online_retail_simple.csv

InvoiceNo,CustomerID,Quantity,UnitPrice,Country

536365,17850,6,2.55,United Kingdom

536366,17851,3,3.39,France

536367,17852,1,5.00,Germany

[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> CREATE TABLE online_retail_simple (
    >     InvoiceNo STRING,
    >     CustomerID STRING,
    >     Quantity INT,
    >     UnitPrice FLOAT,
    >     Country STRING
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE
    > TBLPROPERTIES ("skip.header.line.count"="1");
```

OK

Time taken: 0.736 seconds

```
hive> LOAD DATA LOCAL INPATH '/home/cloudera/online_retail_simple.csv'
    > INTO TABLE online_retail_simple;
```

Loading data to table default.online_retail_simple

Table default.online_retail_simple stats: [numFiles=1, totalSize=139]

OK

Time taken: 0.682 seconds

```
hive> SELECT
```

> SUM(Quantity * UnitPrice) AS TotalSales,

> AVG(Quantity * UnitPrice) AS AvgSales

> FROM

> online_retail_simple;

Query ID = cloudera_20250507080606_b0e4fb03-9cee-4bc9-bf71-c44d0796412e

Total jobs = 1

Launching Job 1 out of 1

Number of reduce tasks determined at compile time: 1

In order to change the average load for a reducer (in bytes):

set hive.exec.reducers.bytes.per.reducer=<number>

In order to limit the maximum number of reducers:

set hive.exec.reducers.max=<number>

In order to set a constant number of reducers:

set mapreduce.job.reduces=<number>

Starting Job = job_1744085536392_0013, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0013/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0013

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

2025-05-07 08:06:41,113 Stage-1 map = 0%, reduce = 0%

2025-05-07 08:06:52,398 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.79 sec

2025-05-07 08:07:01,905 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.23 sec

MapReduce Total cumulative CPU time: 4 seconds 230 msec

Ended Job = job_1744085536392_0013

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.23 sec HDFS Read: 10762 HDFS Write: 37 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 230 msec

OK

30.469999313354492 10.15666643778483

Time taken: 34.117 seconds, Fetched: 1 row(s)

hive> SELECT

> InvoiceNo,

> SUM(Quantity * UnitPrice) AS OrderTotal

> FROM

> online_retail_simple

> GROUP BY

> InvoiceNo

> ORDER BY

> OrderTotal DESC

> LIMIT 1;

Query ID = cloudera_20250507080707_e4951fb8-24b2-410c-884b-3c4e7cefa5f2

Total jobs = 2

Launching Job 1 out of 2

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_1744085536392_0014, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0014/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0014

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

2025-05-07 08:07:45,644 Stage-1 map = 0%, reduce = 0%

2025-05-07 08:07:53,525 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.11 sec

2025-05-07 08:07:59,842 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.29 sec

MapReduce Total cumulative CPU time: 3 seconds 290 msec

Ended Job = job_1744085536392_0014

Launching Job 2 out of 2

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_0015, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0015/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0015

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

2025-05-07 08:08:11,770 Stage-2 map = 0%, reduce = 0%

2025-05-07 08:08:18,502 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.38 sec

2025-05-07 08:08:26,001 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.56 sec

MapReduce Total cumulative CPU time: 2 seconds 560 msec

Ended Job = job_1744085536392_0015

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.29 sec HDFS Read: 9107 HDFS Write: 210 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.56 sec HDFS Read: 5357 HDFS Write: 26 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 850 msec

OK

536365 15.299999237060547

Time taken: 50.303 seconds, Fetched: 1 row(s)

hive> SELECT

> CustomerID,

> SUM(Quantity * UnitPrice) AS TotalSpent

> FROM

> online_retail_simple

> GROUP BY

> CustomerID

> ORDER BY

> TotalSpent DESC

> LIMIT 1;

Query ID = cloudera_20250507080909_dd5e1e6d-b64e-4ae0-8af9-aba6e5fc17e3

Total jobs = 2

Launching Job 1 out of 2

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_1744085536392_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0016/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0016

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

2025-05-07 08:09:54,399 Stage-1 map = 0%, reduce = 0%

2025-05-07 08:10:00,935 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.02 sec

2025-05-07 08:10:09,311 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.52 sec

MapReduce Total cumulative CPU time: 3 seconds 520 msec

Ended Job = job_1744085536392_0016

Launching Job 2 out of 2

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_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0017/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0017

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

2025-05-07 08:10:19,659 Stage-2 map = 0%, reduce = 0%

2025-05-07 08:10:26,264 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.8 sec

2025-05-07 08:10:33,576 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.62 sec

MapReduce Total cumulative CPU time: 3 seconds 620 msec

Ended Job = job_1744085536392_0017

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 3.52 sec   HDFS Read: 9108 HDFS Write: 206 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1   Cumulative CPU: 3.62 sec   HDFS Read: 5348 HDFS Write: 25 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 140 msec

OK

17850   15.299999237060547

Time taken: 49.519 seconds, Fetched: 1 row(s)

hive> -- Max

**   > SELECT Country, SUM(Quantity * UnitPrice) AS TotalSales**

**   > FROM online_retail_simple**

**   > GROUP BY Country**

**   > ORDER BY TotalSales DESC**

**   > LIMIT 1;**

Query ID = cloudera_20250507081313_0329c043-4b60-4ae9-a984-46434307daff

Total jobs = 2

Launching Job 1 out of 2

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_1744085536392_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0018/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0018

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

2025-05-07 08:13:28,472 Stage-1 map = 0%, reduce = 0%

2025-05-07 08:13:36,149 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.69 sec

2025-05-07 08:13:44,566 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.06 sec

MapReduce Total cumulative CPU time: 4 seconds 60 msec

Ended Job = job_1744085536392_0018

Launching Job 2 out of 2

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_0019, Tracking URL =
http://quickstart.cloudera:8088/proxy/application_1744085536392_0019/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0019

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

2025-05-07 08:13:53,571 Stage-2 map = 0%, reduce = 0%

2025-05-07 08:14:01,749 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.86 sec

2025-05-07 08:14:10,085 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.18 sec

MapReduce Total cumulative CPU time: 3 seconds 180 msec

Ended Job = job_1744085536392_0019

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1   Cumulative CPU: 4.06 sec   HDFS Read: 9103 HDFS Write: 218
SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1   Cumulative CPU: 3.18 sec   HDFS Read: 5357 HDFS Write: 34
SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 240 msec

OK

United Kingdom		15.299999237060547

Time taken: 53.98 seconds, Fetched: 1 row(s)

hive>

> -- Min (excluding zero)

> SELECT Country, SUM(Quantity * UnitPrice) AS TotalSales

> FROM online_retail_simple

> GROUP BY Country

> HAVING TotalSales > 0

> ORDER BY TotalSales ASC

> LIMIT 1;

Query ID = cloudera_20250507081414_b217e351-4866-4f8c-8bec-e0a221765db2

Total jobs = 2

Launching Job 1 out of 2

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_1744085536392_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0020/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0020

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

2025-05-07 08:14:22,498 Stage-1 map = 0%, reduce = 0%

2025-05-07 08:14:31,296 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.51 sec

2025-05-07 08:14:42,561 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.01 sec

MapReduce Total cumulative CPU time: 4 seconds 10 msec

Ended Job = job_1744085536392_0020

Launching Job 2 out of 2

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_0021, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1744085536392_0021/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1744085536392_0021

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

2025-05-07 08:14:54,626 Stage-2 map = 0%, reduce = 0%

2025-05-07 08:15:01,847 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.21 sec

2025-05-07 08:15:12,341 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.7 sec

MapReduce Total cumulative CPU time: 3 seconds 700 msec

Ended Job = job_1744085536392_0021

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.01 sec HDFS Read: 9453 HDFS Write: 201 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 3.7 sec HDFS Read: 5341 HDFS Write: 12 SUCCESS

Total MapReduce CPU Time Spent: 7 seconds 710 msec

OK

Germany 5.0

Time taken: 61.18 seconds, Fetched: 1 row(s)

HBASE Table creation

[cloudera@quickstart Desktop]$ hbase shell

2025-05-07 08:17:23,260 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> create 'online_retail_hbase', 'cf1'

0 row(s) in 2.7330 seconds

=> Hbase::Table - online_retail_hbase

HIVE EXTERNAL TABLE

```
hive> CREATE EXTERNAL TABLE hbase_online_retail_simple (
    >    rowkey STRING,
    >    InvoiceNo STRING,
    >    CustomerID STRING,
    >    Country STRING
    > )
    > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
    > WITH SERDEPROPERTIES (
    >    "hbase.columns.mapping" = ":key,cf1:InvoiceNo,cf1:CustomerID,cf1:Country"
    > )
    > TBLPROPERTIES (
    >    "hbase.table.name" = "online_retail_hbase"
    > );
```

OK

Time taken: 0.948 seconds

```
hive> SELECT * FROM hbase_online_retail_simple;
```

OK

Time taken: 0.224 seconds