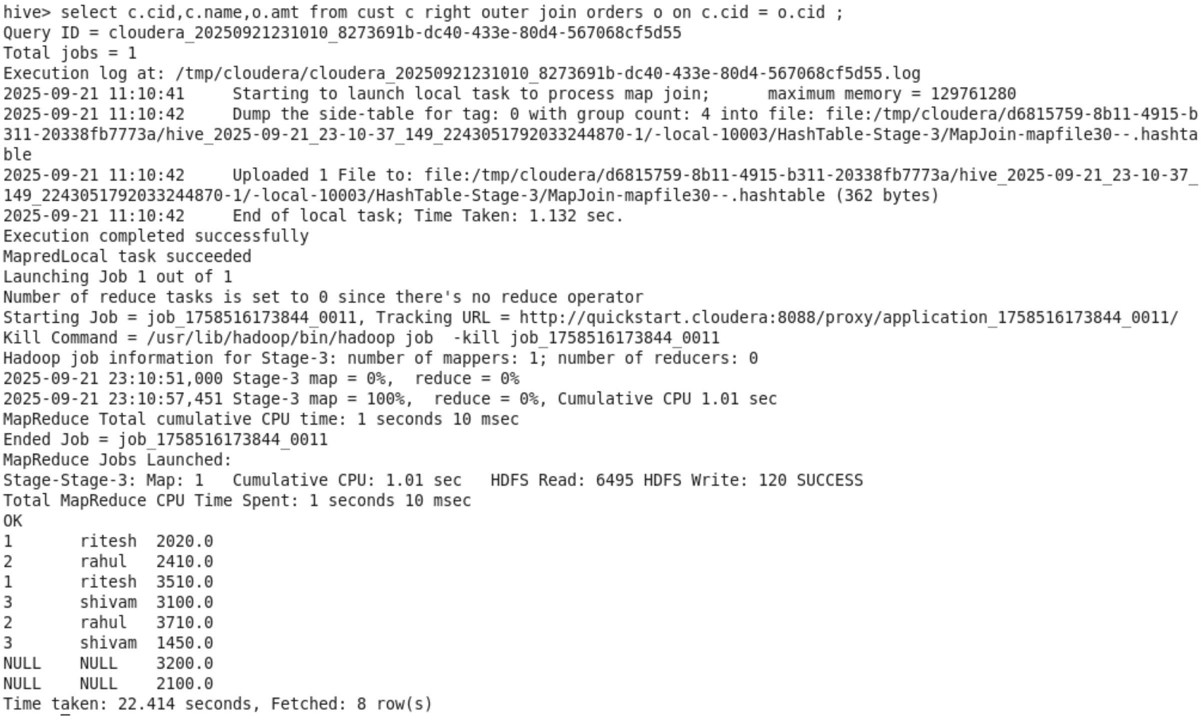
```
hive> select c.name,o.amt from cust c join orders o on c.cid = o.cid ;
Query ID = cloudera_20250921230404_c80874f2-44bb-4273-92fa-fb50a61f3928
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250921230404_c80874f2-44bb-4273-92fa-fb50a61f3928.log
2025-09-21 11:04:16     Starting to launch local task to process map join;      maximum memory = 129761280
2025-09-21 11:04:18     Dump the side-table for tag: 0 with group count: 4 into file: file:/tmp/cloudera/d6815759-8b11-4915-b
311-20338fb7773a/hive_2025-09-21_23-04-11_748_6029512692722399342-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashta
ble
2025-09-21 11:04:18     Uploaded 1 File to: file:/tmp/cloudera/d6815759-8b11-4915-b311-20338fb7773a/hive_2025-09-21_23-04-11_
748_6029512692722399342-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile00--.hashtable (362 bytes)
2025-09-21 11:04:18     End of local task; Time Taken: 1.237 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_1758516173844_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1758516173844_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758516173844_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-09-21 23:04:27,434 Stage-3 map = 0%,  reduce = 0%
2025-09-21 23:04:36,093 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.38 sec
MapReduce Total cumulative CPU time: 1 seconds 380 msec
Ended Job = job_1758516173844_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.38 sec   HDFS Read: 6434 HDFS Write: 82 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 380 msec
OK
ritesh  2020.0
rahul   2410.0
ritesh  3510.0
shivam  3100.0
rahul   3710.0
shivam  1450.0
Time taken: 25.45 seconds, Fetched: 6 row(s)
```



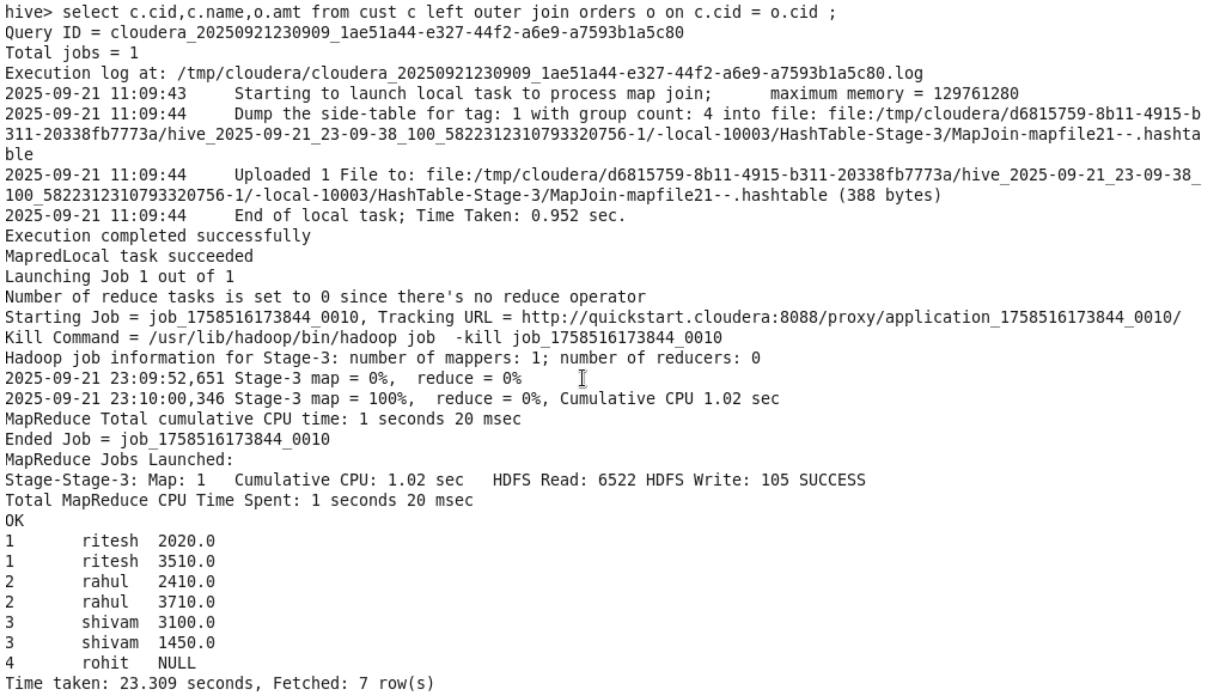
```
hive> select c.cid,c.name,o.amt from cust c join orders o on c.cid = o.cid ;
Query ID = cloudera_20250921230808_74c109be-f3bd-450c-91b0-21335347b3aa
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250921230808_74c109be-f3bd-450c-91b0-21335347b3aa.log
2025-09-21 11:08:39     Starting to launch local task to process map join;      maximum memory = 129761280
2025-09-21 11:08:41     Dump the side-table for tag: 0 with group count: 4 into file: file:/tmp/cloudera/d6815759-8b11-4915-b
311-20338fb7773a/hive_2025-09-21_23-08-35_009_3564783417787990329-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile10--.hashta
ble
2025-09-21 11:08:41     Uploaded 1 File to: file:/tmp/cloudera/d6815759-8b11-4915-b311-20338fb7773a/hive_2025-09-21_23-08-35_
009_3564783417787990329-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile10--.hashtable (362 bytes)
2025-09-21 11:08:41     End of local task; Time Taken: 1.378 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_1758516173844_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1758516173844_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758516173844_0009
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-09-21 23:08:50,014 Stage-3 map = 0%,  reduce = 0%
2025-09-21 23:08:57,599 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.23 sec
MapReduce Total cumulative CPU time: 1 seconds 230 msec
Ended Job = job_1758516173844_0009
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.23 sec   HDFS Read: 6663 HDFS Write: 94 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 230 msec
OK
1       ritesh  2020.0
2       rahul   2410.0
1       ritesh  3510.0
3       shivam  3100.0
2       rahul   3710.0
3       shivam  1450.0
Time taken: 24.732 seconds, Fetched: 6 row(s)
hive> 
```



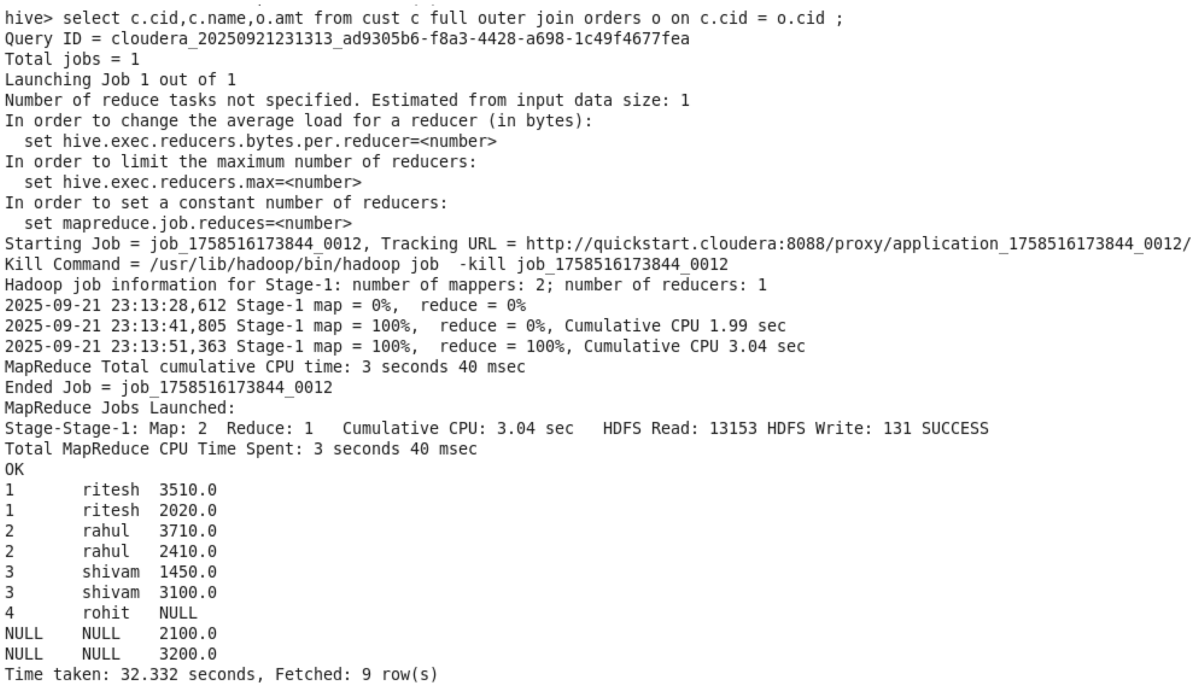
```
hive> select c.cid,c.name,o.amt from cust c left outer join orders o on c.cid = o.cid ;
Query ID = cloudera_20250921230909_1ae51a44-e327-44f2-a6e9-a7593b1a5c80
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250921230909_1ae51a44-e327-44f2-a6e9-a7593b1a5c80.log
2025-09-21 11:09:43     Starting to launch local task to process map join;      maximum memory = 129761280
2025-09-21 11:09:44     Dump the side-table for tag: 1 with group count: 4 into file: file:/tmp/cloudera/d6815759-8b11-4915-b
311-20338fb7773a/hive_2025-09-21_23-09-38_100_5822312310793320756-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile21--.hashta
ble
2025-09-21 11:09:44     Uploaded 1 File to: file:/tmp/cloudera/d6815759-8b11-4915-b311-20338fb7773a/hive_2025-09-21_23-09-38_
100_5822312310793320756-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile21--.hashtable (388 bytes)
2025-09-21 11:09:44     End of local task; Time Taken: 0.952 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_1758516173844_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1758516173844_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758516173844_0010
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-09-21 23:09:52,651 Stage-3 map = 0%,  reduce = 0%
2025-09-21 23:10:00,346 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.02 sec
MapReduce Total cumulative CPU time: 1 seconds 20 msec
Ended Job = job_1758516173844_0010
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.02 sec   HDFS Read: 6522 HDFS Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 20 msec
OK
1       ritesh  2020.0
1       ritesh  3510.0
2       rahul   2410.0
2       rahul   3710.0
3       shivam  3100.0
3       shivam  1450.0
4       rohit   NULL
Time taken: 23.309 seconds, Fetched: 7 row(s)
```



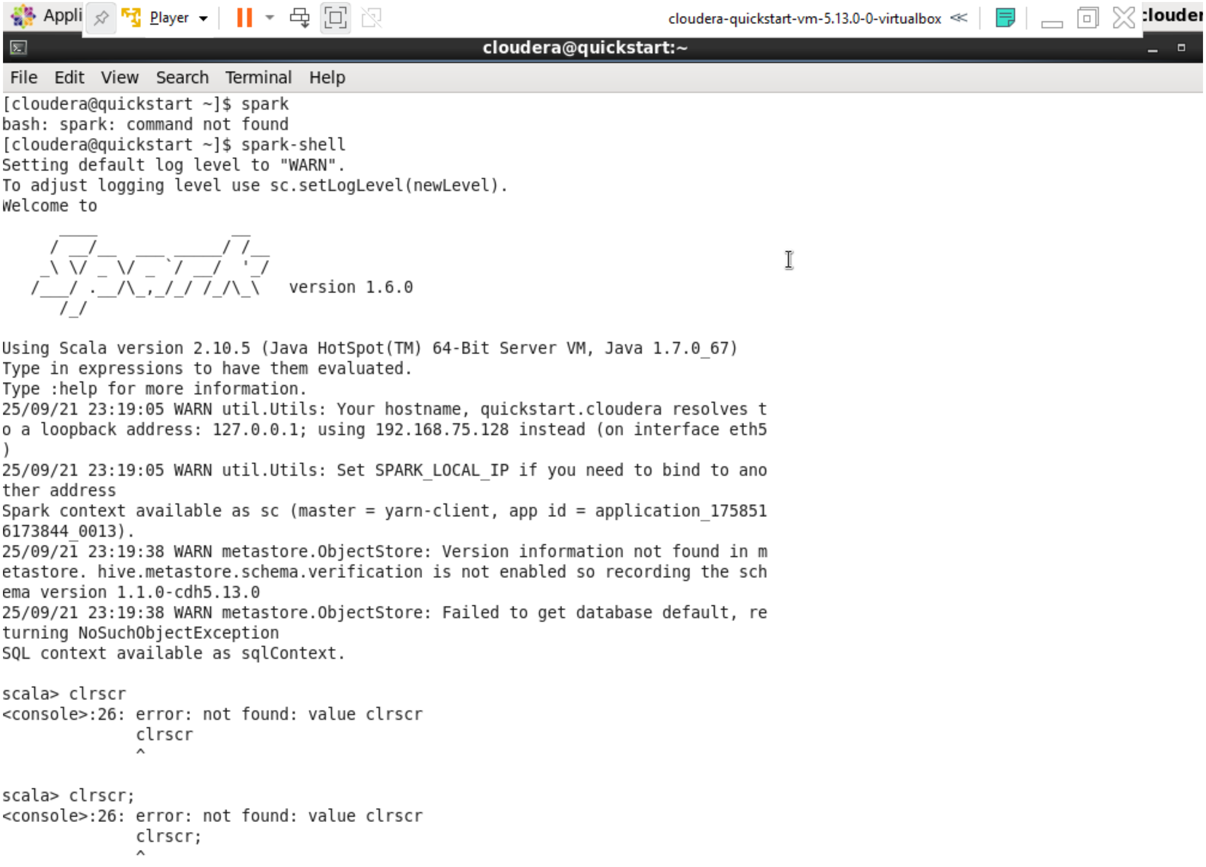
```
hive> select c.cid,c.name,o.amt from cust c right outer join orders o on c.cid = o.cid ;
Query ID = cloudera_20250921231010_8273691b-dc40-433e-80d4-567068cf5d55
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250921231010_8273691b-dc40-433e-80d4-567068cf5d55.log
2025-09-21 11:10:41     Starting to launch local task to process map join;      maximum memory = 129761280
2025-09-21 11:10:42     Dump the side-table for tag: 0 with group count: 4 into file: file:/tmp/cloudera/d6815759-8b11-4915-b
311-20338fb7773a/hive_2025-09-21_23-10-37_149_2243051792033244870-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile30--.hashta
ble
2025-09-21 11:10:42     Uploaded 1 File to: file:/tmp/cloudera/d6815759-8b11-4915-b311-20338fb7773a/hive_2025-09-21_23-10-37_
149_2243051792033244870-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile30--.hashtable (362 bytes)
2025-09-21 11:10:42     End of local task; Time Taken: 1.132 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_1758516173844_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1758516173844_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758516173844_0011
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-09-21 23:10:51,000 Stage-3 map = 0%,  reduce = 0%
2025-09-21 23:10:57,451 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.01 sec
MapReduce Total cumulative CPU time: 1 seconds 10 msec
Ended Job = job_1758516173844_0011
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.01 sec   HDFS Read: 6495 HDFS Write: 120 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 10 msec
OK
1       ritesh  2020.0
2       rahul   2410.0
1       ritesh  3510.0
3       shivam  3100.0
2       rahul   3710.0
3       shivam  1450.0
NULL    NULL    3200.0
NULL    NULL    2100.0
Time taken: 22.414 seconds, Fetched: 8 row(s)
```



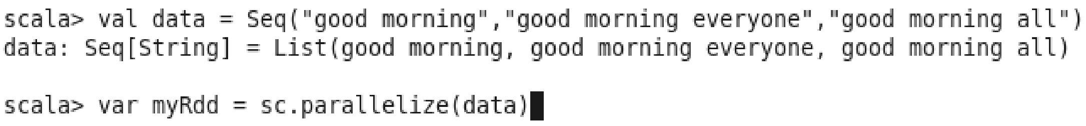
```
hive> select c.cid,c.name,o.amt from cust c left outer join orders o on c.cid = o.cid ;
Query ID = cloudera_20250921230909_1ae51a44-e327-44f2-a6e9-a7593b1a5c80
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250921230909_1ae51a44-e327-44f2-a6e9-a7593b1a5c80.log
2025-09-21 11:09:43     Starting to launch local task to process map join;      maximum memory = 129761280
2025-09-21 11:09:44     Dump the side-table for tag: 1 with group count: 4 into file: file:/tmp/cloudera/d6815759-8b11-4915-b
311-20338fb7773a/hive_2025-09-21_23-09-38_100_5822312310793320756-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile21--.hashta
ble
2025-09-21 11:09:44     Uploaded 1 File to: file:/tmp/cloudera/d6815759-8b11-4915-b311-20338fb7773a/hive_2025-09-21_23-09-38_
100_5822312310793320756-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile21--.hashtable (388 bytes)
2025-09-21 11:09:44     End of local task; Time Taken: 0.952 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_1758516173844_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1758516173844_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758516173844_0010
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2025-09-21 23:09:52,651 Stage-3 map = 0%,  reduce = 0%
2025-09-21 23:10:00,346 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.02 sec
MapReduce Total cumulative CPU time: 1 seconds 20 msec
Ended Job = job_1758516173844_0010
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.02 sec   HDFS Read: 6522 HDFS Write: 105 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 20 msec
OK
1       ritesh  2020.0
1       ritesh  3510.0
2       rahul   2410.0
2       rahul   3710.0
3       shivam  3100.0
3       shivam  1450.0
4       rohit   NULL
Time taken: 23.309 seconds, Fetched: 7 row(s)
```



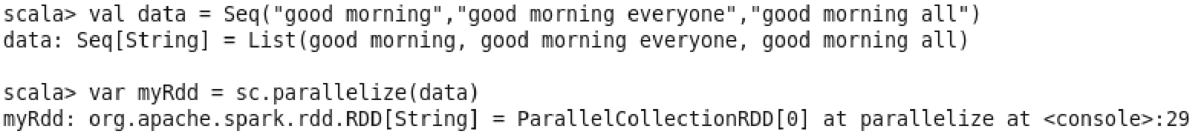
```
hive> select c.cid,c.name,o.amt from cust c full outer join orders o on c.cid = o.cid ;
Query ID = cloudera_20250921231313_ad9305b6-f8a3-4428-a698-1c49f4677fea
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average 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_1758516173844_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1758516173844_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1758516173844_0012
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2025-09-21 23:13:28,612 Stage-1 map = 0%,  reduce = 0%
2025-09-21 23:13:41,805 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.99 sec
2025-09-21 23:13:51,363 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.04 sec
MapReduce Total cumulative CPU time: 3 seconds 40 msec
Ended Job = job_1758516173844_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2  Reduce: 1   Cumulative CPU: 3.04 sec   HDFS Read: 13153 HDFS Write: 131 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 40 msec
OK
1       ritesh  3510.0
1       ritesh  2020.0
2       rahul   3710.0
2       rahul   2410.0
3       shivam  1450.0
3       shivam  3100.0
4       rohit   NULL
NULL    NULL    2100.0
NULL    NULL    3200.0
Time taken: 32.332 seconds, Fetched: 9 row(s)
```


Appli
Player
cloudera-quickstart-vm-5.13.0-0-virtualbox
:louder
cloudera@quickstart:~
File Edit View Search Terminal Help

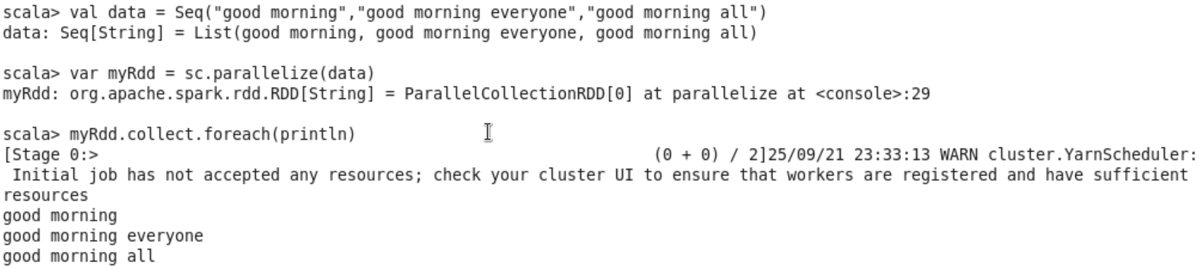
```
[cloudera@quickstart ~]$ spark
bash: spark: command not found
[cloudera@quickstart ~]$ spark-shell
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel).
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 1.6.0
      /_/

Using Scala version 2.10.5 (Java HotSpot(TM) 64-Bit Server VM, Java 1.7.0_67)
Type in expressions to have them evaluated.
Type :help for more information.
25/09/21 23:19:05 WARN util.Utils: Your hostname, quickstart.cloudera resolves t
o a loopback address: 127.0.0.1; using 192.168.75.128 instead (on interface eth5
)
25/09/21 23:19:05 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to ano
ther address
Spark context available as sc (master = yarn-client, app id = application_175851
6173844_0013).
25/09/21 23:19:38 WARN metastore.ObjectStore: Version information not found in m
etastore. hive.metastore.schema.verification is not enabled so recording the sch
ema version 1.1.0-cdh5.13.0
25/09/21 23:19:38 WARN metastore.ObjectStore: Failed to get database default, re
turning NoSuchObjectException
SQL context available as sqlContext.

scala> clrscr
<console>:26: error: not found: value clrscr
              clrscr
              ^

scala> clrscr;
<console>:26: error: not found: value clrscr
              clrscr;
              ^
```



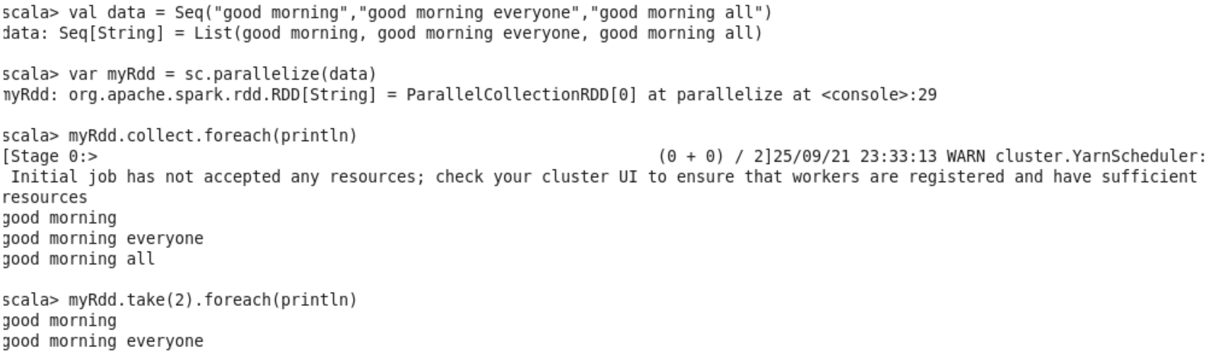
```
scala> val data = Seq("good morning","good morning everyone","good morning all")
data: Seq[String] = List(good morning, good morning everyone, good morning all)

scala> var myRdd = sc.parallelize(data)
```



```
scala> val data = Seq("good morning","good morning everyone","good morning all")
data: Seq[String] = List(good morning, good morning everyone, good morning all)

scala> var myRdd = sc.parallelize(data)
myRdd: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[0] at parallelize at <console>:29
```



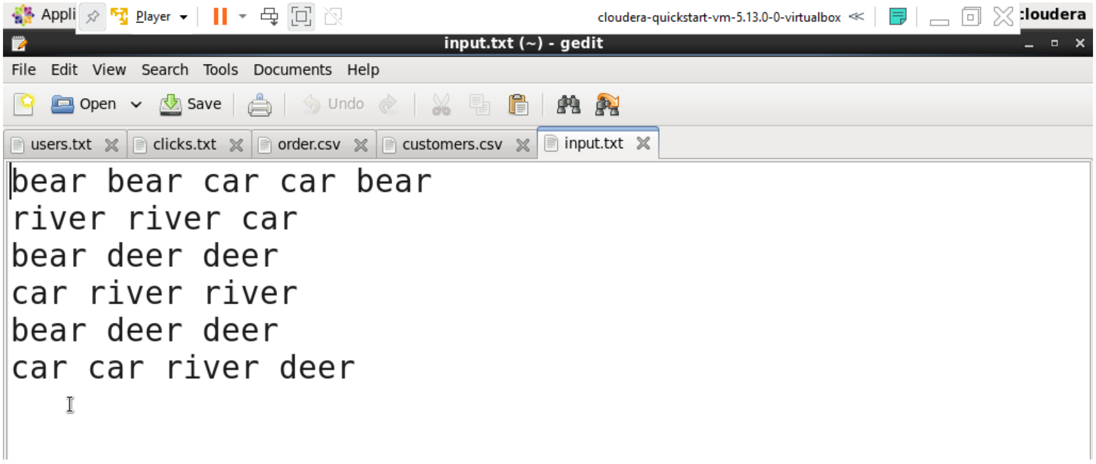
```
scala> val data = Seq("good morning","good morning everyone","good morning all")
data: Seq[String] = List(good morning, good morning everyone, good morning all)

scala> var myRdd = sc.parallelize(data)
myRdd: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[0] at parallelize at <console>:29

scala> myRdd.collect.foreach(println)
[Stage 0:>                                                          (0 + 0) / 2]25/09/21 23:33:13 WARN cluster.YarnScheduler:
 Initial job has not accepted any resources; check your cluster UI to ensure that workers are registered and have sufficient
resources
good morning
good morning everyone
good morning all
```



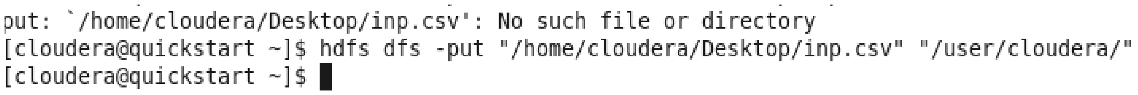
```
scala> val data = Seq("good morning","good morning everyone","good morning all")
data: Seq[String] = List(good morning, good morning everyone, good morning all)

scala> var myRdd = sc.parallelize(data)
myRdd: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[0] at parallelize at <console>:29

scala> myRdd.collect.foreach(println)
[Stage 0:>                                                          (0 + 0) / 2]25/09/21 23:33:13 WARN cluster.YarnScheduler:
 Initial job has not accepted any resources; check your cluster UI to ensure that workers are registered and have sufficient
resources
good morning
good morning everyone
good morning all

scala> myRdd.take(2).foreach(println)
good morning
good morning everyone
```

```
scala> val data = Seq("good morning","good morning everyone","good morning all")
data: Seq[String] = List(good morning, good morning everyone, good morning all)

scala> var myRdd = sc.parallelize(data)
myRdd: org.apache.spark.rdd.RDD[String] = ParallelCollectionRDD[0] at parallelize at <console>:29

scala> myRdd.collect.foreach(println)
[Stage 0:>                                                          (0 + 0) / 2]25/09/21 23:33:13 WARN cluster.YarnScheduler:
 Initial job has not accepted any resources; check your cluster UI to ensure that workers are registered and have sufficient
resources
good morning
good morning everyone
good morning all

scala> myRdd.take(2).foreach(println)
good morning
good morning everyone

scala> myRdd.toDF().show()
25/09/21 23:35:57 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to kill executor 2!
+--------------------+
|                  _1|
+--------------------+
|        good morning|
|good morning ever...|
|    good morning all|
+--------------------+

scala> myRdd.take(2).foreach(println)
good morning
good morning everyone

scala> myRdd.toDF().show()
25/09/21 23:35:57 WARN spark.ExecutorAllocationManager: Unable to reach the cluster manager to kill executor 2!
+--------------------+
|                  _1|
+--------------------+
|        good morning|
|good morning ever...|
|    good morning all|
+--------------------+


scala> var wordsRdd = myRdd.flatMap(f=>f.split(" '))
<console>:1: error: unclosed string literal
       var wordsRdd = myRdd.flatMap(f=>f.split(" '))
                                                ^

scala> var wordsRdd = myRdd.flatMap(f=>f.split(" "))
wordsRdd: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[3] at flatMap at <console>:31

scala> wordsRdd.collect.foreach(println)
good
morning
good
morning
everyone
good
morning
all


scala> var wordsRdd = myRdd.flatMap(f=>f.split(" '))
<console>:1: error: unclosed string literal
       var wordsRdd = myRdd.flatMap(f=>f.split(" '))
                                                ^

scala> var wordsRdd = myRdd.flatMap(f=>f.split(" "))
wordsRdd: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[3] at flatMap at <console>:31

scala> wordsRdd.collect.foreach(println)
good
morning
good
morning
everyone
good
morning
all

scala> val mapRdd = wordsRdd.Map(word=>(word,1))
<console>:33: error: value Map is not a member of org.apache.spark.rdd.RDD[String]
         val mapRdd = wordsRdd.Map(word=>(word,1))
                               ^

scala> val mapRdd = wordsRdd.map(word=>(word,1))
mapRdd: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[4] at map at <console>:33
```

```
scala> val mapRdd = wordsRdd.map(word=>(word,1))
mapRdd: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[4] at map at <console>:33

scala> mapRdd.collect.foreach(println)
(good,1)
(morning,1)
(good,1)
(morning,1)
(everyone,1)
(good,1)
(morning,1)
(all,1)

scala> val mapRdd = wordsRdd.Map(word=>(word,1))
<console>:33: error: value Map is not a member of org.apache.spark.rdd.RDD[String]
         val mapRdd = wordsRdd.Map(word=>(word,1))
                               ^

scala> val mapRdd = wordsRdd.map(word=>(word,1))
mapRdd: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[4] at map at <console>:33

scala> mapRdd.collect.foreach(println)
(good,1)
(morning,1)
(good,1)
(morning,1)
(everyone,1)
(good,1)
(morning,1)
(all,1)

scala> val reduceRdd = mapRdd.reduceByKey(_+_)
reduceRdd: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[5] at reduceByKey at <console>:35

scala> reduceRdd.collect.foreach(println)
(morning,3)
(all,1)
(good,3)
(everyone,1)
```

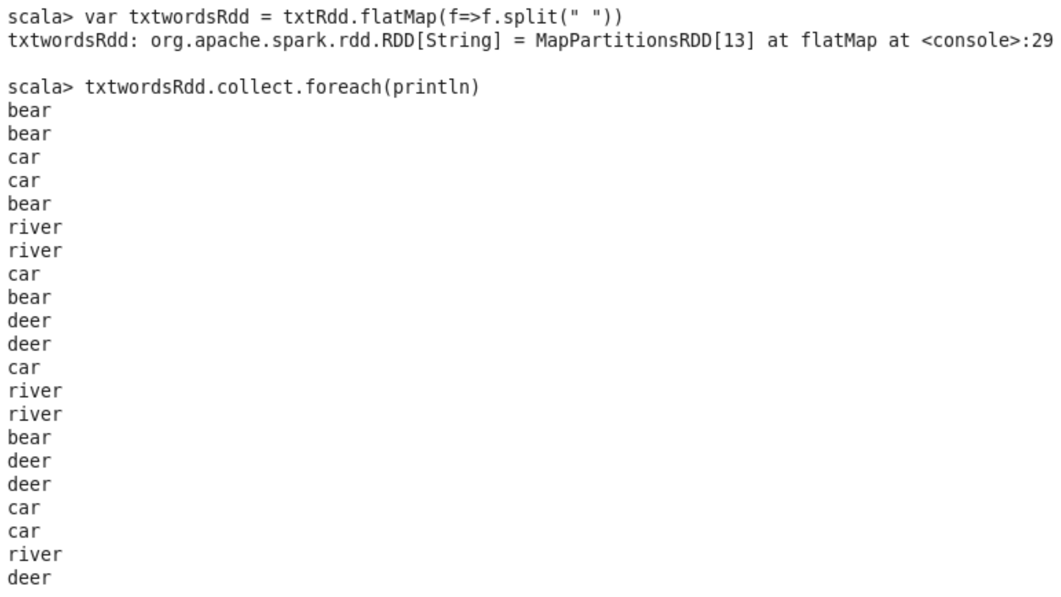


```
bear bear car car bear
river river car
bear deer deer
car river river
bear deer deer
car car river deer
```

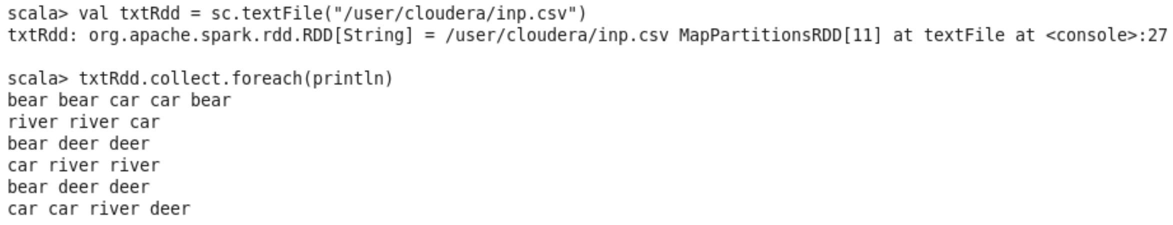


```
put: `/home/cloudera/Desktop/inp.csv': No such file or directory
[cloudera@quickstart ~]$ hdfs dfs -put "/home/cloudera/Desktop/inp.csv" "/user/cloudera/"
[cloudera@quickstart ~]$
```

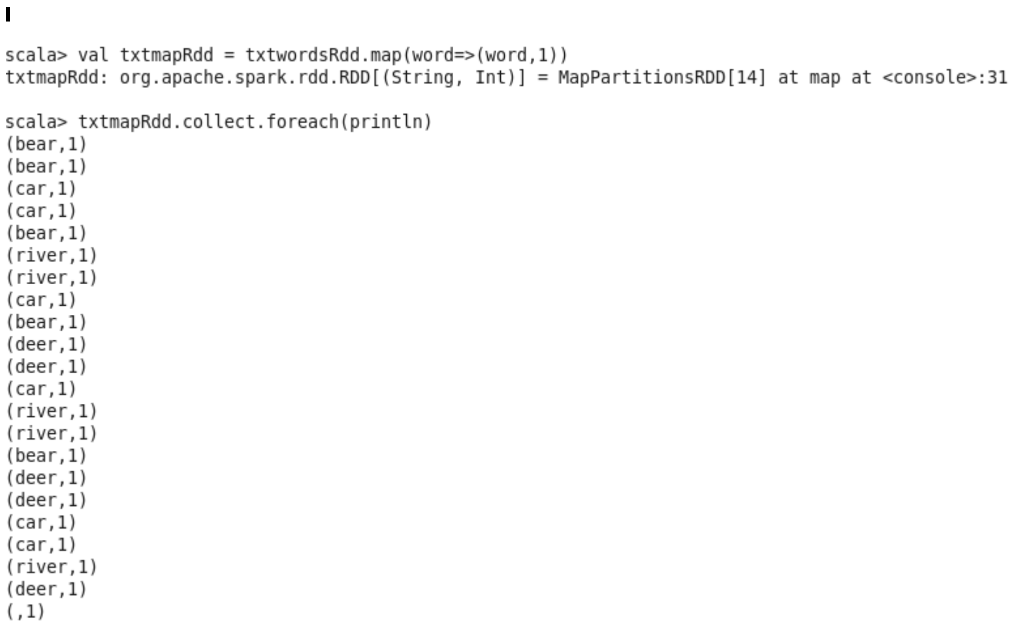
```
scala> val txtRdd = sc.textFile("/user/cloudera/inp.csv")
txtRdd: org.apache.spark.rdd.RDD[String] = /user/cloudera/inp.csv MapPartitionsRDD[11] at textFile at <console>:27

scala> txtRdd.collect.foreach(println)
bear bear car car bear
river river car
bear deer deer
car river river
bear deer deer
car car river deer
```



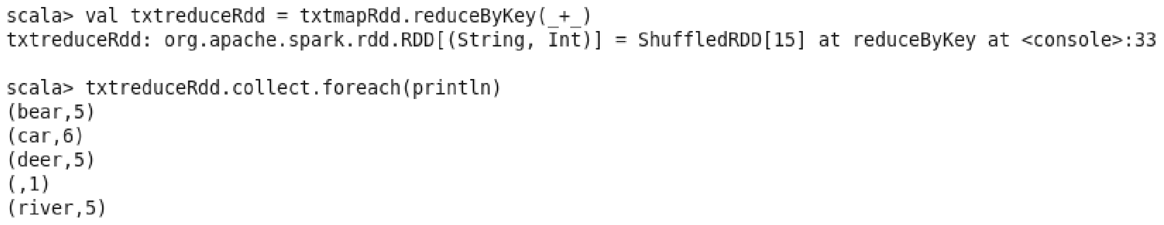
```
scala> var txtwordsRdd = txtRdd.flatMap(f=>f.split(" "))
txtwordsRdd: org.apache.spark.rdd.RDD[String] = MapPartitionsRDD[13] at flatMap at <console>:29

scala> txtwordsRdd.collect.foreach(println)
bear
bear
car
car
bear
river
river
car
bear
deer
deer
car
river
river
bear
deer
deer
car
car
river
deer
```



```
scala> val txtRdd = sc.textFile("/user/cloudera/inp.csv")
txtRdd: org.apache.spark.rdd.RDD[String] = /user/cloudera/inp.csv MapPartitionsRDD[11] at textFile at <console>:27

scala> txtRdd.collect.foreach(println)
bear bear car car bear
river river car
bear deer deer
car river river
bear deer deer
car car river deer
```

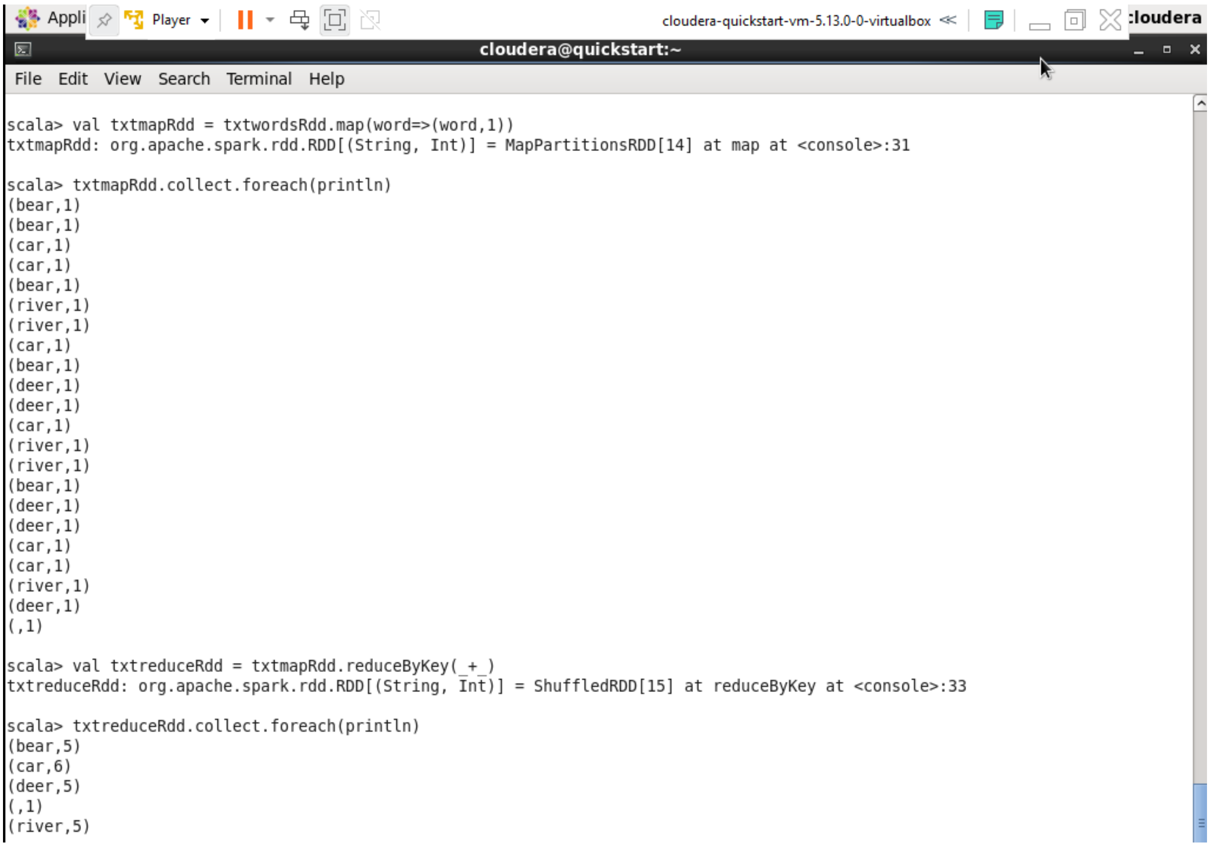


```
scala> val txtmapRdd = txtwordsRdd.map(word=>(word,1))
txtmapRdd: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[14] at map at <console>:31

scala> txtmapRdd.collect.foreach(println)
(bear,1)
(bear,1)
(car,1)
(car,1)
(bear,1)
(river,1)
(river,1)
(car,1)
(bear,1)
(deer,1)
(deer,1)
(car,1)
(river,1)
(river,1)
(bear,1)
(deer,1)
(deer,1)
(car,1)
(car,1)
(river,1)
(deer,1)
(,1)
```



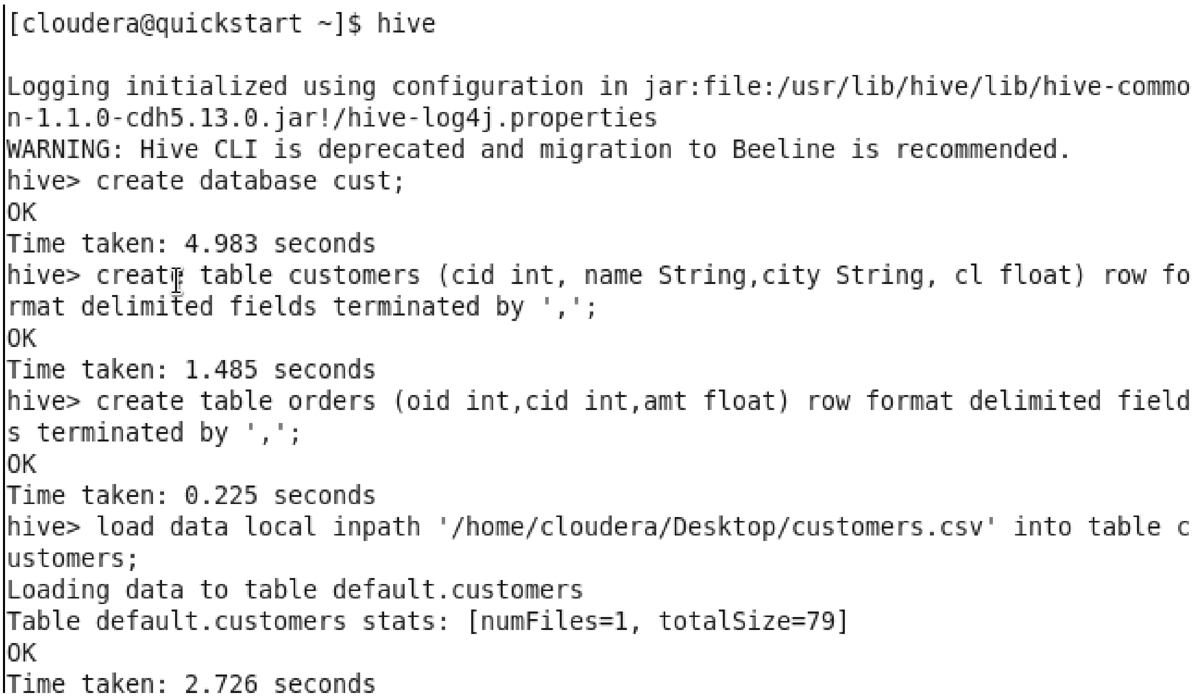
```
scala> val txtreduceRdd = txtmapRdd.reduceByKey(_+_)
txtreduceRdd: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[15] at reduceByKey at <console>:33

scala> txtreduceRdd.collect.foreach(println)
(bear,5)
(car,6)
(deer,5)
(,1)
(river,5)
```

```
scala> txtmapRdd.collect.foreach(println)
(bear,1)
(bear,1)
(car,1)
(car,1)
(bear,1)
(river,1)
(river,1)
(car,1)
(bear,1)
(deer,1)
(deer,1)
(car,1)
(river,1)
(river,1)
(bear,1)
(deer,1)
(deer,1)
(car,1)
(car,1)
(river,1)
(deer,1)
(,1)

scala> val txtreduceRdd = txtmapRdd.reduceByKey(_+_)
txtreduceRdd: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[15] at reduceByKey at <console>:33

scala> txtreduceRdd.collect.foreach(println)
(bear,5)
(car,6)
(deer,5)
(,1)
(river,5)
```

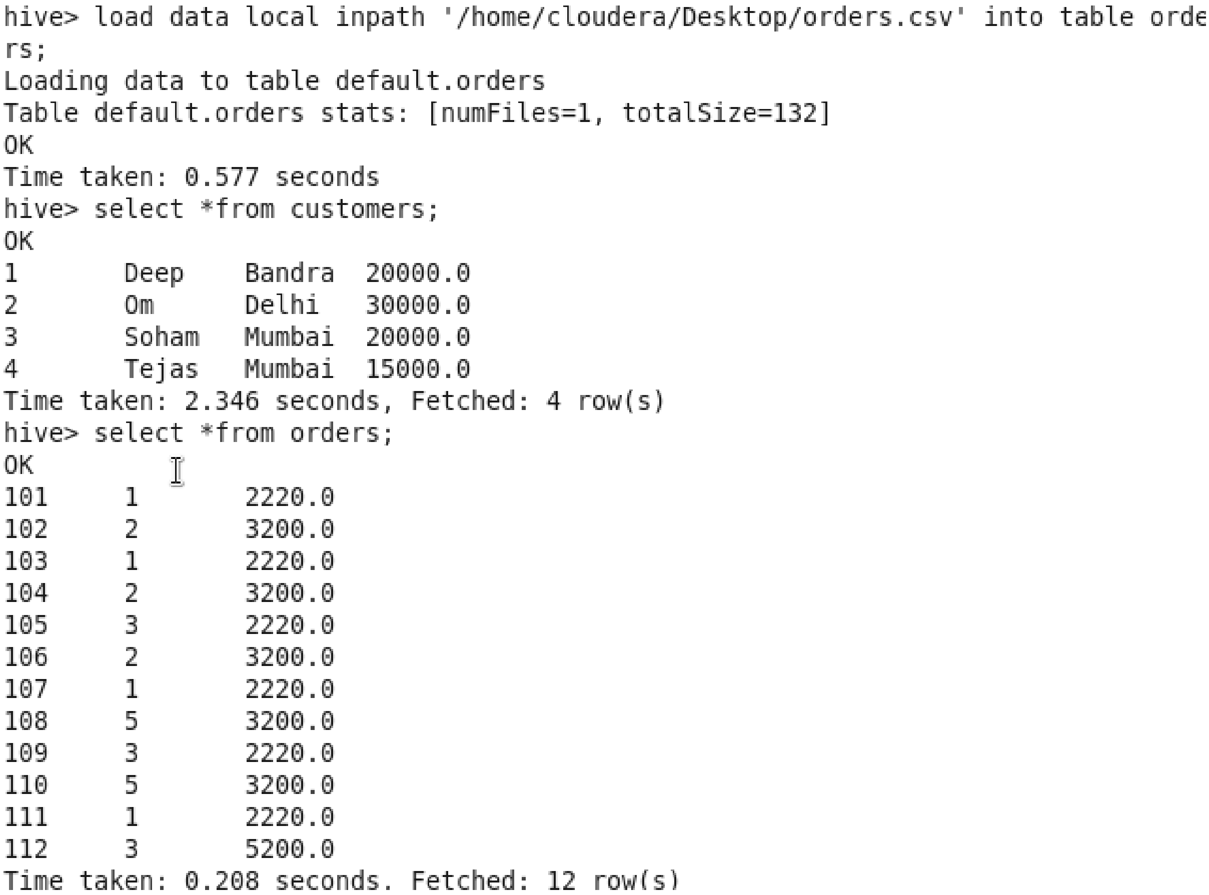


```
scala> val txtmapRdd = txtwordsRdd.map(word=>(word,1))
txtmapRdd: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[14] at map at <console>:31

scala> txtmapRdd.collect.foreach(println)
(bear,1)
(bear,1)
(car,1)
(car,1)
(bear,1)
(river,1)
(river,1)
(car,1)
(bear,1)
(deer,1)
(deer,1)
(car,1)
(river,1)
(river,1)
(bear,1)
(deer,1)
(deer,1)
(car,1)
(car,1)
(river,1)
(deer,1)
(,1)

scala> val txtreduceRdd = txtmapRdd.reduceByKey(_+_)
txtreduceRdd: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[15] at reduceByKey at <console>:33

scala> txtreduceRdd.collect.foreach(println)
(bear,5)
(car,6)
(deer,5)
(,1)
(river,5)
```

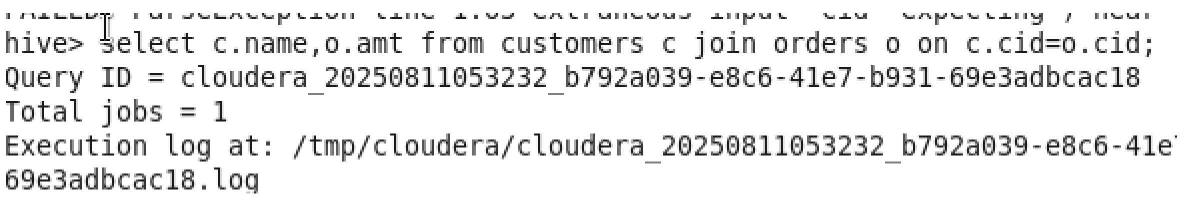
PIG Joins



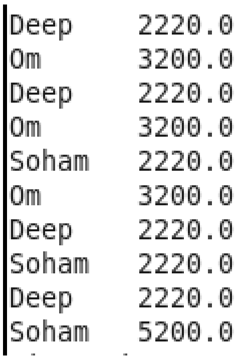
```
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-commo
n-1.1.0-cdh5.13.0.jar!/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create database cust;
OK
Time taken: 4.983 seconds
hive> create table customers (cid int, name String,city String, cl float) row fo
rmat delimited fields terminated by ',';
OK
Time taken: 1.485 seconds
hive> create table orders (oid int,cid int,amt float) row format delimited field
s terminated by ',';
OK
Time taken: 0.225 seconds
hive> load data local inpath '/home/cloudera/Desktop/customers.csv' into table c
ustomers;
Loading data to table default.customers
Table default.customers stats: [numFiles=1, totalSize=79]
OK
Time taken: 2.726 seconds
```



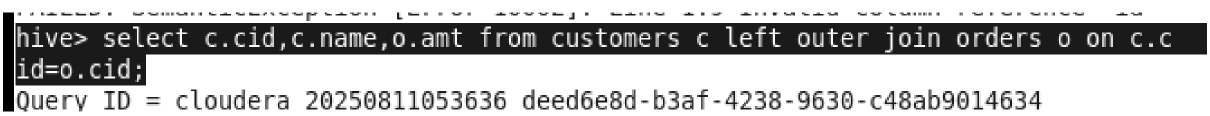
```
hive> load data local inpath '/home/cloudera/Desktop/orders.csv' into table orde
rs;
Loading data to table default.orders
Table default.orders stats: [numFiles=1, totalSize=132]
OK
Time taken: 0.577 seconds
hive> select *from customers;
OK
1       Deep    Bandra  20000.0
2       Om      Delhi   30000.0
3       Soham   Mumbai  20000.0
4       Tejas   Mumbai  15000.0
Time taken: 2.346 seconds, Fetched: 4 row(s)
hive> select *from orders;
OK
101     1       2220.0
102     2       3200.0
103     1       2220.0
104     2       3200.0
105     3       2220.0
106     2       3200.0
107     1       2220.0
108     5       3200.0
109     3       2220.0
110     5       3200.0
111     1       2220.0
112     3       5200.0
Time taken: 0.208 seconds. Fetched: 12 row(s)
```



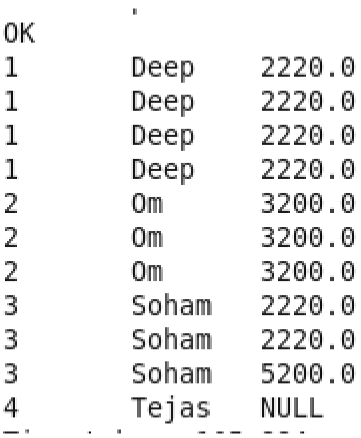
```
FAILED: ParseException line 1:65 extraneous input 'cid' expecting ; near
hive> select c.name,o.amt from customers c join orders o on c.cid=o.cid;
Query ID = cloudera_20250811053232_b792a039-e8c6-41e7-b931-69e3adbcac18
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20250811053232_b792a039-e8c6-41e
69e3adbcac18.log
```



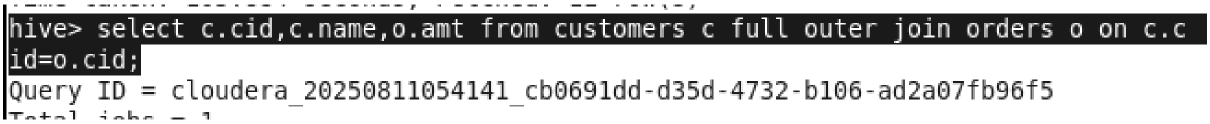
```
Deep    2220.0
Om      3200.0
Deep    2220.0
Om      3200.0
Soham   2220.0
Om      3200.0
Deep    2220.0
Soham   2220.0
Deep    2220.0
Soham   5200.0
```



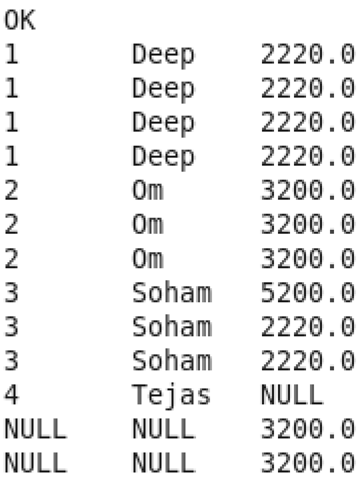
```
FAILED: SemanticException [Error 10002]: Line 1:3 Invalid column reference 'id
hive> select c.cid,c.name,o.amt from customers c left outer join orders o on c.c
id=o.cid;
Query ID = cloudera 20250811053636 deed6e8d-b3af-4238-9630-c48ab9014634
```



```
OK
1       Deep    2220.0
1       Deep    2220.0
1       Deep    2220.0
1       Deep    2220.0
2       Om      3200.0
2       Om      3200.0
2       Om      3200.0
3       Soham   2220.0
3       Soham   2220.0
3       Soham   5200.0
4       Tejas   NULL
```



```
Time taken: [illegible] seconds, Fetched: 11 row(s)
hive> select c.cid,c.name,o.amt from customers c full outer join orders o on c.c
id=o.cid;
Query ID = cloudera_20250811054141_cb0691dd-d35d-4732-b106-ad2a07fb96f5
Total jobs = 1
```



```
OK
1       Deep    2220.0
1       Deep    2220.0
1       Deep    2220.0
1       Deep    2220.0
2       Om      3200.0
2       Om      3200.0
2       Om      3200.0
3       Soham   5200.0
3       Soham   2220.0
3       Soham   2220.0
4       Tejas   NULL
NULL    NULL    3200.0
NULL    NULL    3200.0
```