**1. customer\_info.csv**

1,John,101

2,Alice,102

3,Bob,103

**2. order\_info.csv**

101,501,2

102,502,1

103,503,5

**3. item\_info.csv**

501,Laptop,60000

502,Mouse,500

503,Keyboard,1500  
  
**Save this in /home/cloudera**

**[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>**

**> -- Customer Information Table**

**> CREATE TABLE Customer\_info (**

**> Cust\_ID INT,**

**> Cust\_Name STRING,**

**> OrderID INT**

**> )**

**> ROW FORMAT DELIMITED**

**> FIELDS TERMINATED BY ','**

**> STORED AS TEXTFILE;**

OK

Time taken: 0.728 seconds

**hive>**

**> -- Order Info Table**

**> CREATE TABLE Order\_info (**

**> OrderID INT,**

**> ItemID INT,**

**> Quantity INT**

**> )**

**> ROW FORMAT DELIMITED**

**> FIELDS TERMINATED BY ','**

**> STORED AS TEXTFILE;**

OK

Time taken: 0.093 seconds

**hive>**

**> -- Item Info Table**

**> CREATE TABLE Item\_info (**

**> Item\_ID INT,**

**> Item\_Name STRING,**

**> ItemPrice FLOAT**

**> )**

**> ROW FORMAT DELIMITED**

**> FIELDS TERMINATED BY ','**

**> STORED AS TEXTFILE;**

OK

Time taken: 0.252 seconds

**hive> LOAD DATA LOCAL INPATH '/home/cloudera/customer\_info.csv' INTO TABLE Customer\_info;**

Loading data to table default.customer\_info

Table default.customer\_info stats: [numFiles=1, totalSize=34]

OK

Time taken: 0.736 seconds

**hive> LOAD DATA LOCAL INPATH '/home/cloudera/order\_info.csv' INTO TABLE Order\_info;**

Loading data to table default.order\_info

Table default.order\_info stats: [numFiles=1, totalSize=31]

OK

Time taken: 0.417 seconds

**hive> LOAD DATA LOCAL INPATH '/home/cloudera/item\_info.csv' INTO TABLE Item\_info;**

Loading data to table default.item\_info

Table default.item\_info stats: [numFiles=1, totalSize=50]

OK

Time taken: 0.638 seconds

**hive> SELECT \* FROM Customer\_info;**

OK

1 John 101

2 Alice 102

3 Bob 103

NULL NULL NULL

Time taken: 0.441 seconds, Fetched: 4 row(s)

**hive> SELECT \* FROM Order\_info;**

OK

101 501 2

102 502 1

103 503 5

NULL NULL NULL

Time taken: 0.064 seconds, Fetched: 4 row(s)

**hive> SELECT \* FROM Item\_info;**

OK

501 Laptop 60000.0

502 Mouse 500.0

503 Keyboard 1500.0

NULL NULL NULL

Time taken: 0.074 seconds, Fetched: 4 row(s)

**hive> SELECT**

**> c.Cust\_ID,**

**> c.Cust\_Name,**

**> o.OrderID,**

**> i.Item\_Name,**

**> o.Quantity,**

**> i.ItemPrice,**

**> o.Quantity \* i.ItemPrice AS TotalCost**

**> FROM**

**> Customer\_info c**

**> JOIN**

**> Order\_info o ON c.OrderID = o.OrderID**

**> JOIN**

**> Item\_info i ON o.ItemID = i.Item\_ID;**

Query ID = cloudera\_20250507073939\_4639ce79-2d14-4772-ba51-6733d20f3593

Total jobs = 1

Execution log at: /tmp/cloudera/cloudera\_20250507073939\_4639ce79-2d14-4772-ba51-6733d20f3593.log

2025-05-07 07:39:34 Starting to launch local task to process map join; maximum memory = 932184064

2025-05-07 07:39:36 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-39-28\_145\_8930674938168841766-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile01--.hashtable

2025-05-07 07:39:36 Uploaded 1 File to: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-39-28\_145\_8930674938168841766-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile01--.hashtable (357 bytes)

2025-05-07 07:39:36 Dump the side-table for tag: 1 with group count: 3 into file: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-39-28\_145\_8930674938168841766-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile11--.hashtable

2025-05-07 07:39:36 Uploaded 1 File to: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-39-28\_145\_8930674938168841766-1/-local-10005/HashTable-Stage-5/MapJoin-mapfile11--.hashtable (329 bytes)

2025-05-07 07:39:36 End of local task; Time Taken: 1.791 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\_1744085536392\_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1744085536392\_0010/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1744085536392\_0010

Hadoop job information for Stage-5: number of mappers: 1; number of reducers: 0

2025-05-07 07:39:48,563 Stage-5 map = 0%, reduce = 0%

2025-05-07 07:39:59,618 Stage-5 map = 100%, reduce = 0%, Cumulative CPU 3.02 sec

MapReduce Total cumulative CPU time: 3 seconds 20 msec

Ended Job = job\_1744085536392\_0010

MapReduce Jobs Launched:

Stage-Stage-5: Map: 1 Cumulative CPU: 3.02 sec HDFS Read: 8926 HDFS Write: 104 SUCCESS

Total MapReduce CPU Time Spent: 3 seconds 20 msec

OK

1 John 101 Laptop 2 60000.0 120000.0

2 Alice 102 Mouse 1 500.0 500.0

3 Bob 103 Keyboard 5 1500.0 7500.0

Time taken: 34.896 seconds, Fetched: 3 row(s)

**hive> CREATE INDEX idx\_cust\_id**

**> ON TABLE Customer\_info (Cust\_ID)**

**> AS 'COMPACT'**

**> WITH DEFERRED REBUILD;**

OK

Time taken: 0.73 seconds

**hive> SELECT**

**> SUM(o.Quantity \* i.ItemPrice) AS TotalSales,**

**> AVG(o.Quantity \* i.ItemPrice) AS AverageSales**

**> FROM**

**> Order\_info o**

**> JOIN**

**> Item\_info i ON o.ItemID = i.Item\_ID;**

Query ID = cloudera\_20250507074141\_f0342dcd-43bf-4977-a420-b0676dc697bf

Total jobs = 1

Execution log at: /tmp/cloudera/cloudera\_20250507074141\_f0342dcd-43bf-4977-a420-b0676dc697bf.log

2025-05-07 07:41:16 Starting to launch local task to process map join; maximum memory = 932184064

2025-05-07 07:41:18 Dump the side-table for tag: 0 with group count: 3 into file: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-41-11\_111\_4624799052663818332-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile20--.hashtable

2025-05-07 07:41:18 Uploaded 1 File to: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-41-11\_111\_4624799052663818332-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile20--.hashtable (326 bytes)

2025-05-07 07:41:18 End of local task; Time Taken: 2.28 sec.

Execution completed successfully

MapredLocal task succeeded

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\_1744085536392\_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1744085536392\_0011/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1744085536392\_0011

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2025-05-07 07:41:28,440 Stage-2 map = 0%, reduce = 0%

2025-05-07 07:41:37,108 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.54 sec

2025-05-07 07:41:45,544 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.06 sec

MapReduce Total cumulative CPU time: 4 seconds 60 msec

Ended Job = job\_1744085536392\_0011

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.06 sec HDFS Read: 13782 HDFS Write: 28 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 60 msec

OK

128000.0 42666.666666666664

Time taken: 37.219 seconds, Fetched: 1 row(s)

**hive> SELECT**

**> o.OrderID,**

**> i.Item\_Name,**

**> o.Quantity,**

**> i.ItemPrice,**

**> (o.Quantity \* i.ItemPrice) AS TotalCost**

**> FROM**

**> Order\_info o**

**> JOIN**

**> Item\_info i ON o.ItemID = i.Item\_ID**

**> ORDER BY**

**> TotalCost DESC**

**> LIMIT 1;**

Query ID = cloudera\_20250507074242\_8f0779ab-4bae-45f2-9023-0268147739c0

Total jobs = 1

Execution log at: /tmp/cloudera/cloudera\_20250507074242\_8f0779ab-4bae-45f2-9023-0268147739c0.log

2025-05-07 07:42:54 Starting to launch local task to process map join; maximum memory = 932184064

2025-05-07 07:42:55 Dump the side-table for tag: 0 with group count: 3 into file: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-42-50\_416\_6735928633199794294-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile30--.hashtable

2025-05-07 07:42:55 Uploaded 1 File to: file:/tmp/cloudera/f6334910-0db8-4609-8414-7b60ea21ec06/hive\_2025-05-07\_07-42-50\_416\_6735928633199794294-1/-local-10004/HashTable-Stage-2/MapJoin-mapfile30--.hashtable (329 bytes)

2025-05-07 07:42:55 End of local task; Time Taken: 1.112 sec.

Execution completed successfully

MapredLocal task succeeded

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\_1744085536392\_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1744085536392\_0012/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1744085536392\_0012

Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1

2025-05-07 07:43:05,058 Stage-2 map = 0%, reduce = 0%

2025-05-07 07:43:12,766 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.9 sec

2025-05-07 07:43:23,219 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.2 sec

MapReduce Total cumulative CPU time: 4 seconds 200 msec

Ended Job = job\_1744085536392\_0012

MapReduce Jobs Launched:

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.2 sec HDFS Read: 13406 HDFS Write: 30 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 200 msec

OK

101 Laptop 2 60000.0 120000.0

Time taken: 34.954 seconds, Fetched: 1 row(s)

**hive> CREATE EXTERNAL TABLE hbase\_customer\_info (**

**> rowkey STRING,**

**> cust\_id STRING,**

**> cust\_name STRING,**

**> order\_id STRING**

**> )**

**> STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'**

**> WITH SERDEPROPERTIES (**

**> "hbase.columns.mapping" = ":key,cf1:cust\_id,cf1:cust\_name,cf1:order\_id"**

**> )**

**> TBLPROPERTIES (**

**> "hbase.table.name" = "hbase\_customer\_info"**

**> );**

OK

Time taken: 0.828 seconds

**hive> SELECT \* FROM hbase\_customer\_info;**

OK

row1 1 John 101

row2 2 Alice 102

Time taken: 0.143 seconds, Fetched: 2 row(s)

**#HBASE COMMANDS BEFORE EXTERNAL TABLE CREATION**

**[cloudera@quickstart Desktop]$ hbase shell**

2025-05-07 07:45:12,230 INFO [main] Configuration.deprecation: hadoop.native.lib is deprecated. Instead, use io.native.lib.available

HBase Shell; enter 'help<RETURN>' for list of supported commands.

Type "exit<RETURN>" to leave the HBase Shell

Version 1.2.0-cdh5.13.0, rUnknown, Wed Oct 4 11:16:18 PDT 2017

hbase(main):001:0>

**hbase(main):002:0\* create 'hbase\_customer\_info', 'cf1'**

0 row(s) in 2.7110 seconds

=> Hbase::Table - hbase\_customer\_info

hbase(main):003:0>

**hbase(main):004:0\* put 'hbase\_customer\_info', 'row1', 'cf1:cust\_id', '1'**

0 row(s) in 0.1810 seconds

**hbase(main):005:0> put 'hbase\_customer\_info', 'row1', 'cf1:cust\_name', 'John'**

0 row(s) in 0.0130 seconds

**hbase(main):006:0> put 'hbase\_customer\_info', 'row1', 'cf1:order\_id', '101'**

0 row(s) in 0.0090 seconds

hbase(main):007:0>

**hbase(main):008:0\* put 'hbase\_customer\_info', 'row2', 'cf1:cust\_id', '2'**

0 row(s) in 0.0080 seconds

**hbase(main):009:0> put 'hbase\_customer\_info', 'row2', 'cf1:cust\_name', 'Alice'**

0 row(s) in 0.0060 seconds

**hbase(main):010:0> put 'hbase\_customer\_info', 'row2', 'cf1:order\_id', '102'**

0 row(s) in 0.0060 seconds