```
[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> create database rjc_joins;
OK
Time taken: 1.284 seconds
hive>
```
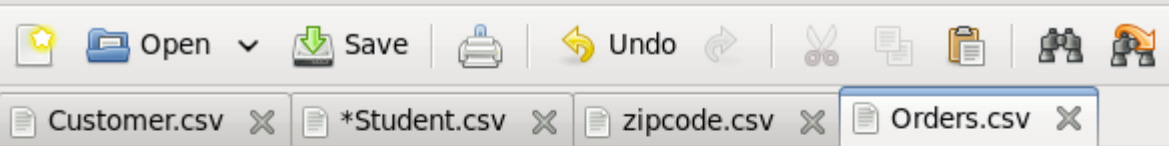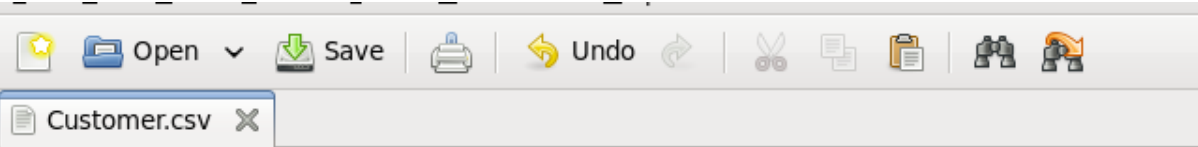
Customer.csv | *Student.csv | zipcode.csv | Orders.csv

```
OID,DATE,CUSTOMER,_ID,AMOUNT
102,2018-08-08,3,3000
100,2018-08-03,3,1500
101,2018-11-02,3,1560
103,2018-11-08,3,2060
```

Customer.csv

```
ID,NAME,AGE,ADDRESS,SALARY
1,REHAN,22,NEW YORK,20000
2,RISHI,21,FLORIDA,15000
3,SHIVAM,23,BOSTON,17000
4,ANAND,24,CALIFORNIA,25000
5,PRINCE,23,CHICAGO,21000
6,AMANDEEP,22,SEATTLE,14000
```

```
hive> show databases;
OK
default
rjc
rjc_joins
rjcstudent
Time taken: 0.673 seconds, Fetched: 4 row(s)
hive>
```

```
hive> use rjc_joins;
OK
Time taken: 0.155 seconds
hive>
```

```
hive> create table customers(ID int, Name string,Age int,Address string,Salary float)
    > row format delimited
    > fields terminated by ','
    > tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.611 seconds
hive> describe customers;
OK
id                      int
name                    string
age                     int
address                 string
salary                  float
Time taken: 0.179 seconds, Fetched: 5 row(s)
hive>
```

```
hive> load data local inpath '/home/cloudera/Documents/Customer.csv' into table customers;
Loading data to table rjc_joins.customers
Table rjc_joins.customers stats: [numFiles=1, totalSize=185]
OK
Time taken: 1.317 seconds
hive>
```

```
hive> select * from customers;
OK
1        REHAN     22      NEW YORK          20000.0
2        RISHI     21      FLORIDA 15000.0
3        SHIVAM    23      BOSTON  17000.0
4        ANAND     24      CALIFORNIA        25000.0
5        PRINCE    23      CHICAGO 21000.0
6        AMANDEEP          22       SEATTLE 14000.0
Time taken: 0.863 seconds, Fetched: 6 row(s)
hive>
```

```
hive> create table orders(oid int,odate date,cid int,amount float)
    > row format delimited
    > fields terminated by ','
    > tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.116 seconds
hive>
```

```
hive> describe orders;
OK
oid                     int
odate                   date
cid                     int
amount                  float
Time taken: 0.152 seconds, Fetched: 4 row(s)
hive>
```

```
hive> load data local inpath '/home/cloudera/Documents/Orders.csv' into table orders;
Loading data to table rjc_joins.orders
Table rjc_joins.orders stats: [numFiles=1, totalSize=117]
OK
Time taken: 0.592 seconds
hive>
```

```
hive> select c.id, c.name, c.age, o.amount
    > from customers c JOIN orders o
    > on (c.id=o.cid);
Query ID = cloudera_20220323005656_7be2d8f0-922f-462a-a992-bc3f36617103
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220323005656_7be2d8f0-922f-462a-a992-bc3f36617103.log
2022-03-23 12:56:29    Starting to launch local task to process map join;     maximum memory = 1013645312
2022-03-23 12:56:34    Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_00-56-17_161_3819264026608354720-1/-local-10003/HashTable-Stage-3/MapJo
le11--.hashtable
2022-03-23 12:56:34    Uploaded 1 File to: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_00-56-17_161_3819264026608354720-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile11--.hashtable (338 bytes)
2022-03-23 12:56:34    End of local task; Time Taken: 4.569 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_1644548343526_0013, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0013/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0013
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-03-23 00:56:54,499 Stage-3 map = 0%,  reduce = 0%
2022-03-23 00:57:10,128 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.33 sec
MapReduce Total cumulative CPU time: 1 seconds 330 msec
Ended Job = job_1644548343526_0013
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.33 sec   HDFS Read: 6698 HDFS Write: 74 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 330 msec
OK
2       RISHI   21      1560.0
3       SHIVAM  23      3000.0
3       SHIVAM  23      1500.0
4       ANAND   24      2060.0
Time taken: 55.208 seconds, Fetched: 4 row(s)
```

```
hive> select c.id, c.name, o.amount, o.odate
    > from customers c LEFT OUTER JOIN orders o
    > on (c.id = o.cid);
Query ID = cloudera_20220323010000_8c0d62fb-060b-45c6-a89c-7196ee19a3ed
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220323010000_8c0d62fb-060b-45c6-a89c-7196ee19a3ed.log
2022-03-23 01:00:53    Starting to launch local task to process map join;    maximum memory = 1013645312
2022-03-23 01:00:57    Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_01-00-42_679_2735558448338020109-1/-local-10003/HashTable-Stage-3/MapJoin-mapfi
le21--.hashtable
2022-03-23 01:00:57    Uploaded 1 File to: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_01-00-42_679_2735558448338020109-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile21--.hashtable (350 bytes)
2022-03-23 01:00:57    End of local task; Time Taken: 3.306 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_1644548343526_0014, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0014/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0014
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-03-23 01:01:16,491 Stage-3 map = 0%,  reduce = 0%
2022-03-23 01:01:32,265 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.05 sec
MapReduce Total cumulative CPU time: 1 seconds 50 msec
Ended Job = job_1644548343526_0014
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.05 sec   HDFS Read: 6546 HDFS Write: 152 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 50 msec
OK
1       REHAN   NULL    NULL
2       RISHI   1560.0  2018-11-02
3       SHIVAM  3000.0  2018-08-08
3       SHIVAM  1500.0  2018-08-03
4       ANAND   2060.0  2018-11-08
5       PRINCE  NULL    NULL
6       AMANDEEP        NULL    NULL
Time taken: 50.742 seconds, Fetched: 7 row(s)
```

```
hive> select c.id, c.name, o.amount, o.odate
    > from customers c RIGHT OUTER JOIN orders o
    > on (c.id = o.cid);
Query ID = cloudera_20220323010505_7a1cc64f-411b-4a3d-9193-d10069b40079
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220323010505_7a1cc64f-411b-4a3d-9193-d10069b40079.log
2022-03-23 01:05:41    Starting to launch local task to process map join;    maximum memory = 1013645312
2022-03-23 01:05:45    Dump the side-table for tag: 0 with group count: 6 into file: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_01-05-28_732_1519338213226978134-1/-local-10003/HashTable-Stage-3/MapJoin-mapfil
e30--.hashtable
2022-03-23 01:05:45    Uploaded 1 File to: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_01-05-28_732_1519338213226978134-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile30--.hashtable (415 bytes)
2022-03-23 01:05:45    End of local task; Time Taken: 3.67 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_1644548343526_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0015
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-03-23 01:06:06,244 Stage-3 map = 0%,  reduce = 0%
2022-03-23 01:06:23,034 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.08 sec
MapReduce Total cumulative CPU time: 1 seconds 80 msec
Ended Job = job_1644548343526_0015
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 1.08 sec   HDFS Read: 6439 HDFS Write: 106 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 80 msec
OK
3       SHIVAM  3000.0  2018-08-08
3       SHIVAM  1500.0  2018-08-03
2       RISHI   1560.0  2018-11-02
4       ANAND   2060.0  2018-11-08
Time taken: 56.034 seconds, Fetched: 4 row(s)
```

```
hive> select c.id, c.name, c.age, o.amount
    > from customers c JOIN orders o
    > on (c.id=o.cid);
Query ID = cloudera_20220323005454_ec8ff044-5965-4c4b-8329-29b2f23dba3e
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20220323005454_ec8ff044-5965-4c4b-8329-29b2f23dba3e.log
2022-03-23 12:54:15    Starting to launch local task to process map join;    maximum memory = 1013645312
2022-03-23 12:54:19    Dump the side-table for tag: 1 with group count: 1 into file: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_00-54-02_203_3306767925110645266-1/-local-10003/HashTable-Stage-3/MapJoin-mapfi
le01--.hashtable
2022-03-23 12:54:19    Uploaded 1 File to: file:/tmp/cloudera/58b13345-b761-4a2a-8df8-b7d448ce5b08/hive_2022-03-23_00-54-02_203_3306767925110645266-1/-local-10003/HashTable-Stage-3/MapJoin-mapfile01--.hashtable (310 bytes)
2022-03-23 12:54:19    End of local task; Time Taken: 4.543 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_1644548343526_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0012
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2022-03-23 00:54:44,089 Stage-3 map = 0%,  reduce = 0%
2022-03-23 00:55:01,616 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 1.38 sec
MapReduce Total cumulative CPU time: 1 seconds 380 msec
```

```
Launching Job 7 out of 7
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_1644548343526_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0020
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-03-23 01:13:11,199 Stage-3 map = 0%,  reduce = 0%
2022-03-23 01:13:27,073 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 0.86 sec
2022-03-23 01:13:45,721 Stage-3 map = 100%,  reduce = 100%, Cumulative CPU 2.03 sec
MapReduce Total cumulative CPU time: 2 seconds 30 msec
Ended Job = job_1644548343526_0020
MapReduce Jobs Launched:
Stage-Stage-4: Map: 1  Reduce: 1   Cumulative CPU: 2.07 sec   HDFS Read: 6422 HDFS Write: 117 SUCCESS
Stage-Stage-5: Map: 1  Reduce: 1   Cumulative CPU: 2.52 sec   HDFS Read: 7522 HDFS Write: 114 SUCCESS
Stage-Stage-8: Map: 1   Cumulative CPU: 1.06 sec   HDFS Read: 5443 HDFS Write: 222 SUCCESS
Stage-Stage-6: Map: 1   Cumulative CPU: 1.41 sec   HDFS Read: 5284 HDFS Write: 117 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 2.03 sec   HDFS Read: 4649 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 9 seconds 90 msec
OK
21000.0
Time taken: 278.646 seconds, Fetched: 1 row(s)
hive>
```

```
hive> Select salary from (select salary from customers sort by salary desc limit 4) result sort by salary asc limit 1;
Query ID = cloudera_20220323011818_8599de0c-40b4-4127-bcb8-4cae452e503b
Total jobs = 4
Launching Job 1 out of 4
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_1644548343526_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0021/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0021
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-03-23 01:18:20,398 Stage-1 map = 0%,   reduce = 0%
2022-03-23 01:18:37,497 Stage-1 map = 100%,   reduce = 0%, Cumulative CPU 0.92 sec
2022-03-23 01:18:55,016 Stage-1 map = 100%,   reduce = 100%, Cumulative CPU 2.05 sec
MapReduce Total cumulative CPU time: 2 seconds 50 msec
Ended Job = job_1644548343526_0021
Launching Job 2 out of 4
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_1644548343526_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0022
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-03-23 01:19:14,939 Stage-2 map = 0%,   reduce = 0%
2022-03-23 01:19:29,297 Stage-2 map = 100%,   reduce = 0%, Cumulative CPU 0.89 sec
2022-03-23 01:19:47,879 Stage-2 map = 100%,   reduce = 100%, Cumulative CPU 1.99 sec
MapReduce Total cumulative CPU time: 1 seconds 990 msec
Ended Job = job_1644548343526_0022
```

```
  set mapreduce.job.reduces=<number>
Starting Job = job_1644548343526_0022, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0022/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0022
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2022-03-23 01:19:14,939 Stage-2 map = 0%,   reduce = 0%
2022-03-23 01:19:29,297 Stage-2 map = 100%,   reduce = 0%, Cumulative CPU 0.89 sec
2022-03-23 01:19:47,879 Stage-2 map = 100%,   reduce = 100%, Cumulative CPU 1.99 sec
MapReduce Total cumulative CPU time: 1 seconds 990 msec
Ended Job = job_1644548343526_0022
Launching Job 3 out of 4
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_1644548343526_0023, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0023/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0023
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2022-03-23 01:20:07,119 Stage-3 map = 0%,   reduce = 0%
2022-03-23 01:20:21,448 Stage-3 map = 100%,   reduce = 0%, Cumulative CPU 0.76 sec
2022-03-23 01:20:38,968 Stage-3 map = 100%,   reduce = 100%, Cumulative CPU 1.86 sec
MapReduce Total cumulative CPU time: 1 seconds 860 msec
Ended Job = job_1644548343526_0023
Launching Job 4 out of 4
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_1644548343526_0024, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1644548343526_0024/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1644548343526_0024
Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 1
2022-03-23 01:20:59,449 Stage-4 map = 0%,   reduce = 0%
2022-03-23 01:21:13,923 Stage-4 map = 100%,   reduce = 0%, Cumulative CPU 0.81 sec
2022-03-23 01:21:32,381 Stage-4 map = 100%,   reduce = 100%, Cumulative CPU 1.92 sec
MapReduce Total cumulative CPU time: 1 seconds 920 msec
Ended Job = job_1644548343526_0024
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.05 sec   HDFS Read: 5446 HDFS Write: 180 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.99 sec   HDFS Read: 3837 HDFS Write: 180 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 1.86 sec   HDFS Read: 3849 HDFS Write: 117 SUCCESS
Stage-Stage-4: Map: 1  Reduce: 1   Cumulative CPU: 1.92 sec   HDFS Read: 4330 HDFS Write: 8 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 820 msec
OK
17000.0
Time taken: 212.098 seconds, Fetched: 1 row(s)
hive>
```