HIVE ECOMMERCE PROJECT

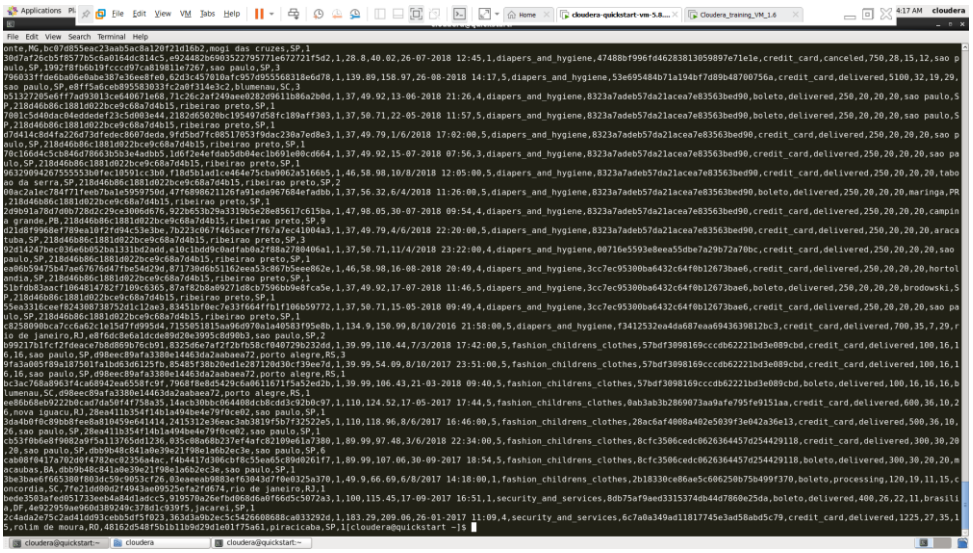
Hive DB used: ecommerce
MySQL DB used: ecommerce

1. After Importing to HDFS:

hdfs dfs -mkdir ecom_Pj
hdfs dfs -put ecom_data.csv ecom_Pj

```
t_card,delivered,1225,27,35,15,rolim de moura,RO,48162d548f5b1b11b9d29d1e01f75a
cloudera@quickstart ~]$ hdfs dfs -ls ecom_Pj/ecom_data.csv
-rw-r--r--   1 cloudera cloudera   28483259 2023-05-19 03:56 ecom_Pj/ecom_data.csv
cloudera@quickstart ~]$
```

hdfs dfs -cat ecom_Pj/ ecom_data.csv



2. Ecommerce table created in ecommerce database in Hive:

create database ecommerce;

```
hive> show databases;
OK
airport
default
ecommerce
sales
Time taken: 0.603 seconds, Fetched: 4 row(s)
hive> use ecommerce;
OK
Time taken: 0.027 seconds
```

create table ecom (order_id string, cust_id string, qty int, MRP float, payment float, order_pchTime string, rating int, prodCat string, prodID string, paymt_type string, order_status string, prod_weightGm int, prod_len_cm int, prod_height_cm int, prod_width_cm int, cust_city string, cust_state string,
seller_id string, seller_city string, seller_state string, paymt_insmt int)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES ("skip.header.line.count" = "1")

```
hive> create table ecom(order_id string, cust_id string, qty int, MRP float, payment float, order_pchTime string, rating int, prodCat string, prodID string,
    > paymt_type string, order_status string, prod_weightGm int, prod_len_cm int, prod_height_cm int, prod_width_cm int,cust_city string, cust_state string,
    > seller_id string, seller_city string, seller_state string, paymt_insmt int)
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  TBLPROPERTIES ("skip.header.line.count" = "1")
    > ;
OK
Time taken: 0.18 seconds
```

```
hive> describe ecom;
OK
order_id                string
cust_id                 string
qty                     int
mrp                     float
payment                 float
order_pchtime           string
rating                  int
prodcat                 string
prodid                  string
paymt_type              string
order_status            string
prod_weightgm           int
prod_len_cm             int
prod_height_cm          int
prod_width_cm           int
cust_city               string
cust_state              string
seller_id               string
seller_city             string
seller_state            string
paymt_insmt             int
Time taken: 0.134 seconds, Fetched: 21 row(s)
```

3. LOAD DATA INPATH 'ecom_Pj/ecom_data.csv' INTO TABLE ecom

Select * from ecom limit 5;

```
hive> LOAD DATA INPATH 'ecom_pr/ecom_data.csv' INTO TABLE ecom;
Loading data to table ecommerce.ecom
chgrp: changing ownership of 'hdfs://quickstart.cloudera:8020/user/hive/warehouse/ecommerce.db/ecom/ecom_data.csv': User does not belong to supergroup
Table ecommerce.ecom stats: [numFiles=1, totalSize=28483259]
OK
Time taken: 0.647 seconds
hive> select * from ecom limit 5;
OK
9045fa841de3514ce900f5fc869d9b1f        2f8d6af8cb3875d6964eb4b5eff3a718        1       1074.38 1095.65 21-11-2017 22:34        5       health_beauty   0a37e05528984f47c7
54837344732fbb  credit_card     delivered       800     40      20      30      osasco  SP      ccc4bbb5f32a6ab2b7066a4130f114e3        curitiba        PR      5
404c087c1f661898fcfe9d4c4b1480d7        1ce70910bc16481433b771949f74552c        1       145.0   161.71  24-02-2017 22:55        5       health_beauty   67473aa97e981b3b8a
baa6a8a768eb26  credit_card     delivered       400     38      12      25      campos dos goytacazes   RJ      ccc4bbb5f32a6ab2b7066a4130f114e3        curitiba        PR
7
d6d7c431275f0029dcc3538850930046        5b477d525dd0c1eb93148e22ce5d3ba6        1       145.0   161.71  19-01-2017 14:28        4       health_beauty   67473aa97e981b3b8a
baa6a8a768eb26  credit_card     delivered       400     38      12      25      nova friburgo   RJ      ccc4bbb5f32a6ab2b7066a4130f114e3        curitiba        PR      1
c0e02613bf15378f18d3408361a76c9d        c2f181778a12039532d79460d4ba59e2        1       555.0   585.95  1/6/2018 22:22:00       4       health_beauty   b60a0c8bd0333c69ee
7056597309bbf  boleto  delivered       650     16      10      11      porto velho     RO      ccc4bbb5f32a6ab2b7066a4130f114e3        curitiba        PR      1
1bf38e3450f15beeafbebe6e0665d612        21a9772e10934cf4598522d6ec827697        1       226.8   250.32  15-06-2018 20:36        5       health_beauty   08462528607b71ea62
cbc019d91b001e  credit_card     delivered       650     16      10      11      montes claros   MG      ccc4bbb5f32a6ab2b7066a4130f114e3        curitiba        PR      8
Time taken: 0.327 seconds, Fetched: 5 row(s)
```

OR

4. Putting data in MYSQL:

3.1.1 Creating Table in MySQL console:

CREATE TABLE ecom(order_id VARCHAR(100), cust_id VARCHAR(100), qty int, MRP float, payment float, order_pchTime VARCHAR(100), rating int, prodCat VARCHAR(100), prodID VARCHAR(100),
paymt_type VARCHAR(100), order_status VARCHAR(100), prod_weightGm int, prod_len_cm int, prod_height_cm int, prod_width_cm int,cust_city VARCHAR(100), cust_state VARCHAR(100),
seller_id VARCHAR(100), seller_city VARCHAR(100), seller_state VARCHAR(100), paymt_insmt int)

3.1.2: LOADING Data FROM LFS ----------> MySQL table:

LOAD DATA LOCAL INFILE '/home/cloudera/ecom_data.csv' INTO TABLE ecom FIELDS TERMINATED BY ',';

3.2. SQOOP Pipeline for MySQL ----------> Hive:

sqoop import --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table ecom --hive-import -m 1

==========

4-> HIVE JOBS

==========

JOB 1: Customer Segmentation: Categorizing customers based on their spendings

```
SELECT cust_id, ROUND(SUM(payment),3) AS spendings,
CASE
WHEN ROUND(SUM(payment),3) > 20000 THEN 'GROUP 1'
WHEN ROUND(SUM(payment),3) > 5000 THEN 'GROUP 2'
WHEN ROUND(SUM(payment),3) > 1500 THEN 'GROUP 3'
WHEN ROUND(SUM(payment),3) > 100 THEN 'GROUP 4'
ELSE 'GROUP 5'
END AS cust_cat
FROM ecom GROUP BY cust_id ORDER BY spendings DESC;
```

1.1.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job1_op(cust_id string, spendings float, cust_cat string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION '/user/hive/warehouse/ecommerce/job1_op.txt';
```

```
hive> CREATE EXTERNAL TABLE job1_op(cust_id string, spendings float, cust_cat string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
    > LOCATION '/user/hive/warehouse/ecommerce/job1_op.txt';
OK
Time taken: 0.193 seconds
```

1.1.2-> Transferring o/p to ext table:

```
insert overwrite table job1_op SELECT cust_id, ROUND(SUM(payment),3) AS spendings,
CASE
WHEN ROUND(SUM(payment),3) > 20000 THEN 'GROUP 1'
WHEN ROUND(SUM(payment),3) > 5000 THEN 'GROUP 2'
WHEN ROUND(SUM(payment),3) > 1500 THEN 'GROUP 3'
WHEN ROUND(SUM(payment),3) > 100 THEN 'GROUP 4'
ELSE 'GROUP 5'
END AS cust_cat
FROM ecom GROUP BY cust_id ORDER BY spendings DESC;
```

```
Time taken: 22.012 seconds, Fetched: 1 row(s)
hive> insert overwrite table job1_op SELECT cust_id, ROUND(SUM(payment),3) AS spendings,
    > CASE
    > WHEN ROUND(SUM(payment),3) > 20000 THEN 'GROUP 1'
    > WHEN ROUND(SUM(payment),3) > 5000 THEN 'GROUP 2'
    > WHEN ROUND(SUM(payment),3) > 1500 THEN 'GROUP 3'
    > WHEN ROUND(SUM(payment),3) > 100 THEN 'GROUP 4'
    > ELSE 'GROUP 5'
    > END AS cust_cat
    > FROM ecom GROUP BY cust_id ORDER BY spendings DESC;
Query ID = cloudera_20230521215757_d5732c1e-2070-4185-bc2d-9de764190ab4
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_1684729337865_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684729337865_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684729337865_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-21 21:57:50,889 Stage-1 map = 0%,  reduce = 0%
```

```
hive> select * from job1_op limit 50;
OK
1617b1357756262bfa56ab541c47bc16        109312.64       GROUP 1
bd5d39761aa56689a265d95d8d32b8be        45256.0 GROUP 1
be1b70680b9f9694d8c70f41fa3dc92b        44047.996       GROUP 1
05455dfa7cd02f13d132aa7a6a9729c6        36489.24        GROUP 1
1ff773612ab8934db89fd5afa8afe506        30186.002       GROUP 1
ec5b2ba62e574342386871631fafd3fc        29099.52        GROUP 1
e7d6802668de6e74d0d6c56565bf2a24        22346.6 GROUP 1
8c20d9bfbc96c5d39025d77a3ba83d7f        21874.05        GROUP 1
f7622098214b4634b7fe7eee269b5426        19457.041       GROUP 2
71901689c5f3e5adc27b1dd16b33f0b8        19174.38        GROUP 2
be1c4e52bb71e0c54b11a26b8e8d59f2        18786.9 GROUP 2
10de381f8a8d23fff822753305f71cae        18384.75        GROUP 2
91f92cfee46b79581b05aa974dd57ce5        17786.88        GROUP 2
4a60b2ce1ee8c7b828e4bbcca5b86b41        17671.0 GROUP 2
ed583a2a1eaf0dedc33af816153e8b8c        17069.76        GROUP 2
0c792d32a3251b4f69dae8646dfbedbc        16313.6 GROUP 2
daf15f1b940cc6a72ba558f093dc00dd        14963.64        GROUP 2
88324c93ce11436ae046563bf0da285c        14577.57        GROUP 2
78fc46047c4a639e81ff65f0396e02fe        14401.0 GROUP 2
d5f2b3f597c7ccafbb5cac0bcc3d6024        14196.28        GROUP 2
03a6f3a3935165f5d26f4797ba6039bb        12834.5 GROUP 2
6152fbfc8a92ee25fd821740bd33b089        12490.88        GROUP 2
e2f063c52fa92b54f186c62b90699c10        12300.0 GROUP 2
5c51fe11e0c3a016201653955c6090ee        11832.18        GROUP 2
adb32467ecc74b53576d9d13a5a55891        11745.0 GROUP 2
e91409937a0fd26a57310548858d1762        11697.12        GROUP 2
0d861a5e4dd6a9079d89e1330848f0ab        11572.8 GROUP 2
f4c13379ddd0ed4f4fc1c0b49c2c5e51        11531.411       GROUP 2
159132ab31eb3c8a1061ed18967f720b        11468.45        GROUP 2
a1a52dda6ac52aee936ad06712027b83        11380.5 GROUP 2
cb87122c4871e202777cf243fbea2d12        11200.2 GROUP 2
7d03bf20fa96e80468bbf678eebbcb3f        10999.26        GROUP 2
068830daf2867cadbfc655e6a413388e        9759.75 GROUP 2
727ab4449f530b9af13f40fa8e528bff        9751.0  GROUP 2
e0a2412720e9ea4f26c1ac985f6a7358        9618.88 GROUP 2
```

1.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job1(customer_id VARCHAR(33), Spendings Float, Customer_category VARCHAR(10))
```

****

JOB 2: Monthly Trend Forecasting: the monthly trend of sales

```
SELECT yr, CASE WHEN mon='01' THEN 'January'
WHEN mon='02' THEN 'February'
WHEN mon='03' THEN 'March'
WHEN mon='04' THEN 'April'
WHEN mon='05' THEN 'May'
WHEN mon='06' THEN 'June'
WHEN mon='07' THEN 'July'
WHEN mon='08' THEN 'August'
WHEN mon='09' THEN 'September'
WHEN mon='10' THEN 'October'
WHEN mon='11' THEN 'November'
WHEN mon='12' THEN 'December'
ELSE mon END AS month, sales ,
ROUND((sales - (LAG(sales) OVER()))) AS sales_trend
from(SELECT SUBSTRING(tstamp,7,4) as yr, SUBSTRING(tstamp,4,2) as mon, ROUND(SUM(qty*mrp),3) as sales from ecom
GROUP BY SUBSTRING(tstamp,7,4), SUBSTRING(tstamp,4,2) ORDER BY yr DESC, mon DESC) t;
```

2.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job2_op(year INT, month string, sales float, sales_trend INT);
```

2.2-> Transferring o/p to ext table:

```
insert overwrite table job2_op
SELECT yr, CASE WHEN mon='01' THEN 'January'
WHEN mon='02' THEN 'February'
WHEN mon='03' THEN 'March'
WHEN mon='04' THEN 'April'
WHEN mon='05' THEN 'May'
WHEN mon='06' THEN 'June'
WHEN mon='07' THEN 'July'
WHEN mon='08' THEN 'August'
WHEN mon='09' THEN 'September'
WHEN mon='10' THEN 'October'
WHEN mon='11' THEN 'November'
WHEN mon='12' THEN 'December'
ELSE mon END AS month, sales ,
ROUND((sales - (LAG(sales) OVER()))) AS sales_trend
from(SELECT SUBSTRING(tstamp,7,4) as yr, SUBSTRING(tstamp,4,2) as mon, ROUND(SUM(qty*mrp),3) as sales from ecom
GROUP BY SUBSTRING(tstamp,7,4), SUBSTRING(tstamp,4,2) ORDER BY yr DESC, mon DESC) t;
```

```
hive> insert overwrite table job2_op
    > SELECT yr, CASE WHEN mon='01' THEN 'January'
    > WHEN mon='02' THEN 'February'
    > WHEN mon='03' THEN 'March'
    > WHEN mon='04' THEN 'April'
    > WHEN mon='05' THEN 'May'
    > WHEN mon='06' THEN 'June'
    > WHEN mon='07' THEN 'July'
    > WHEN mon='08' THEN 'August'
    > WHEN mon='09' THEN 'September'
    > WHEN mon='10' THEN 'October'
    > WHEN mon='11' THEN 'November'
    > WHEN mon='12' THEN 'December'
    > ELSE mon END AS month, sales ,
    > ROUND((sales - (LAG(sales) OVER()))) AS sales_trend
    > from(SELECT SUBSTRING(tstamp,7,4) as yr, SUBSTRING(tstamp,4,2) as mon, ROUND(SUM(qty*mrp),3) as sales from ecom
    > GROUP BY SUBSTRING(tstamp,7,4), SUBSTRING(tstamp,4,2) ORDER BY yr DESC, mon DESC) t;
Query ID = cloudera_20230520095454_4b13861d-5751-4e8c-a23b-879a8bf999c5
Total jobs = 3
Launching Job 1 out of 3
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_1684567613417_0055, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0055/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0055
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 09:54:48,376 Stage-1 map = 0%,  reduce = 0%
MapReduce Total cumulative CPU time: 2 seconds 260 msec
Ended Job = job_1684567613417_0057
Loading data to table ecommerce.job2_op
Table ecommerce.job2_op stats: [numFiles=1, numRows=24, totalSize=681, rawDataSize=657]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.61 sec   HDFS Read: 28435198 HDFS Write: 888 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 1.78 sec   HDFS Read: 5044 HDFS Write: 888 SUCCESS
Stage-Stage-3: Map: 1  Reduce: 1   Cumulative CPU: 2.26 sec   HDFS Read: 9619 HDFS Write: 755 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 650 msec
OK
Time taken: 74.672 seconds
hive> select * from job2_op;
OK
2016    September       165.29  NULL
2016    October 58085.85        57921
2016    December        10.9    -58075
2017    January 148449.92       148439
2017    February        276258.16       127808
2017    March   431746.97       155489
2017    April   421562.66       -10184
2017    May     595436.94       173874
2017    June    489713.5        -105723
2017    July    596519.75       106806
2017    August  689648.9        93129
2017    September       792979.94       103331
2017    October 782517.5        -10462
2017    November        1209636.2       427119
2017    December        829291.0        -380345
2018    January 1095322.8       266032
2018    February        1010978.75      -84344
2018    March   1138198.2       127219
2018    April   1158672.4       20474
2018    May     1171104.1       12432
2018    June    1023277.5       -147827
2018    July    1043661.06      20384
2018    August  976499.75       -67161
2018    September       145.0   -976355
Time taken: 0.055 seconds, Fetched: 24 row(s)
```

2.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job2(Year INT, Month VARCHAR(10), sales FLOAT, sales_trend INT);
```

****

JOB 3 : Hourly Sales Analysis: Which hour has more no. of sales?

```
SELECT CASE WHEN hr ='00' THEN '12 AM'
WHEN hr>='01' AND  hr<=11 THEN CONCAT(hr,' AM')
WHEN hr='12' THEN '12 PM'
WHEN hr>'12' THEN CONCAT(CAST(hr-12 AS INT),' PM')
ELSE CONCAT(hr-12,' PM') END AS hr, no_sales
FROM(SELECT SUBSTRING(tstamp,12,2) as hr, COUNT(order_id) as no_sales from ecom
GROUP BY SUBSTRING(tstamp,12,2) ORDER BY no_sales DESC)t;
```

3.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job3_op(hour string, no_of_sales INT);
```

3.2-> Transferring o/p to ext table:

```
insert overwrite table job3_op  SELECT CASE WHEN hr ='00' THEN '12 AM'
WHEN hr>='01' AND  hr<=11 THEN CONCAT(hr,' AM')
WHEN hr='12' THEN '12 PM'
WHEN hr>'12' THEN CONCAT(CAST(hr-12 AS INT),' PM')
ELSE CONCAT(hr-12,' PM') END AS hr, no_sales
FROM(SELECT SUBSTRING(tstamp,12,2) as hr, COUNT(order_id) as no_sales from ecom
GROUP BY SUBSTRING(tstamp,12,2) ORDER BY no_sales DESC)t;
```

```
hive> CREATE EXTERNAL TABLE job3_op(hour string, no_of_sales INT)
    > ;
OK
Time taken: 0.119 seconds
hive> insert overwrite table job3_op  SELECT CASE WHEN hr ='00' THEN '12 AM'
    > WHEN hr>='01' AND  hr<=11 THEN CONCAT(hr,' AM')
    > WHEN hr='12' THEN '12 PM'
    > WHEN hr>'12' THEN CONCAT(CAST(hr-12 AS INT),' PM')
    > ELSE CONCAT(hr-12,' PM') END AS hr, no_sales
    > FROM(SELECT SUBSTRING(tstamp,12,2) as hr, COUNT(order_id) as no_sales from ecom
    > GROUP BY SUBSTRING(tstamp,12,2) ORDER BY no_sales DESC)t
    > ;
Query ID = cloudera_20230520120303_4d34aee3-b843-452a-bfa6-92d8db86a7cc
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_1684567613417_0076, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0076/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0076
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
Time taken: 53.093 seconds
hive> select * from job3_op;
OK
4 PM    7901
2 PM    7833
11 AM   7688
1 PM    7629
3 PM    7558
10 AM   7256
5 PM    7193
8 PM    7190
9 PM    7170
12 PM   7115
7 PM    6923
10 PM   6869
6 PM    6791
09 AM   5573
11 PM   4853
08 AM   3487
12 AM   2871
07 AM   1405
01 AM   1324
02 AM   600
06 AM   565
03 AM   320
04 AM   252
05 AM   214
Time taken: 0.065 seconds, Fetched: 24 row(s)
hive>
```

OBSERVATION: Most no. of sales happens at afternoon 4 PM, followed by 2 PM .
(Top 5 hours when no. of sales is most:

4 PM	7901
2 PM	7833
11 AM	7688
1 PM	7629
3 PM	7558

)

3.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job3(hour VARCHAR(6), No_of_Sales INT);
```

****

JOB 4 : Product Based Analysis:

4.1-> Which category product has sold more?

```
SELECT prodCat, SUM(qty) AS no_sold FROM ecom GROUP BY prodCat ORDER BY no_sold DESC;
```

4.1.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job4a_op(prod_category string, no_sold INT)
```

4.1.2-> Transferring o/p to ext table:

insert overwrite table job4a_op SELECT prodCat, SUM(qty) AS no_sold FROM ecom GROUP BY prodCat ORDER BY no_sold DESC

```
Time taken: 0.049 seconds, Fetched: 10 row(s)
hive> CREATE EXTERNAL TABLE job4a_op(prod_category string, no_sold INT)
    > ;
OK
Time taken: 0.07 seconds
hive> insert overwrite table job4a_op SELECT prodCat, SUM(qty) AS no_sold FROM ecom GROUP BY prodCat ORDER BY no_sold DESC
    > ;
Query ID = cloudera_20230520130707_c31d22a7-0d0a-4647-9556-5ac9f7d1beb7
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_1684567613417_0081, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0081/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0081
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 13:07:32,173 Stage-1 map = 0%,  reduce = 0%
```

```
hive> SELECT * FROM job4a_op;
OK
bed_bath_table  14797
furniture_decor 12186
health_beauty   11486
sports_leisure  10349
computers_accessories   10273
housewares      9559
watches_gifts   6823
garden_tools    6152
telephony       5354
auto    5067
Time taken: 0.05 seconds, Fetched: 10 row(s)
```

OBSERVATION: According to client's data, 'Bed Bath Table' product category has been sold the most.
(Top 3 product categories sold most:
bed_bath_table 14797
furniture_decor 12186
health_beauty 11486
)

4.1.3-> MySQL Table Created (Client's DB):
CREATE TABLE job4a(Product_Category VARCHAR(50), sold_count INT);

4.2-> Which category product has more rating?

SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count DESC;

4.2.1-> Ext Table Creation:
CREATE EXTERNAL TABLE job4b_op(prod_category string, rating INT, rt_count INT);

4.2.2-> Transferring o/p to ext table:

insert overwrite table job4b_op SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count DESC LIMIT 13;

```
hive> CREATE EXTERNAL TABLE job4b_op(prod_category string, rating INT, rt_count INT)
    > ;
OK
Time taken: 0.136 seconds
hive> insert overwrite table job4b_op SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count DESC LIMIT 13;
Query ID = cloudera_20230520134646_2a45c22a-716f-4e03-b56f-cc8aafde5672
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_1684567613417_0096, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0096/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0096
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 13:46:18,405 Stage-1 map = 0%,  reduce = 0%
Table ecommerce.job4b_op stats: [numFiles=1, numRows=1
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.1
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 3.1
Total MapReduce CPU Time Spent: 7 seconds 290 msec
OK
Time taken: 59.648 seconds
hive> select * from job4b_op;
OK
bed_bath_table  5       6179
health_beauty   5       6036
sports_leisure  5       5292
furniture_decor 5       4708
computers_accessories   5       4351
housewares      5       4206
watches_gifts   5       3454
toys    5       2623
garden_tools    5       2571
auto    5       2456
telephony       5       2416
cool_stuff      5       2332
perfumery       5       2244
Time taken: 0.094 seconds, Fetched: 13 row(s)
```

OBSERVATION: 'Bed Bath Table' product category has the most rating, followed by 'Health Beauty'.
(Top 5 product categories having most rating:
bed_bath_table 5 6179
health_beauty 5 6036
sports_leisure 5 5292
furniture_decor 5 4708
computers_accessories 5 4351

)

4.2.3-> MySQL Table Created (Client's DB):
CREATE TABLE job4b(Product_Category VARCHAR(50), Rating INT, rating_count INT);

4.3-> Which product has sold more?
SELECT prodID, prodCat, SUM(qty) AS no_sold FROM ecom GROUP BY prodID, prodCat ORDER BY no_sold DESC LIMIT 40

4.3.1-> Ext Table Creation:
CREATE EXTERNAL TABLE job4c_op(pID string, pcategory string, no_sold INT)

4.3.2-> Transferring o/p to ext table:

insert overwrite table job4c_op SELECT prodID, prodCat, SUM(qty) AS no_sold FROM ecom GROUP BY prodID, prodCat ORDER BY no_sold DESC LIMIT 10;

```
hive> CREATE EXTERNAL TABLE job4c_op(pID string, pcategory string, no_sold INT);
OK
Time taken: 0.054 seconds
hive> insert overwrite table job4c_op SELECT prodID, prodCat, SUM(qty) AS no_sold FROM ecom GROUP BY prodID, prodCat ORDER BY no_sold DESC LIMIT 10;
Query ID = cloudera_20230520143636_d6b11486-f9c5-42b1-b8a5-f33c80fb29af
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_1684567613417_0106, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0106/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0106
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 14:36:23,219 Stage-1 map = 0%,  reduce = 0%
```

OBSERVATION: Product from Garden Tools has been sold the most, followed by Furniture Decor.
(Top 5 Products sold the most are from categories:

422879e10f46682990de24d770e7f83d	garden_tools	826
aca2eb7d00ea1a7b8ebd4e68314663af	furniture_decor	650
99a4788cb24856965c36a24e339b6058	bed_bath_table	588
368c6c730842d78016ad823897a372db	garden_tools	564
53759a2ecddad2bb87a079a1f1519f73	garden_tools	563

)

4.3.3-> MySQL Table Created (Client's DB):
CREATE TABLE job4c(Product_ID VARCHAR(33), Product_Category VARCHAR(50), sold_count INT);

4.4-> Top 10 highest & least product rating?
Least 10:
SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count LIMIT 10
Top 10:
SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count DESC LIMIT 10

4.4.1-> Ext Table Creation:
Least 10:
CREATE EXTERNAL TABLE job4dLeast10_op(prod_category string, rating INT, rt_count INT)
Top 10:
CREATE EXTERNAL TABLE job4dTop10_op(prod_category string, rating INT, rt_count INT)

4.4.2-> Transferring o/p to ext table:

Least 10:
insert overwrite table job4dLeast10_op SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count LIMIT 10;
Top 10:
insert overwrite table job4dTop10_op SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count DESC LIMIT 10;

```
hive> CREATE EXTERNAL TABLE job4dLeast10_op(prod_category string, rating INT, rt_count INT)
    > ;
OK
Time taken: 0.284 seconds
hive> CREATE EXTERNAL TABLE job4dTop10_op(prod_category string, rating INT, rt_count INT)
    > ;
OK
Time taken: 0.096 seconds
hive> insert overwrite table job4dLeast10_op SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count LIMIT 10
    > ;
Query ID = cloudera_20230520145353_490e2e57-57f8-4eb0-b63a-c99ed4c43c3a
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_1684567613417_0112, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0112/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0112
```

```
hive> insert overwrite table job4dTop10_op SELECT prodCat, rating, COUNT(rating) as rt_count FROM ecom GROUP BY prodCat, rating ORDER BY rt_count DESC LIMIT 10
    > ;
Query ID = cloudera_20230520145656_9d242582-9fec-4acf-9bdf-d2987a9af0c9
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_1684567613417_0114, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0114/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0114
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 14:56:35,539 Stage-1 map = 0%,  reduce = 0%
2023-05-20 14:56:51,171 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.58 sec
2023-05-20 14:57:12,393 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.47 sec
Time taken: 112.277 seconds
```

```
hive> select * from job4dLeast10 op;
OK
books_imported  2       1
fashion_childrens_clothes       1       1
fashion_sport   2       1
dvds_blu_ray    3       1
furniture_mattress_and_upholstery       2       1
dvds_blu_ray    2       1
costruction_tools_tools 3       1
costruction_tools_tools 2       1
security_and_services   1       1
security_and_services   4       1
Time taken: 0.097 seconds, Fetched: 10 row(s)
```

```
hive> select * from job4dTop10 op;
OK
bed_bath_table  5       6179
health_beauty   5       6036
sports_leisure  5       5292
furniture_decor 5       4708
computers_accessories   5       4351
housewares      5       4206
watches_gifts   5       3454
toys    5       2623
garden_tools    5       2571
auto    5       2456
Time taken: 0.103 seconds, Fetched: 10 row(s)
```

4.4.3-> MySQL Table Created (Client's DB):

Least 10:

CREATE TABLE job4d_Least10(Product_Category VARCHAR(50), Rating INT, rating_count INT)

Top 10:

CREATE TABLE job4d_Top10(Product_Category VARCHAR(50), Rating INT, rating_count INT)

4.5-> Order Count for each rating

SELECT rating, COUNT(order_id) AS order_count FROM ecom GROUP BY rating ORDER BY order_count DESC;

4.5.1-> Ext Table Creation:

CREATE EXTERNAL TABLE job4e(rating INT, order_count INT);

4.5.2-> Transferring o/p to ext table:

insert overwrite table job4e SELECT rating, COUNT(order_id) AS order_count FROM ecom GROUP BY rating ORDER BY order_count DESC;

```
hive> CREATE EXTERNAL TABLE job4e(rating INT, order_count INT)
    > ;
OK
Time taken: 0.247 seconds
hive> insert overwrite table job4e SELECT rating, COUNT(order_id) AS order_count FROM ecom GROUP BY rating ORDER BY order_count DESC
    > ;
Query ID = cloudera_20230520150909_82720054-d9c8-45d0-8c1b-f75f8b0d4628
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_1684567613417_0117, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0117/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0117
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 15:09:47,940 Stage-1 map = 0%,  reduce = 0%
```

```
Table ecommerce.job4e stats: [numFiles=1, nu
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative
Total MapReduce CPU Time Spent: 13 seconds 23
OK
Time taken: 115.73 seconds
hive> select * from job4e;
OK
5       65478
4       22017
1       15110
3       9842
2       4133
Time taken: 0.088 seconds, Fetched: 5 row(s)
hive>
```

4.5.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job4e(Rating INT, No_of_Orders INT);
```

****

JOB 5 : Payment Preference:

5.1-> What are the most commonly used payment types?

```
SELECT paymt_type, COUNT(order_id) AS no_orders FROM ecom GROUP BY paymt_type ORDER BY no_orders;
```

5.1.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job5a_op(pay_type string, orders_count INT);
```

5.1.2-> Transferring o/p to ext table:

```
insert overwrite table job5a_op SELECT paymt_type, COUNT(order_id) AS no_orders FROM ecom GROUP BY paymt_type ORDER BY no_orders DESC;
```

```
Time taken: 0.086 seconds, Fetched: 1 row(s)
hive> CREATE EXTERNAL TABLE job5a_op(pay_type string, orders_count INT)
    > ;
OK
Time taken: 0.086 seconds
hive> insert overwrite table job5a_op SELECT paymt_type, COUNT(order_id) AS no_orders FROM ecom GROUP BY paymt_type ORDER BY no_orders DESC
    > ;
Query ID = cloudera_20230520163535_3bf5f53f-a0c6-488f-9a1f-274b7cdc1e54
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_1684567613417_0121, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0121/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0121
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 16:35:49,786 Stage-1 map = 0%,  reduce = 0%
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU
Total MapReduce CPU Time Spent: 7 seconds 790 msec
OK
Time taken: 64.369 seconds
hive> select * from job5a_op;
OK
credit_card     86011
boleto  22692
voucher 6211
debit_card      1666
Time taken: 0.063 seconds, Fetched: 4 row(s)
hive>
```

OBSERVATION: Most Commonly used payment type is Credit Card, followed by Boleto.

5.1.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job5a(Payment_Type VARCHAR(20), No_of_Orders INT);
```

5.2-> Count of Orders With each No. of Payment Installments

```
SELECT paymt_insmt, COUNT( DISTINCT order_id) AS no_orders FROM ecom GROUP BY paymt_insmt ORDER BY paymt_insmt;
```

5.2.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job5b_op(payment_installments INT, no_orders INT);
```

5.2.2-> Transferring o/p to ext table:

```
insert overwrite table job5b_op SELECT paymt_insmt, COUNT( DISTINCT order_id) AS no_orders FROM ecom GROUP BY paymt_insmt ORDER BY paymt_insmt;
```

```
hive> CREATE EXTERNAL TABLE job5b_op(payment_installments INT, no_orders INT)
    > ;
OK
Time taken: 0.251 seconds
hive> insert overwrite table job5b_op SELECT paymt_insmt, COUNT( DISTINCT order_
id)Display all 469 possibilities? (y or n)
hive> insert overwrite table job5b_op SELECT paymt_insmt, COUNT( DISTINCT order_
id)o_orders FROM ecom GROUP BY paymt_insmt ORDER BY paymt_insmt
    > ;
Query ID = cloudera_20230521023939_301b6ec0-da57-4ade-8527-92c853a1fc9a
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_1684660565171_0002, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1684660565171_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684660565171_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-21 02:39:11,689 Stage-1 map = 0%,  reduce = 0%
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative C
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative C
Total MapReduce CPU Time Spent: 7 seconds 880 m
OK
Time taken: 54.223 seconds
hive> select * from job5_op;
FAILED: SemanticException [Error 10001]: Line 1
hive> select * from job5b_op;
OK
NULL    3
0       2
1       47885
2       12091
3       10236
4       6955
5       5130
6       3837
7       1584
8       4194
9       630
10      5216
11      22
12      131
13      16
14      15
15      74
16      5
17      7
18      27
20      17
21      3
22      1
23      1
24      18
Time taken: 0.118 seconds, Fetched: 25 row(s)
hive>
```

5.2.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job5b(Payment_Installments INT, No_of_Orders INT);
```

****

JOB 6 : Potential Customer's Location: Where do most customers come from?

```
SELECT cust_state, cust_city, COUNT(cust_id) AS no_cust FROM ecom GROUP BY cust_state, cust_city ORDER BY no_cust DESC LIMIT 10;
```

1.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job6_op(cust_state string, cust_city string, no_cust INT)
```

1.2-> Transferring o/p to ext table:

```
insert overwrite table job6_op SELECT cust_state, cust_city, COUNT(cust_id) AS no_cust FROM ecom GROUP BY cust_state, cust_city ORDER BY no_cust DESC LIMIT 10;
```

```
hive> CREATE EXTERNAL TABLE job6_op(cust_state string, cust_city string, no_cust INT)
    > ;
OK
Time taken: 0.128 seconds
hive> insert overwrite table job6_op SELECT cust_state, cust_city, COUNT(cust_id) AS no_cust FROM ecom GROUP BY cust_state, cust_city ORDER BY no_cust DESC LIMIT 10
    > ;
Query ID = cloudera_20230520170303_d0091cfe-631f-415f-ad52-b25b1cdb570c
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_1684567613417_0134, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0134/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0134
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 17:04:06,360 Stage-1 map = 0%,  reduce = 0%
Time taken: 0.058 seconds, Fetched: 10 row(s)
hive> select * from job6_op;
OK
SP      sao paulo       18412
RJ      rio de janeiro  8123
MG      belo horizonte  3218
DF      brasilia        2452
PR      curitiba        1799
SP      campinas        1722
RS      porto alegre    1640
BA      salvador        1498
SP      guarulhos       1388
SP      sao bernardo do campo   1100
Time taken: 0.152 seconds, Fetched: 10 row(s)
```

6.1.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job6(Customer_State VARCHAR(5), Customer_City VARCHAR(55), No_Of_Customers INT );
```

****

JOB 7 : Seller Rating:

7.1-> Which seller sold more?

```
SELECT seller_id, ROUND(SUM(qty*mrp), 2) AS sales FROM ecom GROUP BY seller_id ORDER BY sales DESC LIMIT 10;
```

7.1.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job7a_op(seller_id string, sales INT);
```

7.1.2-> Transferring o/p to ext table:

insert overwrite table job7a_op SELECT seller_id, ROUND(SUM(qty*mrp), 2) AS sales FROM ecom GROUP BY seller_id ORDER BY sales DESC LIMIT 10;

```
hive> CREATE EXTERNAL TABLE job7a_op(seller_id string, sales INT)
    > ;
OK
Time taken: 0.096 seconds
hive> insert overwrite table job7a_op SELECT seller_id, ROUND(SUM(qty*mrp), 2) AS sales FROM ecom GROUP BY seller_id ORDER BY sales DESC LIMIT 10;
Query ID = cloudera_20230520171616_bc0fd662-88f1-4547-959b-09796dc4cce2
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_1684567613417_0138, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0138/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0138
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
Time taken: 70.240 seconds
hive> select * from job7a_op;
OK
7c67e1448b00f6e969d365cea6b010ab        308853
53243585a1d6dc2643021fd1853d8905        267642
4a3ca9315b744ce9f8e9374361493884        244215
4869f7a5dfa277a7dca6462dcf3b52b2        244023
da8622b14eb17ae2831f4ac5b9dab84a        210603
fa1c13f2614d7b5c4749cbc52fecda94        205545
1025f0e2d44d7041d6cf58b6550e0bfa        197499
7e93a43ef30c4f03f38b393420bc753a        184351
955fee9216a65b617aa5c0531780ce60        169398
1f50f920176fa81dab994f9023523100        168425
Time taken: 0.058 seconds, Fetched: 10 row(s)
```

7.1.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job7a(Seller_ID VARCHAR(33), Sales_amt INT);
```

7.2-> Which seller got more rating?

```
SELECT seller_id, rating, COUNT(*) AS rt_count FROM ecom GROUP BY seller_id, rating ORDER BY rt_count DESC LIMIT 10;
```

7.2.1-> Ext Table Creation:

```
CREATE EXTERNAL TABLE job7b_op(seller_id string, rating INT, rt_count INT);
```

7.2.2-> Transferring o/p to ext table:

insert overwrite table job7b_op SELECT seller_id, rating, COUNT(*) AS rt_count FROM ecom GROUP BY seller_id, rating ORDER BY rt_count DESC LIMIT 10;

```
hive> CREATE EXTERNAL TABLE job7b_op(seller_id string, rating INT, rt_count INT)
    > ;
OK
Time taken: 0.144 seconds
hive> insert overwrite table job7b_op SELECT seller_id, rating, COUNT(*) AS rt_count FROM ecom GROUP BY seller_id, rating ORDER BY rt_count DESC LIMIT 10;
Query ID = cloudera_20230520172323_957f2e03-f65b-4f73-8236-5bd1a73771af
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_1684567613417_0142, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684567613417_0142/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684567613417_0142
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-20 17:23:58,864 Stage-1 map = 0%,  reduce = 0%
Time taken: 69.728 seconds
hive> select * from job7b_op;
OK
1f50f920176fa81dab994f9023523100        5       1134
cc419e0650a3c5ba77189a1882b7556a        5       1091
6560211a19b47992c3666cc44a7e94c0        5       1070
4a3ca9315b744ce9f8e9374361493884        5       1021
da8622b14eb17ae2831f4ac5b9dab84a        5       950
955fee9216a65b617aa5c0531780ce60        5       818
7a67c85e85bb2ce8582c35f2203ad736        5       753
1025f0e2d44d7041d6cf58b6550e0bfa        5       751
4869f7a5dfa277a7dca6462dcf3b52b2        5       700
3d871de0142ce09b7081e2b9d1733cb1        5       658
Time taken: 0.092 seconds, Fetched: 10 row(s)
```

7.2.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job7b(Seller_ID VARCHAR(33), Rating INT, rating_count INT);
```

****

JOB 8 : Logistics based Optimization Insights:

8.1-> Which city buys heavy weight products and low weight products?

Low Weight Products:

```
WITH cte AS (
```

```
SELECT order_id, cust_id, cust_city, prod_weightgm, CASE WHEN prod_weightgm > 22000 THEN 'High Weight Product'
WHEN prod_weightgm < 2000 THEN 'Low Weight Product'
ELSE 'Medium Weight Product'
END AS pwt_cat
FROM ecom
) SELECT cust_city, pwt_cat, COUNT(*) AS count from cte  where pwt_cat='Low Weight Product' GROUP BY cust_city, pwt_cat ORDER BY count desc;
```

High Weight Products:

```
WITH cte AS (
SELECT order_id, cust_id, cust_city, prod_weightgm, CASE WHEN prod_weightgm > 22000 THEN 'High Weight Product'
WHEN prod_weightgm < 2000 THEN 'Low Weight Product'
ELSE 'Medium Weight Product'
END AS pwt_cat
FROM ecom
) SELECT cust_city, pwt_cat, COUNT(*) AS count from cte  where pwt_cat='High Weight Product' GROUP BY cust_city, pwt_cat ORDER BY count desc;
```

8.1.1-> Ext Table Creation:

Low Weight Products:

```
CREATE EXTERNAL TABLE job8a_op(cust_city string, prod_weightGm INT, count INT);
```

High Weight Products:

```
CREATE EXTERNAL TABLE job8b_op(cust_city string, prod_weightGm INT, count INT);
```

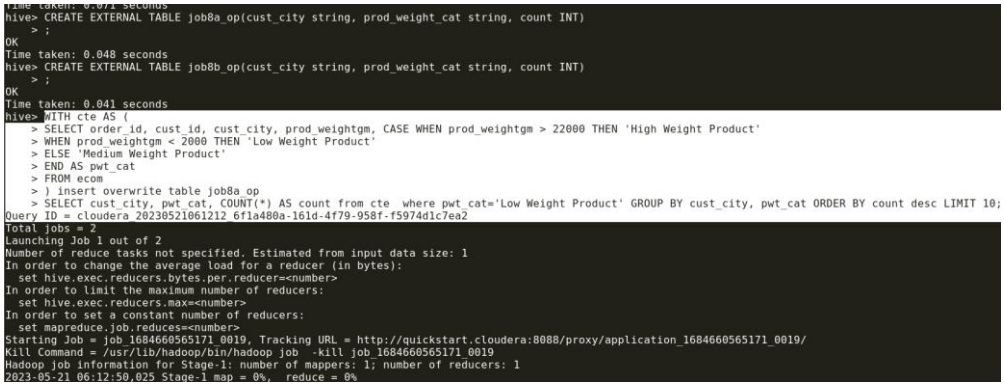
8.1.2-> Transferring o/p to ext table:

Low Weight Products:

```
WITH cte AS (
SELECT order_id, cust_id, cust_city, prod_weightgm, CASE WHEN prod_weightgm > 22000 THEN 'High Weight Product'
WHEN prod_weightgm < 2000 THEN 'Low Weight Product'
ELSE 'Medium Weight Product'
END AS pwt_cat
FROM ecom
) insert overwrite table job8a_op
SELECT cust_city, pwt_cat, COUNT(*) AS count from cte  where pwt_cat='Low Weight Product' GROUP BY cust_city, pwt_cat ORDER BY count desc LIMIT 10;
```

High Weight Products:

```
WITH cte AS (
SELECT order_id, cust_id, cust_city, prod_weightgm, CASE WHEN prod_weightgm > 22000 THEN 'High Weight Product'
WHEN prod_weightgm < 2000 THEN 'Low Weight Product'
ELSE 'Medium Weight Product'
END AS pwt_cat
FROM ecom
) insert overwrite table job8b_op
SELECT cust_city, pwt_cat, COUNT(*) AS count from cte  where pwt_cat='High Weight Product' GROUP BY cust_city, pwt_cat ORDER BY count desc LIMIT 10;
```



```
hive> WITH cte AS (
    > SELECT order_id, cust_id, cust_city, prod_weightgm, CASE WHEN prod_weightg
m > 22000 THEN 'High Weight Product'
    > WHEN prod_weightgm < 2000 THEN 'Low Weight Product'
    > ELSE 'Medium Weight Product'
    > END AS pwt_cat
    > FROM ecom
    > ) insert overwrite table job8b_op
    > SELECT cust_city, pwt_cat, COUNT(*) AS count from cte  where pwt_cat='High
 Weight Product' GROUP BY cust_city, pwt_cat ORDER BY count desc LIMIT 10;
Query ID = cloudera_20230521055757_0f16b2fb-29fb-436c-ae76-8428442943a5
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_1684660565171_0015, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1684660565171_0015/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684660565171_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-21 05:57:13,381 Stage-1 map = 0%,  reduce = 0%
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2
Total MapReduce CPU Time Spent: 5 seconds 390 msec
OK
Time taken: 44.422 seconds
hive> select * from job8a_op;
OK
sao paulo       Low Weight Product      14745
rio de janeiro  Low Weight Product      6057
belo horizonte  Low Weight Product      2511
brasilia        Low Weight Product      1979
curitiba        Low Weight Product      1424
campinas        Low Weight Product      1338
porto alegre    Low Weight Product      1251
salvador        Low Weight Product      1140
guarulhos       Low Weight Product      1041
sao bernardo do campo   Low Weight Product      878
Time taken: 0.04 seconds, Fetched: 10 row(s)
hive> select * from job8b_op;
OK
sao paulo       High Weight Product     102
rio de janeiro  High Weight Product     53
palmas  High Weight Product     16
belo horizonte  High Weight Product     14
niteroi High Weight Product     13
brasilia        High Weight Product     9
campinas        High Weight Product     9
santos  High Weight Product     9
salvador        High Weight Product     9
registro        High Weight Product     8
Time taken: 0.037 seconds, Fetched: 10 row(s)
hive>
```

OBSERVATION: Sao Paulo city buys both Heavy Weight Products and Low Weight Products.

8.1.3-> MySQL Table Created (Client's DB):
Low Weight Products:
CREATE TABLE job8a_lowWt(Customer_City VARCHAR(55), Product_weightCategory VARCHAR(20), No_of_Orders INT);
High Weight Products:
CREATE TABLE job8a_HighWt(Customer_City VARCHAR(55), Product_weightCategory VARCHAR(20), No_of_Orders INT);

8.2-> How much products sold within seller state?

SELECT COALESCE(prodCat, 'TOTAL') AS prodCat, COUNT(order_id) as no_orders FROM ecom WHERE cust_state = seller_state GROUP BY prodCat WITH ROLLUP ;

8.2.1-> Ext Table Creation:

CREATE EXTERNAL TABLE job8c_op(prod_cat string, no_orders INT);

8.2.2-> Transferring o/p to ext table:

insert overwrite table job8c_op SELECT COALESCE(prodCat, 'TOTAL') AS prodCat, COUNT(order_id) as no_orders FROM ecom WHERE cust_state = seller_state
GROUP BY prodCat WITH ROLLUP;

```
hive> CREATE EXTERNAL TABLE job8c_op(prod_cat string, no_orders INT);
OK
Time taken: 0.073 seconds
hive> insert overwrite table job8c_op SELECT COALESCE(prodCat, 'TOTAL') AS prodCat, COUNT(order_id) as no_orders FROM ecom WHERE cust_state = seller_state
    > GROUP BY prodCat WITH ROLLUP;
Query ID = cloudera_20230521072626_26cba286-583e-41f0-925d-37c4b46d825a
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_1684660565171_0035, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1684660565171_0035/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684660565171_0035
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-21 07:26:31,740 Stage-1 map = 0%,  reduce = 0%
```

```
hive> select * from job8c_op;
OK
TOTAL   42052
agro_industry_and_commerce      67
air_conditioning        107
art     96
arts_and_craftmanship   17
audio   146
auto    1655
baby    1124
bed_bath_table  5202
books_general_interest  203
books_imported  37
books_technical 111
cds_dvds_musicals       6
christmas_supplies      49
cine_photo      29
computers       41
computers_accessories   2177
consoles_games  413
construction_tools_construction 392
construction_tools_lights       147
construction_tools_safety       59
cool_stuff      1227
costruction_tools_garden        89
Time taken: 0.712 seconds, Fetched: 71 row(s)
```

8.2.3-> MySQL Table Created (Client's DB):

```
CREATE TABLE job8b(Product_Category VARCHAR(50) , No_of_Orders INT);
```

****

MySQL Table Creation ScreenShots:

****

```
mysql> CREATE TABLE job1(customer_id VARCHAR(33), Spendings Float, Customer_category VARCHAR(10))
    -> ;
Query OK, 0 rows affected (0.03 sec)

mysql> CREATE TABLE job2(Year INT, Month VARCHAR(10), sales FLOAT, sales_trend INT)
    -> ;
Query OK, 0 rows affected (0.01 sec)

mysql> CREATE TABLE job3(hour VARCHAR(6), No_of_Sales INT)
    -> ;
Query OK, 0 rows affected (0.02 sec)

mysql> CREATE TABLE job4a(Product_Category VARCHAR(50), sold_count INT)
    -> ;
Query OK, 0 rows affected (0.01 sec)

mysql> CREATE TABLE job4b(Product_Category VARCHAR(50), Rating INT, rating_count INT)
    -> ;
Query OK, 0 rows affected (0.03 sec)

mysql> CREATE TABLE job4c(Product_ID VARCHAR(33), Product_Category VARCHAR(50), sold_count INT)
    -> ;
Query OK, 0 rows affected (0.00 sec)

mysql> CREATE TABLE job4d_Least10(Product_Category VARCHAR(50), Rating INT, rating_count INT)
    -> ;
Query OK, 0 rows affected (0.04 sec)

mysql> CREATE TABLE job4d_Top10(Product_Category VARCHAR(50), Rating INT, rating_count INT)
    -> ;
Query OK, 0 rows affected (0.01 sec)

mysql> CREATE TABLE job4e(Rating INT, No_of_Orders INT)
    -> ;
Query OK, 0 rows affected (0.01 sec)
```

```
mysql> CREATE TABLE job5a(Payment_Type VARCHAR(20), No_of_Orders INT)
    -> ;
Query OK, 0 rows affected (0.03 sec)

mysql> CREATE TABLE job5b(Payment_Installments INT, No_of_Orders INT)
    -> ;
Query OK, 0 rows affected (0.01 sec)

mysql> CREATE TABLE job6(Customer_State VARCHAR(5), Customer_City VARCHAR(55), No_Of_Customers INT )
    -> ;
Query OK, 0 rows affected (0.03 sec)

mysql> CREATE TABLE job7a(Seller_ID VARCHAR(33), Sales_amt INT)
    -> ;
Query OK, 0 rows affected (0.00 sec)

mysql> CREATE TABLE job7b(Seller_ID VARCHAR(33), Rating INT, rating_count INT)
    -> ;
Query OK, 0 rows affected (0.00 sec)

mysql> CREATE TABLE job8a_lowWt(Customer_City VARCHAR(55), Product_weightCategory VARCHAR(20), No_of_Orders INT)
    -> ;
Query OK, 0 rows affected (0.01 sec)

mysql> CREATE TABLE job8a_HighWt(Customer_City VARCHAR(55), Product_weightCategory VARCHAR(20), No_of_Orders INT)
    -> ;
Query OK, 0 rows affected (0.02 sec)

mysql> CREATE TABLE job8b(Product_Category VARCHAR(50) , No_of_Orders INT)
    -> ;
Query OK, 0 rows affected (0.03 sec)
```

Result tables in MySQL:

```
mysql> show tables;
+---------------------+
| Tables_in_ecommerce |
+---------------------+
| job1                |
| job2                |
| job3                |
| job4a               |
| job4b               |
| job4c               |
| job4d_Least10       |
| job4d_Top10         |
| job4e               |
| job5a               |
| job5b               |
| job6                |
| job7a               |
| job7b               |
| job8a_HighWt        |
| job8a_lowWt         |
| job8b               |
+---------------------+
17 rows in set (0.00 sec)
```

```
mysql> select * from job1 limit 10;;
+----------------------------------+-----------+-------------------+
| customer_id                      | Spendings | Customer_category |
+----------------------------------+-----------+-------------------+
| d2a88d4cd954a41795819308c136a315 |     63.33 | GROUP 5           |
| 028ef41231e80887d8be309e5a28aaf2 |     63.33 | GROUP 5           |
| 744a80b67af618adfecb19c2b77e3618 |     63.33 | GROUP 5           |
| b52dc9b0b38a7c414ffb6da8aec8d27d |     63.33 | GROUP 5           |
| d30a43ef4613415982f5d40a4d0e47bf |     63.31 | GROUP 5           |
| 5667ce9b9fb068ffcc82902f8ad64c63 |     63.31 | GROUP 5           |
| 1addc7d0ea1d5587f4cec79230ef78c5 |     63.31 | GROUP 5           |
| 68f1d888d94100d13b58f401de3bb5b4 |     63.31 | GROUP 5           |
| c42eeec226b4a998f0655d5bd633cbb3 |     63.31 | GROUP 5           |
| 1471f2464edb13ceccc7b52cb1a9def4 |     63.31 | GROUP 5           |
+----------------------------------+-----------+-------------------+
10 rows in set (0.00 sec)
```

```
mysql> select * from job2;
+------+----------+------------+-------------+
| Year | Month    | sales      | sales_trend |
+------+----------+------------+-------------+
| 2018 | April    | 1.15867e+06 |       20474 |
| 2018 | May      |  1.1711e+06 |       12432 |
| 2018 | June     | 1.02328e+06 |     -147827 |
| 2018 | July     | 1.04366e+06 |       20384 |
| 2018 | August   |      976500 |      -67161 |
| 2018 | September |         145 |     -976355 |
| 2018 | April    | 1.15867e+06 |       20474 |
| 2018 | May      |  1.1711e+06 |       12432 |
| 2018 | June     | 1.02328e+06 |     -147827 |
| 2018 | July     | 1.04366e+06 |       20384 |
| 2018 | August   |      976500 |      -67161 |
| 2018 | September |         145 |     -976355 |
| 2016 | September |      165.29 |        NULL |
| 2016 | October  |     58085.9 |       57921 |
| 2016 | December |        10.9 |      -58075 |
| 2017 | January  |      148450 |      148439 |
| 2017 | February |      276258 |      127808 |
| 2017 | March    |      431747 |      155489 |
| 2017 | April    |      421563 |      -10184 |
| 2017 | May      |      595437 |      173874 |
| 2017 | June     |      489714 |     -105723 |
| 2017 | July     |      596520 |      106806 |
| 2017 | August   |      689649 |       93129 |
| 2017 | September |      792980 |      103331 |
| 2017 | October  |      782518 |      -10462 |
| 2017 | November | 1.20964e+06 |      427119 |
| 2017 | December |      829291 |     -380345 |
| 2018 | January  | 1.09532e+06 |      266032 |
| 2018 | February | 1.01098e+06 |      -84344 |
| 2018 | March    |  1.1382e+06 |      127219 |
+------+----------+------------+-------------+
30 rows in set (0.00 sec)
```

```
mysql> select * from job3;
+-------+-------------+
| hour  | No_of_Sales |
+-------+-------------+
| 01 AM |        1324 |
| 4 PM  |        7901 |
| 02 AM |         600 |
| 2 PM  |        7833 |
| 11 AM |        7688 |
| 1 PM  |        7629 |
| 3 PM  |        7558 |
| 10 AM |        7256 |
| 5 PM  |        7193 |
| 06 AM |         565 |
| 03 AM |         320 |
| 04 AM |         252 |
| 05 AM |         214 |
| 8 PM  |        7190 |
| 9 PM  |        7170 |
| 12 PM |        7115 |
| 7 PM  |        6923 |
| 10 PM |        6869 |
| 6 PM  |        6791 |
| 09 AM |        5573 |
| 11 PM |        4853 |
| 08 AM |        3487 |
| 12 AM |        2871 |
| 07 AM |        1405 |
+-------+-------------+
24 rows in set (0.00 sec)
```

```
mysql> select * from job4a;
+-----------------------+-------------+
| Product_Category      | sold_count |
+-----------------------+-------------+
| garden_tools          |        6152 |
| telephony             |        5354 |
| auto                  |        5067 |
| housewares            |        9559 |
| watches_gifts         |        6823 |
| sports_leisure        |       10349 |
| computers_accessories |       10273 |
| bed_bath_table        |       14797 |
| furniture_decor       |       12186 |
| health_beauty         |       11486 |
+-----------------------+-------------+
10 rows in set (0.00 sec)
```

```
mysql> select * from job4b;
+-----------------------+--------+--------------+
| Product_Category      | Rating | rating_count |
+-----------------------+--------+--------------+
| watches_gifts         |      5 |         3454 |
| furniture_decor       |      5 |         4708 |
| bed_bath_table        |      5 |         6179 |
| computers_accessories |      5 |         4351 |
| housewares            |      5 |         4206 |
| health_beauty         |      5 |         6036 |
| sports_leisure        |      5 |         5292 |
| toys                  |      5 |         2623 |
| garden_tools          |      5 |         2571 |
| auto                  |      5 |         2456 |
| telephony             |      5 |         2416 |
| cool_stuff            |      5 |         2332 |
| perfumery             |      5 |         2244 |
+-----------------------+--------+--------------+
13 rows in set (0.00 sec)

mysql> select * from job4c;
+----------------------------------+-----------------------+------------+
| Product_ID                       | Product_Category      | sold_count |
+----------------------------------+-----------------------+------------+
| 53b36df67ebb7c41585e8d54d6772e08 | watches_gifts         |        363 |
| b532349fe46b38fbc7bb3914c1bdae07 | furniture_decor       |        324 |
| 368c6c730842d78016ad823897a372db | garden_tools          |        564 |
| 53759a2ecddad2bb87a079a1f1519f73 | garden_tools          |        563 |
| 389d119b48cf3043d311335e499d9c6b | garden_tools          |        553 |
| 422879e10f46682990de24d770e7f83d | garden_tools          |        826 |
| aca2eb7d00ea1a7b8ebd4e68314663af | furniture_decor       |        650 |
| 99a4788cb24856965c36a24e339b6058 | bed_bath_table        |        588 |
| d1c427060a0f73f6b889a5c7c61f2ac4 | computers_accessories |        383 |
| a62e25e09e05e6faf31d90c6ec1aa3d1 | watches_gifts         |        369 |
+----------------------------------+-----------------------+------------+
10 rows in set (0.00 sec)
```

```
mysql> select * from job4d_Top10;
+----------------------+--------+--------------+
| Product_Category     | Rating | rating_count |
+----------------------+--------+--------------+
| bed_bath_table       |      5 |         6179 |
| health_beauty        |      5 |         6036 |
| sports_leisure       |      5 |         5292 |
| furniture_decor      |      5 |         4708 |
| computers_accessories |      5 |         4351 |
| housewares           |      5 |         4206 |
| watches_gifts        |      5 |         3454 |
| toys                 |      5 |         2623 |
| garden_tools         |      5 |         2571 |
| auto                 |      5 |         2456 |
+----------------------+--------+--------------+
10 rows in set (0.00 sec)

mysql> select * from job4d_Least10;
+-----------------------------------+--------+--------------+
| Product_Category                  | Rating | rating_count |
+-----------------------------------+--------+--------------+
| security_and_services             |      1 |            1 |
| security_and_services             |      4 |            1 |
| books_imported                    |      2 |            1 |
| fashion_childrens_clothes         |      1 |            1 |
| fashion_sport                     |      2 |            1 |
| dvds_blu_ray                      |      3 |            1 |
| furniture_mattress_and_upholstery |      2 |            1 |
| dvds_blu_ray                      |      2 |            1 |
| costruction_tools_tools           |      3 |            1 |
| costruction_tools_tools           |      2 |            1 |
| security_and_services             |      1 |            1 |
| security_and_services             |      4 |            1 |
| books_imported                    |      2 |            1 |
```

```
mysql> select * from job5a;
+--------------+--------------+
| Payment_Type | No_of_Orders |
+--------------+--------------+
| credit_card  |        86011 |
| boleto       |        22692 |
| voucher      |         6211 |
| debit_card   |         1666 |
+--------------+--------------+
4 rows in set (0.00 sec)

mysql> select * from job5b;
+----------------------+--------------+
| Payment_Installments | No_of_Orders |
+----------------------+--------------+
|                 NULL |            3 |
|                    0 |            2 |
|                    1 |        47885 |
|                    2 |        12091 |
|                    3 |        10236 |
|                    4 |         6955 |
|                    5 |         5130 |
|                    6 |         3837 |
|                    7 |         1584 |
|                    8 |         4194 |
|                    9 |          630 |
|                   10 |         5216 |
|                   11 |           22 |
|                   12 |          131 |
```

```
mysql> select * from job6;
+----------------+-----------------------+-----------------+
| Customer_State | Customer_City         | No_Of_Customers |
+----------------+-----------------------+-----------------+
| SP             | sao paulo             |           18412 |
| RJ             | rio de janeiro        |            8123 |
| MG             | belo horizonte        |            3218 |
| SP             | guarulhos             |            1388 |
| SP             | sao bernardo do campo |            1100 |
| DF             | brasilia              |            2452 |
| PR             | curitiba              |            1799 |
| SP             | campinas              |            1722 |
| RS             | porto alegre          |            1640 |
| BA             | salvador              |            1498 |
+----------------+-----------------------+-----------------+
10 rows in set (0.00 sec)
```

```
mysql> select * from job7a;;
+----------------------------------+-----------+
| Seller_ID                        | Sales_amt |
+----------------------------------+-----------+
| 7c67e1448b00f6e969d365cea6b010ab |    308853 |
| 53243585a1d6dc2643021fd1853d8905 |    267642 |
| 4a3ca9315b744ce9f8e9374361493884 |    244215 |
| 4869f7a5dfa277a7dca6462dcf3b52b2 |    244023 |
| da8622b14eb17ae2831f4ac5b9dab84a |    210603 |
| fa1c13f2614d7b5c4749cbc52fecda94 |    205545 |
| 1025f0e2d44d7041d6cf58b6550e0bfa |    197499 |
| 7e93a43ef30c4f03f38b393420bc753a |    184351 |
| 955fee9216a65b617aa5c0531780ce60 |    169398 |
| 1f50f920176fa81dab994f9023523100 |    168425 |
+----------------------------------+-----------+
10 rows in set (0.00 sec)

ERROR:
No query specified

mysql> select * from job7b;
+----------------------------------+--------+--------------+
| Seller_ID                        | Rating | rating_count |
+----------------------------------+--------+--------------+
| 4869f7a5dfa277a7dca6462dcf3b52b2 |      5 |          700 |
| 3d871de0142ce09b7081e2b9d1733cb1 |      5 |          658 |
| 1f50f920176fa81dab994f9023523100 |      5 |         1134 |
| cc419e0650a3c5ba77189a1882b7556a |      5 |         1091 |
| 6560211a19b47992c3666cc44a7e94c0 |      5 |         1070 |
| 4a3ca9315b744ce9f8e9374361493884 |      5 |         1021 |
| da8622b14eb17ae2831f4ac5b9dab84a |      5 |          950 |
| 955fee9216a65b617aa5c0531780ce60 |      5 |          818 |
| 7a67c85e85bb2ce8582c35f2203ad736 |      5 |          753 |
| 1025f0e2d44d7041d6cf58b6550e0bfa |      5 |          751 |
+----------------------------------+--------+--------------+
10 rows in set (0.00 sec)
```

```
mysql> select * from job8a_HighWt;
+----------------------+-----------------------+--------------+
| Customer_City        | Product_weightCategory | No_of_Orders |
+----------------------+-----------------------+--------------+
| guarulhos            | Low Weight Product     |         1041 |
| sao bernardo do campo | Low Weight Product     |          878 |
| sao paulo            | Low Weight Product     |        14745 |
| rio de janeiro       | Low Weight Product     |         6057 |
| belo horizonte       | Low Weight Product     |         2511 |
| brasilia             | Low Weight Product     |         1979 |
| curitiba             | Low Weight Product     |         1424 |
| campinas             | Low Weight Product     |         1338 |
| porto alegre         | Low Weight Product     |         1251 |
| salvador             | Low Weight Product     |         1140 |
+----------------------+-----------------------+--------------+
10 rows in set (0.00 sec)

mysql> select * from job8a_lowWt;
+---------------+-----------------------+--------------+
| Customer_City  | Product_weightCategory | No_of_Orders |
+---------------+-----------------------+--------------+
| sao paulo      | High Weight Product    |          102 |
| rio de janeiro | High Weight Product    |           53 |
| palmas         | High Weight Product    |           16 |
| belo horizonte | High Weight Product    |           14 |
| niteroi        | High Weight Product    |           13 |
| brasilia       | High Weight Product    |            9 |
| campinas       | High Weight Product    |            9 |
| santos         | High Weight Product    |            9 |
| salvador       | High Weight Product    |            9 |
| registro       | High Weight Product    |            8 |
+---------------+-----------------------+--------------+
10 rows in set (0.00 sec)
```

```
mysql> select * from job8b;
+---------------------------------------+-------------+
| Product_Category                      | No_of_Orders |
+---------------------------------------+-------------+
| TOTAL                                 |        42052 |
| agro_industry_and_commerce            |           67 |
| air_conditioning                      |          107 |
| art                                   |           96 |
| arts_and_craftmanship                 |           17 |
| audio                                 |          146 |
| auto                                  |         1655 |
| baby                                  |         1124 |
| bed_bath_table                        |         5202 |
| books_general_interest                |          203 |
| books_imported                        |           37 |
| books_technical                       |          111 |
| cds_dvds_musicals                     |            6 |
| christmas_supplies                    |           49 |
| cine_photo                            |           29 |
| computers                             |           41 |
| computers_accessories                 |         2177 |
| consoles_games                        |          413 |
| construction_tools_construction       |          392 |
| construction_tools_lights             |          147 |
| flowers                               |           11 |
| food                                  |          280 |
| food_drink                            |          108 |
| furniture_bedroom                     |           51 |
| furniture_decor                       |         3202 |
| furniture_living_room                 |          168 |
| furniture_mattress_and_upholstery     |           24 |
| garden_tools                          |         1424 |
| health_beauty                         |         3709 |
| home_appliances                       |          355 |
| home_appliances_2                     |           74 |
```

****

Sqoop Export Commands

****

```
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job1 --export-dir
/user/hive/warehouse/ecommerce/job1_op.txt;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job2 --export-dir
/user/hive/warehouse/ecommerce.db/job2_op --input-null-string '\\N'  --input-null-non-string '\\N';
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job3 --export-dir
/user/hive/warehouse/ecommerce.db/job3_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job4a --export-dir
/user/hive/warehouse/ecommerce.db/job4a_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job4b --export-dir
/user/hive/warehouse/ecommerce.db/job4b_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job4c --export-dir
/user/hive/warehouse/ecommerce.db/job4c_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job4d_Least10 --export-dir
/user/hive/warehouse/ecommerce.db/job4dleast10_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job4d_Top10 --export-dir
/user/hive/warehouse/ecommerce.db/job4dtop10_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job4e --export-dir
/user/hive/warehouse/ecommerce.db/job4e;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job5a --export-dir
/user/hive/warehouse/ecommerce.db/job5a_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job5b --export-dir
/user/hive/warehouse/ecommerce.db/job5b_op --input-null-string '\\N'  --input-null-non-string '\\N'; ;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job6 --export-dir
/user/hive/warehouse/ecommerce.db/job6_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job7a --export-dir
/user/hive/warehouse/ecommerce.db/job7a_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job7b --export-dir
/user/hive/warehouse/ecommerce.db/job7b_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job8a_HighWt --export-dir
/user/hive/warehouse/ecommerce.db/job8a_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job8a_lowWt --export-dir
/user/hive/warehouse/ecommerce.db/job8b_op;
sqoop export --connect jdbc:mysql://localhost:3306/ecommerce --username root --password cloudera --table job8b --export-dir
/user/hive/warehouse/ecommerce.db/job8c_op;
```

****

HDFS to Local File System

```
hdfs dfs -get /user/hive/warehouse/ecommerce/job1_op.txt/000000_0 /home/cloudera/ecommerce/job1.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job2_op/000000_0 /home/cloudera/ecommerce/job2.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job3_op/000000_0 /home/cloudera/ecommerce/job3.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job4a_op/000000_0 /home/cloudera/ecommerce/job4a.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job4b_op/000000_0 /home/cloudera/ecommercejob4b.csv
```

```
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job4c_op/000000_0 /home/cloudera/ecommerce/job4c.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job4dleast10_op/000000_0 /home/cloudera/ecommerce/job4dleast10.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job4dtop10_op/000000_0 /home/cloudera/ecommerce/job4dtop10.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job4e/000000_0 /home/cloudera/ecommerce/job4e.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job5a_op/000000_0 /home/cloudera/ecommerce/job5a.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job5b_op/000000_0 /home/cloudera/ecommerce/job5b.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job6_op/000000_0 /home/cloudera/ecommerce/job6.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job7a_op/000000_0 /home/cloudera/ecommerce/job7a.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job7b_op/000000_0 /home/cloudera/ecommerce/job7b.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job8a_op/000000_0 /home/cloudera/ecommerce/job8a.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job8b_op/000000_0 /home/cloudera/ecommerce/job8b.csv
hdfs dfs -get /user/hive/warehouse/ecommerce.db/job8c_op/000000_0 /home/cloudera/ecommerce/job8c.csv
```

```
cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -get /user/hive/warehouse/ecommerce.db/job5a_op/000000_0 /home/cloudera/ecommerce/job5a.csv
cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -get /user/hive/warehouse/ecommerce.db/job5b_op/000000_0 /home/cloudera/ecommerce/job5b.csv
cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -get /user/hive/warehouse/ecommerce.db/job6_op/000000_0 /home/cloudera/ecommerce/job6.csv
cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -get /user/hive/warehouse/ecommerce.db/job7a_op/000000_0 /home/cloudera/ecommerce/job7a.csv
cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -get /user/hive/warehouse/ecommerce.db/job7b_op/000000_0 /home/cloudera/ecommerce/job7b.csv
cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -get /user/hive/warehouse/ecommerce.db/job8a_op/000000_0 /home/cloudera/ecommerce/job8a.csv
cloudera@quickstart ~]$
cloudera@quickstart ~]$ hdfs dfs -get /user/hive/warehouse/ecommerce.db/job8b_op/000000_0 /home/cloudera/ecommerce/job8b.csv
```