[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> show tables;

OK

customer\_info

default\_\_flight\_info\_hive\_flight\_delay\_index\_\_

employee\_hbase

flight\_info\_hive

flighthive

item\_info

online\_retail

online\_retail\_hbase

order\_info

Time taken: 1.013 seconds, Fetched: 9 row(s)

hive> create database online;

OK

Time taken: 2.695 seconds

hive> CREATE TABLE online\_retail (

> InvoiceNo STRING,

> StockCode STRING,

> Description STRING,

> Quantity INT,

> InvoiceDate STRING,

> UnitPrice DOUBLE,

> CustomerID STRING,

> Country STRING

> )

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY ','

> STORED AS TEXTFILE;

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table online\_retail already exists)

hive> CREATE TABLE online\_retail (

> InvoiceNo STRING,

> StockCode STRING,

> Description STRING,

> Quantity INT,

> InvoiceDate STRING,

> UnitPrice DOUBLE,

> CustomerID STRING,

> Country STRING

> )

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY ','

> STORED AS TEXTFILE;

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table online\_retail already exists)

hive> CREATE TABLE retail (

> InvoiceNo STRING,

> StockCode STRING,

> Description STRING,

> Quantity INT,

> InvoiceDate STRING,

> UnitPrice DOUBLE,

> CustomerID STRING,

> Country STRING

> )

> ROW FORMAT DELIMITED

> FIELDS TERMINATED BY ','

> STORED AS TEXTFILE;

OK

Time taken: 0.19 seconds

hive> load data local inpath 'Desktop/retail.csv' into table retail;

FAILED: SemanticException Line 1:23 Invalid path ''Desktop/retail.csv'': No files matching path file:/home/cloudera/Desktop/Desktop/retail.csv

hive> LOAD DATA LOCAL INPATH '/home/cloudera/Desktop/retail.txt' INTO TABLE reatil;

FAILED: SemanticException [Error 10001]: Line 1:70 Table not found 'reatil'

hive> LOAD DATA LOCAL INPATH '/home/cloudera/Desktop/retail.txt' INTO TABLE retail;

Loading data to table default.retail

Table default.retail stats: [numFiles=1, totalSize=604]

OK

Time taken: 0.973 seconds

hive> select \* from retail;

OK

536365 85123A WHITE HANGING HEART T-LIGHT HOLDER 6 12/1/2010 8:26 2.55 17850 United Kingdom

536366 71053 WHITE METAL LANTERN 6 12/1/2010 8:28 3.39 17850 United Kingdom

536367 84406B CREAM CUPID HEARTS COAT HANGER 8 12/1/2010 8:34 2.75 13047 United Kingdom

536368 84029G KNITTED UNION FLAG HOT WATER BOTTLE 6 12/1/2010 8:34 3.39 13047 United Kingdom

536369 21777 BOX OF 6 ASSORTED COLOUR TEASPOONS 6 12/1/2010 8:35 4.25 13047 United Kingdom

536370 22728 ALARM CLOCK BAKELIKE PINK 24 12/1/2010 8:45 3.75 12583 France

536371 22086 PAPER CHAIN KIT 50'S CHRISTMAS 80 12/1/2010 9:00 2.55 13748 United Kingdom

Time taken: 0.55 seconds, Fetched: 7 row(s)

hive> CREATE TABLE combined\_info AS

> SELECT c.Cust\_ID, c.Cust\_Name, c.Order\_ID, o.Item\_ID, o.Quantity, i.Item\_Name, i.Item\_Price

> FROM customer\_info c

> JOIN order\_info o ON c.Order\_ID = o.Order\_ID

> JOIN item\_info i ON o.Item\_ID = i.Item\_ID;

Query ID = cloudera\_20240508231212\_50b65260-caac-4503-8a73-26cd21465bfc

Total jobs = 1

Execution log at: /tmp/cloudera/cloudera\_20240508231212\_50b65260-caac-4503-8a73-26cd21465bfc.log

2024-05-08 11:13:05 Starting to launch local task to process map join; maximum memory = 1013645312

2024-05-08 11:13:07 Dump the side-table for tag: 1 with group count: 0 into file: file:/tmp/cloudera/36a369e6-bd40-4e41-9331-806afcc6f9dd/hive\_2024-05-08\_23-12-52\_128\_7039638056699373558-1/-local-10005/HashTable-Stage-6/MapJoin-mapfile01--.hashtable

2024-05-08 11:13:07 Uploaded 1 File to: file:/tmp/cloudera/36a369e6-bd40-4e41-9331-806afcc6f9dd/hive\_2024-05-08\_23-12-52\_128\_7039638056699373558-1/-local-10005/HashTable-Stage-6/MapJoin-mapfile01--.hashtable (260 bytes)

2024-05-08 11:13:07 Dump the side-table for tag: 1 with group count: 0 into file: file:/tmp/cloudera/36a369e6-bd40-4e41-9331-806afcc6f9dd/hive\_2024-05-08\_23-12-52\_128\_7039638056699373558-1/-local-10005/HashTable-Stage-6/MapJoin-mapfile11--.hashtable

2024-05-08 11:13:07 Uploaded 1 File to: file:/tmp/cloudera/36a369e6-bd40-4e41-9331-806afcc6f9dd/hive\_2024-05-08\_23-12-52\_128\_7039638056699373558-1/-local-10005/HashTable-Stage-6/MapJoin-mapfile11--.hashtable (260 bytes)

2024-05-08 11:13:07 End of local task; Time Taken: 2.334 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\_1715234211897\_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0001/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0001

Hadoop job information for Stage-6: number of mappers: 1; number of reducers: 0

2024-05-08 23:13:35,584 Stage-6 map = 0%, reduce = 0%

2024-05-08 23:13:51,595 Stage-6 map = 100%, reduce = 0%, Cumulative CPU 2.68 sec

MapReduce Total cumulative CPU time: 2 seconds 680 msec

Ended Job = job\_1715234211897\_0001

Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/combined\_info

Table default.combined\_info stats: [numFiles=1, numRows=0, totalSize=0, rawDataSize=0]

MapReduce Jobs Launched:

Stage-Stage-6: Map: 1 Cumulative CPU: 2.68 sec HDFS Read: 8259 HDFS Write: 47 SUCCESS

Total MapReduce CPU Time Spent: 2 seconds 680 msec

OK

Time taken: 62.31 seconds

hive> CREATE INDEX idx\_customer\_id ON TABLE customer\_info(Cust\_ID) AS 'COMPACT' WITH DEFERRED REBUILD;

OK

Time taken: 0.58 seconds

hive> SELECT SUM(Quantity \* Item\_Price) AS total\_sales,

> AVG(Quantity \* Item\_Price) AS average\_sales

> FROM combined\_info;

Query ID = cloudera\_20240508231616\_fcfeb39d-f0bc-4aab-a438-f23df191bef6

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\_1715234211897\_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0002

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

2024-05-08 23:16:52,097 Stage-1 map = 0%, reduce = 0%

2024-05-08 23:17:06,317 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.2 sec

2024-05-08 23:17:20,991 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.32 sec

MapReduce Total cumulative CPU time: 4 seconds 320 msec

Ended Job = job\_1715234211897\_0002

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.32 sec HDFS Read: 9813 HDFS Write: 6 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 320 msec

OK

NULL NULL

Time taken: 51.13 seconds, Fetched: 1 row(s)

hive> CREATE TABLE order\_info (

> Order\_ID STRING,

> Item\_ID STRING,

> Quantity INT

> );

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table order\_info already exists)

hive>

> CREATE TABLE item\_info (

> Item\_ID STRING,

> Item\_Name STRING,

> Item\_Price DOUBLE

>

> ;

MismatchedTokenException(-1!=296)

at org.antlr.runtime.BaseRecognizer.recoverFromMismatchedToken(BaseRecognizer.java:617)

at org.antlr.runtime.BaseRecognizer.match(BaseRecognizer.java:115)

at org.apache.hadoop.hive.ql.parse.HiveParser.createTableStatement(HiveParser.java:4846)

at org.apache.hadoop.hive.ql.parse.HiveParser.ddlStatement(HiveParser.java:2355)

at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:1579)

at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1057)

at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:199)

at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:166)

at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:393)

at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:307)

at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1110)

at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1158)

at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1047)

at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1037)

at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:207)

at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:159)

at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:370)

at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:756)

at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:675)

at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:615)

at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)

at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)

at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)

at java.lang.reflect.Method.invoke(Method.java:606)

at org.apache.hadoop.util.RunJar.run(RunJar.java:221)

at org.apache.hadoop.util.RunJar.main(RunJar.java:136)

FAILED: ParseException line 6:0 mismatched input '<EOF>' expecting ) near 'DOUBLE' in create table statement

hive> SELECT SUM(Quantity \* UnitPrice) AS total\_sales,

> AVG(Quantity \* UnitPrice) AS average\_sales

> FROM online\_retail;

Query ID = cloudera\_20240508232020\_6fcb89e5-e784-4edc-96b9-2d1e301f62c2

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\_1715234211897\_0003, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0003/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0003

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

2024-05-08 23:20:29,612 Stage-1 map = 0%, reduce = 0%

2024-05-08 23:20:42,145 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.02 sec

2024-05-08 23:20:54,721 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.18 sec

MapReduce Total cumulative CPU time: 4 seconds 180 msec

Ended Job = job\_1715234211897\_0003

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.18 sec HDFS Read: 11256 HDFS Write: 37 SUCCESS

Total MapReduce CPU Time Spent: 4 seconds 180 msec

OK

358.5999984741211 19.922222137451172

Time taken: 44.0 seconds, Fetched: 1 row(s)

hive> SELECT \*

> FROM online\_retail

> ORDER BY (Quantity \* UnitPrice) DESC

> LIMIT 1;

Query ID = cloudera\_20240508232121\_ea99ae3e-42dc-461d-8fa5-c9714f177608

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\_1715234211897\_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0004/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0004

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

2024-05-08 23:21:25,123 Stage-1 map = 0%, reduce = 0%

2024-05-08 23:21:54,100 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.54 sec

2024-05-08 23:22:21,291 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 4.12 sec

2024-05-08 23:22:25,567 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.03 sec

MapReduce Total cumulative CPU time: 5 seconds 30 msec

Ended Job = job\_1715234211897\_0004

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.03 sec HDFS Read: 11061 HDFS Write: 89 SUCCESS

Total MapReduce CPU Time Spent: 5 seconds 30 msec

OK

536367 84879 ASSORTED COLOUR BIRD ORNAMENT 32 01/12/2010 08:34 1.69 13047 United Kingdom

Time taken: 87.068 seconds, Fetched: 1 row(s)

hive> SELECT CustomerID, SUM(Quantity \* UnitPrice) AS total\_order

> FROM online\_retail

> GROUP BY CustomerID

> ORDER BY total\_order DESC

> LIMIT 1;

Query ID = cloudera\_20240508232222\_87b21518-20d8-46fc-860d-5216ba6085df

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\_1715234211897\_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0005/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0005

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

2024-05-08 23:23:04,327 Stage-1 map = 0%, reduce = 0%

2024-05-08 23:23:21,960 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.65 sec

2024-05-08 23:23:38,754 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.89 sec

MapReduce Total cumulative CPU time: 4 seconds 890 msec

Ended Job = job\_1715234211897\_0005

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\_1715234211897\_0006, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0006/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0006

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

2024-05-08 23:23:55,419 Stage-2 map = 0%, reduce = 0%

2024-05-08 23:24:09,218 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.47 sec

2024-05-08 23:24:27,244 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 3.71 sec

MapReduce Total cumulative CPU time: 3 seconds 710 msec

Ended Job = job\_1715234211897\_0006

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.89 sec HDFS Read: 9780 HDFS Write: 203 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 3.71 sec HDFS Read: 4701 HDFS Write: 25 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 600 msec

OK

13047 197.27999782562256

Time taken: 105.215 seconds, Fetched: 1 row(s)

hive> SELECT Country, SUM(Quantity \* UnitPrice) AS total\_sales

> FROM online\_retail

> GROUP BY Country

> ORDER BY total\_sales DESC

> LIMIT 1;

Query ID = cloudera\_20240508233131\_fa60ad91-4d17-4e14-bb41-83baa5eabe4c

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\_1715234211897\_0008, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0008/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0008

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

2024-05-08 23:31:47,054 Stage-1 map = 0%, reduce = 0%

2024-05-08 23:32:15,532 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.39 sec

2024-05-08 23:32:40,088 Stage-1 map = 100%, reduce = 67%, Cumulative CPU 4.12 sec

2024-05-08 23:32:41,383 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.75 sec

MapReduce Total cumulative CPU time: 4 seconds 750 msec

Ended Job = job\_1715234211897\_0008

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\_1715234211897\_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application\_1715234211897\_0009/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job\_1715234211897\_0009

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

2024-05-08 23:33:04,549 Stage-2 map = 0%, reduce = 0%

2024-05-08 23:33:17,962 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.65 sec

2024-05-08 23:33:33,668 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.05 sec

MapReduce Total cumulative CPU time: 4 seconds 50 msec

Ended Job = job\_1715234211897\_0009

MapReduce Jobs Launched:

Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.75 sec HDFS Read: 9779 HDFS Write: 178 SUCCESS

Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.05 sec HDFS Read: 4667 HDFS Write: 33 SUCCESS

Total MapReduce CPU Time Spent: 8 seconds 800 msec

OK

United Kingdom 358.5999984741211

Time taken: 149.405 seconds, Fetched: 1 row(s)

hive>