

Walmart Data Analysis

Section A-

Identify the relation between different columns of 5 tables (product_master, sales_order_detail, sales_order, salesman_master, client_master) in csv format. Create the structure of these tables in HIVE and load the data from the respective files. Assist the manager of the Marketing and Sales Department by executing the following queries.

Step 1: Uploading all the data files on the virtual machines and then uploading them on Hadoop file system

```
SSH-in-browser
UPLOAD FILE
DOWNLOAD FILE
!
[]
⚙️

rakshandahedawoo@cluster-b00d-m:~$ ls
Departmentdata.json      client_master.csv      sales_order.csv        salesman_master.csv
LogisticsandOperations.csv  product_master.csv    sales_order_details.csv
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -mkdir /user/rakshandahedawoo
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -ls /user/
Found 12 items
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/dataproc
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/hbase
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/hdfs
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/hive
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/mapred
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/pig
drwxr-xr-x - rakshandahedawoo hadoop          0 2024-02-16 04:06 /user/rakshandahedawoo
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/solr
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/spark
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/yarn
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/zeppelin
drwxrwxrwt - hdfs          hadoop          0 2024-02-16 04:00 /user/zookeeper
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -put Departmentdata.json
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -put client_master.csv
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -put sales_order.csv
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -put salesman_master.csv
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -put LogisticsandOperations.csv
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -put product_master.csv
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -put sales_order_details.csv
rakshandahedawoo@cluster-b00d-m:~$ hadoop fs -ls /user/rakshandahedawoo/
Found 7 items
-rw-r--r-- 2 rakshandahedawoo hadoop      3518 2024-02-16 04:06 /user/rakshandahedawoo/Departmentdata.json
-rw-r--r-- 2 rakshandahedawoo hadoop     1327 2024-02-16 04:07 /user/rakshandahedawoo/LogisticsandOperation
s.csv
-rw-r--r-- 2 rakshandahedawoo hadoop      170 2024-02-16 04:07 /user/rakshandahedawoo/client_master.csv
-rw-r--r-- 2 rakshandahedawoo hadoop      338 2024-02-16 04:07 /user/rakshandahedawoo/product_master.csv
-rw-r--r-- 2 rakshandahedawoo hadoop      325 2024-02-16 04:07 /user/rakshandahedawoo/sales_order.csv
-rw-r--r-- 2 rakshandahedawoo hadoop      225 2024-02-16 04:08 /user/rakshandahedawoo/sales_order_details.c
sv
-rw-r--r-- 2 rakshandahedawoo hadoop       68 2024-02-16 04:07 /user/rakshandahedawoo/salesman_master.csv
rakshandahedawoo@cluster-b00d-m:~$ hive
```

Step 2: Creating a database Walmart in Hive and then creating all the tables and loading the data in those tables from the datafiles.

```

hive> show databases;
OK
default
Time taken: 0.582 seconds, Fetched: 1 row(s)
hive> create database walmart;
OK
Time taken: 0.294 seconds
hive> show databases;
OK
default
walmart
Time taken: 0.039 seconds, Fetched: 2 row(s)
hive> create table Client(Clientno String, Name String, State String)ROW FORMAT DELIMITED FIELDS TERMINATED BY
',' TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 0.496 seconds
hive> load data inpath '/user/rakshandahedawoo/client_master.csv' into table Client;
Loading data to table default.client
OK
Time taken: 0.599 seconds
hive> Select * from Client;
OK
1      Ivan Bayross      New York
2      Ven Simmons      California
3      Pramada Jaguste   New York
4      Basu Navindgi     New York
5      Ravi Sreedharan   Texas
6      Mary Sen          New York
Time taken: 2.234 seconds, Fetched: 6 row(s)

```

```

hive> create table LogisticsOperations(Order_Date Date, Order_Quantity INT, Product_Container String, Sales INT
, Ship_Mode String)ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES("skip.header.line.count"="1", "date.format" = "yyyy-mm-dd");
OK
Time taken: 0.129 seconds
hive> load data inpath '/user/rakshandahedawoo/LogisticsandOperations.csv' into table LogisticsOperations;
Loading data to table default.logisticsoperations
OK
Time taken: 0.37 seconds
hive> Select * from LogisticsOperations;
OK
2007-01-27      31      Medium Box      6567      Express Air
2007-01-28      39      Large Box      1780      Regular Air
2007-01-29      15      Jumbo Drum      578      Delivery Truck
2007-01-30      30      Small Pack      611      Regular Air
2007-01-31      10      Small Box      517      Regular Air
2007-02-01      41      Small Box      283      Regular Air
2007-02-02      42      Jumbo Drum      7373      Delivery Truck
2007-02-04      15      Jumbo Box      9572      Delivery Truck
2007-02-06      6      Medium Box      463      Regular Air
2007-02-09      16      Small Pack      142      Regular Air
2007-03-01      29      Small Box      3212      Regular Air
2007-03-02      46      Small Box      1603      Express Air
2007-03-03      23      Small Pack      214      Express Air
2007-03-04      18      Small Box      2253      Regular Air
2007-03-05      38      Small Box      1498      Express Air
2007-03-02      25      Small Pack      709      Regular Air
2007-03-03      8      Small Box      351      Regular Air
2007-03-04      37      Small Box      512      Regular Air
2007-03-05      15      Jumbo Drum      1044      Delivery Truck
2007-03-06      10      Small Box      34      Regular Air

```

```

hive> create table Product(ProductNo String, Description String, ProfitPercent String, UOM String, QTYonHand IN
T, SellPrice INT, CostPrice INT)ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES("skip.header.line.c
ount"="1");
OK
Time taken: 0.116 seconds
hive> load data inpath '/user/rakshandahedawoo/product_master.csv' into table Product;
Loading data to table default.product
OK
Time taken: 0.286 seconds
hive> Select * from Product;
OK
11      TV      5      Piece      100      525      500
12      Laptop      6      Piece      10      1200      1128
13      Mouse      5      Piece      20      20      18
14      Monitor      5      Piece      100      52      50
15      Keyboard      2      Piece      10      40      32
16      HardDisk      2.5      Piece      10      200      180
17      Mobile      4      Piece      10      800      700
18      PenDrive      5      Piece      10      10      9
19      Headphone      5      Piece      2      20      18
Time taken: 0.306 seconds, Fetched: 9 row(s)
hive> create table SalesOrder(SOrderNo String, SOrderDate Date, ClientNo String, SalesmanNo String, DelyDate Da
te, OrderStatus String)ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES("skip.header.line.count"="1"
, "date.format" = "yyyy-mm-dd");
OK
Time taken: 0.099 seconds
hive> load data inpath '/user/rakshandahedawoo/sales_order.csv' into table SalesOrder;
Loading data to table default.salesorder
OK
Time taken: 0.319 seconds
hive> Select * From SalesOrder;
OK
2001      2021-01-12      1      101      2021-01-20      In Process
2002      2021-01-25      2      102      2021-01-27      Cancelled
2003      2021-01-18      3      103      2021-02-20      Fulfilled
2004      2021-04-03      1      104      2021-04-07      Fulfilled

```

```

hive> create table SalesOrderDetails(SOrderNo String, ProductNo String, QtyOrdered INT, QtyDisp INT)ROW FORMAT
DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 0.099 seconds
hive> load data inpath '/user/rakshandahedawoo/sales_order_details.csv' into table SalesOrderDetails;
Loading data to table default.salesorderdetails
OK
Time taken: 0.306 seconds
hive> Select * From SalesOrderDetails;
OK
2001    11      4      4
2001    12      2      1
2001    13      2      1
2002    14     10      0
2003    15      3      3
2003    14      3      1
2003    16     10     10
2003    17      4      4
2004    13      2      2
2004    18      1      1
2005    16      1      0
2005    17      1      0
2005    18     10      5
2006    12      5      3
Time taken: 0.314 seconds, Fetched: 14 row(s)
hive> create table Salesman(SalesmanNo String, SalesmanName String)ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 0.108 seconds
hive> load data inpath '/user/rakshandahedawoo/salesman_master.csv' into table Salesman;
Loading data to table default.salesman
OK
Time taken: 0.305 seconds

```

```

hive> Select * from Salesman;
OK
101      Kia
102      Manish
103      Remo
104      Angelina
Time taken: 0.148 seconds, Fetched: 4 row(s)
hive> 

```

Code:

Client Table:

```

create table Client(Clientno String, Name String, State String)ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',' TBLPROPERTIES("skip.header.line.count"="1");
create table Client(Clientno String, Name String, State String)ROW FORMAT DELIMITED
FIELDS TERMINATED BY ',' TBLPROPERTIES("skip.header.line.count"="1");

```

load data inpath '/user/rakshandahedawoo/client_master.csv' into table Client;

LogisticsOperations Table:

```

create table LogisticsOperations(Order_Date Date, Order_Quantity INT, Product_Container
String, Sales INT, Ship_Mode String)ROW FORMAT DELIMITED FIELDS TERMINATED
BY ',' TBLPROPERTIES("skip.header.line.count"="1", "date.format" = "yyyy-mm-dd");

```

load data inpath '/user/rakshandahedawoo/LogisticsandOperations.csv' into table
LogisticsOperations;

Product Table:

```

create table Product(ProductNo String, Description String, ProfitPercent String, UOM String,
QTYonHand INT, SellPrice INT, CostPrice INT)ROW FORMAT DELIMITED FIELDS
TERMINATED BY ',' TBLPROPERTIES("skip.header.line.count"="1");

```

load data inpath '/user/rakshandahedawoo/product_master.csv' into table Product;

SalesOrder Table:

create table SalesOrder(SOrderNo String, SOrderDate Date, ClientNo String, SalesmanNo String, DelyDate Date, OrderStatus String)ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES(("skip.header.line.count"="1", "date.format" = "yyyy-mm-dd");

load data inpath '/user/rakshandahedawoo/sales_order.csv' into table SalesOrder;

SalesOrderDetails Table:

create table SalesOrderDetails(SOrderNo String, ProductNo String, QtyOrdered INT, QtyDisp INT)ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES(("skip.header.line.count"="1");

load data inpath '/user/rakshandahedawoo/sales_order_details.csv' into table SalesOrderDetails;

Salesman Table:

create table Salesman(SalesmanNo String, SalesmanName String)ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' TBLPROPERTIES(("skip.header.line.count"="1");

load data inpath '/user/rakshandahedawoo/salesman_master.csv' into table Salesman;

Queries:

1. Display the client name and product name involved in the order '2001'

Code:

Select c.Name, p.Description

From Client c

INNER JOIN SalesOrder so ON so.ClientNo = c.ClientNo

INNER JOIN SalesOrderDetails sod ON sod.SOrderNo = so.SOrderNo

INNER JOIN Product p ON p.ProductNo = sod.ProductNo

Where so.SOrderNo='2001';

```
hive> Select c.Name, p.Description From Client c INNER JOIN SalesOrder so ON so.ClientNo = c.C
lientNo INNER JOIN SalesOrderDetails sod ON sod.SOrderNo = so.SOrderNo INNER JOIN Product p ON
p.ProductNo = sod.ProductNo Where so.SOrderNo='2001';
No Stats for walmart@client, Columns: name, clientno
No Stats for walmart@salesorder, Columns: sorderno, clientno
No Stats for walmart@salesorderdetails, Columns: sorderno, productno
No Stats for walmart@product, Columns: description, productno
Warning: Map Join MAPJOIN[34][bigTable=2] in task 'Map 3' is a cross product
Query ID = rakshandahedawoo_20240216043225_b6f309aa-545c-4262-84d4-4db2231f8eea
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708056013283_0005)

-----
VERTICES      MODE           STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container  SUCCEEDED   1         1         0         0         0         0
Map 2 ..... container  SUCCEEDED   1         1         0         0         0         0
Map 4 ..... container  SUCCEEDED   1         1         0         0         0         0
Map 3 ..... container  SUCCEEDED   1         1         0         0         0         0
-----
VERTICES: 04/04  [=====>>>] 100% ELAPSED TIME: 15.80 s
-----
OK
Ivan Bayross      Mouse
Ivan Bayross      Laptop
Ivan Bayross      TV
Time taken: 36.051 seconds, Fetched: 3 row(s)
hive>
```

Output:

2. Display the products and name of salesman who sold to 'Ven Simmons'.

Code:

```
Select s.SalesmanName, p.Description
From Product p
INNER JOIN SalesOrderDetails sod ON sod.ProductNo = p.ProductNo
INNER JOIN SalesOrder so ON so.SOrderNo = sod.SOrderNo
INNER JOIN Client c ON c.ClientNo = so.ClientNo
Where c.Name = 'Ven Simmons';
```

Output:

```
hive> Select s.SalesmanName, p.Description From Product p INNER JOIN SalesOrderDetails sod ON
sod.ProductNo = p.ProductNo INNER JOIN SalesOrder so ON so.SOrderNo = sod.SOrderNo INNER JOIN
Salesman s ON s.SalesmanNo = so.SalesmanNo INNER JOIN Client c ON c.ClientNo = so.ClientNo Whe
re c.Name = 'Ven Simmons';
No Stats for walmart@product, Columns: description, productno
No Stats for walmart@salesorderdetails, Columns: sorderno, productno
No Stats for walmart@salesorder, Columns: salesmanno, sorderno, clientno
No Stats for walmart@salesman, Columns: salesmanno, salesmanname
No Stats for walmart@client, Columns: name, clientno
Query ID = rakshandahedawoo_20240216043823_49a127e4-adcd-445a-a533-clfbc0bd203e
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1708056013283_0006)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 2 ..... container  SUCCEEDED    1          1          0          0          0          0
Map 3 ..... container  SUCCEEDED    1          1          0          0          0          0
Map 5 ..... container  SUCCEEDED    1          1          0          0          0          0
Map 4 ..... container  SUCCEEDED    1          1          0          0          0          0
Map 1 ..... container  SUCCEEDED    1          1          0          0          0          0
-----
VERTICES: 05/05 [=====>>>] 100% ELAPSED TIME: 18.86 s
-----
OK
Manish Monitor
Time taken: 31.46 seconds, Fetched: 1 row(s)
hive> █
```

3. Display the salesman's name and delivery date of the customers who reside in New York.

Code:

```
Select s.SalesmanName, so.DelyDate
From Client c
INNER JOIN SalesOrder so ON so.ClientNo = s.ClientNo
INNER JOIN Salesman s ON s.SalesmanNo = so.SalesmanNo
Where c.State = 'New York';
```

Output:

```
hive> Select s.SalesmanName, so.DelyDate From Client c INNER JOIN SalesOrder so ON so.ClientNo
= c.ClientNo INNER JOIN Salesman s ON s.SalesmanNo = so.SalesManNo Where c.State='New York'
> ;
No Stats for walmart@client, Columns: clientno, state
No Stats for walmart@salesorder, Columns: delydate, salesmanno, clientno
No Stats for walmart@salesman, Columns: salesmanno, salesmanname
Query ID = rakshandahedawoo_20240216044119_9dc5fc28-fc26-4e43-80da-40bdd3073788
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708056013283_0006)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 3 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 03/03  [=====>>>] 100%  ELAPSED TIME: 13.34 s
-----
OK
Angelina      2021-04-07
Remo         2021-02-20
Kia          2021-01-20
Manish       2021-05-22
Time taken: 14.542 seconds, Fetched: 4 row(s)
hive> 
```

4. Display names of product sold by 'Kia'.

Code:

Select p.Description

From Product p

INNER JOIN SalesOrderDetails sod ON sod.ProductNo = p.ProductNo

INNER JOIN SalesOrder so ON so.SOrderNo = sod.SOrderNo

INNER JOIN Salesman s ON s.SalesmanNo = so.SalesmanNo

Where SalesmanName = 'Kia';

Output:

```
hive> Select p.Description From Product p INNER JOIN SalesOrderDetails sod ON sod.ProductNo =
p.ProductNo INNER JOIN SalesOrder so ON so.SOrderNo=sod.SOrderNo INNER JOIN Salesman s ON s.Sa
lesmanNo = so.SalesmanNo Where SalesmanName = 'Kia';
No Stats for walmart@product, Columns: description, productno
No Stats for walmart@salesorderdetails, Columns: sorderno, productno
No Stats for walmart@salesorder, Columns: salesmanno, sorderno
No Stats for walmart@salesman, Columns: salesmanno, salesmanname
Query ID = rakshandahedawoo_20240216044353_093171dd-01ef-41b9-b7b8-186c65d2099f
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708056013283_0006)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 4 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 3 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 04/04  [=====>>>] 100%  ELAPSED TIME: 14.60 s
-----
OK
TV
Mouse
Laptop
Time taken: 16.018 seconds, Fetched: 3 row(s)
hive> 
```

5. Display the state of clients whom 'Kia' has sold monitors

Code:

```
Select c.State
From Product p
INNER JOIN SalesOrderDetails sod ON sod.ProductNo = p.ProductNo
INNER JOIN SalesOrder so ON so.SOrderNo = sod.SOrderNo
INNER JOIN Salesman s ON s.SalesmanNo = so.SalesmanNo
INNER JOIN Client c ON c.ClientNo = so.ClientNo
Where SalesmanName = 'Kia' and p.Description='Monitor';
```

Output:

```
hive> Select c.State From Product p INNER JOIN SalesOrderDetails sod ON sod.ProductNo = p.Prod
uctNo INNER JOIN SalesOrder so ON so.SOrderNo=sod.SOrderNo INNER JOIN Salesman s ON s.Salesman
No = so.SalesmanNo INNER JOIN Client c ON c.ClientNo = so.ClientNo Where SalesmanName = 'Kia'
and p.Description = 'Monitor';
No Stats for walmart@product, Columns: description, productno
No Stats for walmart@salesorderdetails, Columns: sorderno, productno
No Stats for walmart@salesorder, Columns: salesmanno, sorderno, clientno
No Stats for walmart@salesman, Columns: salesmanno, salesmanname
No Stats for walmart@client, Columns: clientno, state
Query ID = rakshandahedawoo_20240216044953_61ffflc7-d2c3-4612-a1b8-1b45b3b5c893
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1708056013283_0007)

-----
      VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 2 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 3 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 5 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 4 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 05/05 [=====>>] 100% ELAPSED TIME: 15.63 s
-----
OK
Time taken: 27.487 seconds
hive> 
```

6. Display the total quantity purchased by different clients (Individual information of all clients)

Code:

```
Select c.ClientNo, c.Name, c.State, SUM(sod.QtyOrdered) AS Total_Quantity
FROM Client c
INNER JOIN SalesOrder so ON so.ClientNo = c.ClientNo
INNER JOIN SalesOrderDetails sod ON sod.SOrderNo = so.SOrderNo
Group By c.clientNo, c.Name, c.State;
```

Output:

```
hive> Select c.ClientNo, c.Name, c.State, SUM(sod.QtyOrdered) AS Total_Quantity FROM Client c
INNER JOIN SalesOrder so ON so.ClientNo = c.ClientNo INNER JOIN SalesOrderDetails sod ON sod.S
OrderNo = so.SOrderNo Group By c.ClientNo, c.Name, c.State;
No Stats for walmart@client, Columns: name, clientno, state
No Stats for walmart@salesorder, Columns: sorderno, clientno
No Stats for walmart@salesorderdetails, Columns: sorderno, qtyordered
Query ID = rakshandahedawoo_20240216045718_d3115384-ff78-4e43-9553-134c6bd4ba0f
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1708056013283_0008)

-----
      VERTICES      MODE      STATUS      TOTAL      COMPLETED      RUNNING      PENDING      FAILED      KILLED
-----
Map 3 ..... container      SUCCEEDED      1              1              0              0              0              0
Map 4 ..... container      SUCCEEDED      1              1              0              0              0              0
Map 1 ..... container      SUCCEEDED      1              1              0              0              0              0
Reducer 2 ..... container      SUCCEEDED      1              1              0              0              0              0
-----
VERTICES: 04/04  [=====>>>] 100%  ELAPSED TIME: 16.19 s
-----
OK
1      Ivan Bayross      New York      11
2      Ven Simmons      California      10
3      Pramada Jaguste      New York      20
4      Basu Navindgi      New York      12
5      Ravi Sreedharan      Texas      5
Time taken: 28.366 seconds, Fetched: 5 row(s)
hive> 
```

7. Display the total sales amount received by selling different products. (Individual information of all products)

Code:

```
Select p.*, SUM(sod.QtyOrdered)*(p.SellPrice)
From Product p
INNER JOIN SalesOrderDetails sod.ProductNo = p.ProductNo
Group By p.ProductNo, p.Description, p.ProditPercent, p.UOM, p.QTYonHand, p.SellPrice,
p.CostPrice;
```


Output:

```
hive> Select p.*, SUM(sod.QtyOrdered)*(p.SellPrice) from Product p INNER JOIN SalesOrderDetail
s sod ON sod.ProductNo = p.ProductNo Group by p.ProductNo,p.Description, p.ProfitPercent, P.UO
M, P.QTYonHand, p.SellPrice, p.CostPrice;
Query ID = rakshandahedawoo_20240216051433_58161b96-8891-42c8-829b-53f653acbd55
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708056013283_0009)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 3 ..... container  SUCCEEDED    1         1         0         0         0         0
Map 1 ..... container  SUCCEEDED    1         1         0         0         0         0
Reducer 2 ..... container  SUCCEEDED    1         1         0         0         0         0
-----
VERTICES: 03/03  [=====>>>] 100% ELAPSED TIME: 14.98 s
-----
OK
11      TV      5      Piece  100    525    500    2100
12      Laptop  6      Piece  10    1200   1128   8400
13      Mouse   5      Piece  20     20     18     80
14      Monitor 5      Piece  100    52     50    676
15      Keyboard 2      Piece  10     40     32    120
16      HardDisk 2.5    Piece  10    200    180   2200
17      Mobile  4      Piece  10    800    700   4000
18      PenDrive 5      Piece  10     10     9     110
Time taken: 15.505 seconds, Fetched: 8 row(s)
hive> 
```

Section B-

Assist the manager of the Operations Department by executing the following queries considering the attached file “LogisticsandOperations.csv”. Create the structure of table and load data from the file in HIVE

Queries:

1. Display the rank of different orders on the basis of sales for different ship modes.
Display all the fields and records.

Code:

```
Select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode,
RANK() OVER(PARTITION BY Ship_Mode ORDER BY Sales)
FROM LogisticsOperations;
```

Output:

```
hive> select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode, RANK() OVER (PARTITION BY Ship_Mode order by Sales) FROM LogisticsOperations;
Query ID = rakshandahedawoo_20240216052450_fef287c6-15c9-4d2f-8751-370059f9df4f
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1708056013283_0010)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 02/02 [=====>>>] 100% ELAPSED TIME: 8.56 s
```

```
OK
2007-03-10      1      Jumbo Box      226      Delivery Truck      1
2007-01-29     15      Jumbo Drum      578      Delivery Truck      2
2007-03-05     15      Jumbo Drum     1044      Delivery Truck      3
2007-03-13     14      Jumbo Drum     4566      Delivery Truck      4
2007-02-02     42      Jumbo Drum     7373      Delivery Truck      5
2007-02-04     15      Jumbo Box      9572      Delivery Truck      6
2007-03-03     23      Small Pack      214      Express Air          1
2007-03-09     47      Small Pack      561      Express Air          2
2007-03-05     38      Small Box      1498      Express Air          3
2007-03-02     46      Small Box      1603      Express Air          4
2007-03-15     36      Small Box      2554      Express Air          5
2007-03-07      1      Large Box      4663      Express Air          6
2007-01-27     31      Medium Box     6567      Express Air          7
2007-03-12      1      Wrap Bag        14      Regular Air          1
2007-03-06     10      Small Box        34      Regular Air          2
2007-03-11     34      Small Box        88      Regular Air          3
2007-02-09     16      Small Pack      142      Regular Air          4
2007-02-01     41      Small Box      283      Regular Air          5
2007-03-03      8      Small Box      351      Regular Air          6
2007-02-06      6      Medium Box      463      Regular Air          7
2007-03-04     37      Small Box      512      Regular Air          8
2007-01-31     10      Small Box      517      Regular Air          9
2007-01-30     30      Small Pack      611      Regular Air         10
2007-03-02     25      Small Pack      709      Regular Air         11
```

```
2007-03-10      1      Jumbo Box      226      Delivery Truck      1
2007-01-29     15      Jumbo Drum      578      Delivery Truck      2
2007-03-05     15      Jumbo Drum     1044      Delivery Truck      3
2007-03-13     14      Jumbo Drum     4566      Delivery Truck      4
2007-02-02     42      Jumbo Drum     7373      Delivery Truck      5
2007-02-04     15      Jumbo Box      9572      Delivery Truck      6
2007-03-03     23      Small Pack      214      Express Air          1
2007-03-09     47      Small Pack      561      Express Air          2
2007-03-05     38      Small Box      1498      Express Air          3
2007-03-02     46      Small Box      1603      Express Air          4
2007-03-15     36      Small Box      2554      Express Air          5
2007-03-07      1      Large Box      4663      Express Air          6
2007-01-27     31      Medium Box     6567      Express Air          7
2007-03-12      1      Wrap Bag        14      Regular Air          1
2007-03-06     10      Small Box        34      Regular Air          2
2007-03-11     34      Small Box        88      Regular Air          3
2007-02-09     16      Small Pack      142      Regular Air          4
2007-02-01     41      Small Box      283      Regular Air          5
2007-03-03      8      Small Box      351      Regular Air          6
2007-02-06      6      Medium Box      463      Regular Air          7
2007-03-04     37      Small Box      512      Regular Air          8
2007-01-31     10      Small Box      517      Regular Air          9
2007-01-30     30      Small Pack      611      Regular Air         10
2007-03-02     25      Small Pack      709      Regular Air         11
2007-03-08     29      Small Box      891      Regular Air         12
2007-03-14     10      Small Box     1082      Regular Air         13
2007-01-28     39      Large Box     1780      Regular Air         14
2007-03-04     18      Small Box     2253      Regular Air         15
2007-03-01     29      Small Box     3212      Regular Air         16
2007-03-16     29      Jumbo Drum     3212      Regular Air         16
Time taken: 19.456 seconds, Fetched: 30 row(s)
hive> 
```

2. Display the rank of different orders on the basis of order quantity for different containers. The rank should be in the descending order. If the quantity is same, the next rank should not be skipped. Display all the fields and records.

Code:

```
Select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode,
DENSE_RANK() OVER (PARTITION BY Product_Container Order By Order_Quantity desc)
From LogisticsOperations;
```

Output:

```
hive> select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode, DENSE_RANK() OVE
R (PARTITION BY Product_Container order by Order_Quantity desc) FROM LogisticsOperations;
Query ID = rakshandahedawoo_20240216053425_4c0963cd-1e3b-4577-ae62-a23e7b994e81
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1708056013283_0011)
```

VERTICES	MODE	STATUS	TOTAL	COMPLETED	RUNNING	PENDING	FAILED	KILLED
Map 1	container	SUCCEEDED	1	1	0	0	0	0
Reducer 2	container	SUCCEEDED	1	1	0	0	0	0

```
VERTICES: 02/02 [ 100% ] ELAPSED TIME: 9.85 s
```

Order_Date	Order_Quantity	Product_Container	Sales	Ship_Mode
2007-02-04	15	Jumbo Box	9572	Delivery Truck
2007-03-10	1	Jumbo Box	226	Delivery Truck
2007-02-02	42	Jumbo Drum	7373	Delivery Truck
2007-03-16	29	Jumbo Drum	3212	Regular Air
2007-01-29	15	Jumbo Drum	578	Delivery Truck
2007-03-05	15	Jumbo Drum	1044	Delivery Truck
2007-03-13	14	Jumbo Drum	4566	Delivery Truck
2007-01-28	39	Large Box	1780	Regular Air
2007-03-07	1	Large Box	4663	Express Air
2007-01-27	31	Medium Box	6567	Express Air
2007-02-06	6	Medium Box	463	Regular Air
2007-03-02	46	Small Box	1603	Express Air
2007-02-01	41	Small Box	283	Regular Air
2007-03-05	38	Small Box	1498	Express Air
2007-03-04	37	Small Box	512	Regular Air
2007-03-15	36	Small Box	2554	Express Air
2007-03-11	34	Small Box	88	Regular Air
2007-03-01	29	Small Box	3212	Regular Air
2007-03-08	29	Small Box	891	Regular Air
2007-03-04	18	Small Box	2253	Regular Air
2007-03-06	10	Small Box	34	Regular Air
2007-01-31	10	Small Box	517	Regular Air
2007-03-14	10	Small Box	1082	Regular Air
2007-03-03	8	Small Box	351	Regular Air

Order_Date	Order_Quantity	Product_Container	Sales	Ship_Mode
2007-02-04	15	Jumbo Box	9572	Delivery Truck
2007-03-10	1	Jumbo Box	226	Delivery Truck
2007-02-02	42	Jumbo Drum	7373	Delivery Truck
2007-03-16	29	Jumbo Drum	3212	Regular Air
2007-01-29	15	Jumbo Drum	578	Delivery Truck
2007-03-05	15	Jumbo Drum	1044	Delivery Truck
2007-03-13	14	Jumbo Drum	4566	Delivery Truck
2007-01-28	39	Large Box	1780	Regular Air
2007-03-07	1	Large Box	4663	Express Air
2007-01-27	31	Medium Box	6567	Express Air
2007-02-06	6	Medium Box	463	Regular Air
2007-03-02	46	Small Box	1603	Express Air
2007-02-01	41	Small Box	283	Regular Air
2007-03-05	38	Small Box	1498	Express Air
2007-03-04	37	Small Box	512	Regular Air
2007-03-15	36	Small Box	2554	Express Air
2007-03-11	34	Small Box	88	Regular Air
2007-03-01	29	Small Box	3212	Regular Air
2007-03-08	29	Small Box	891	Regular Air
2007-03-04	18	Small Box	2253	Regular Air
2007-03-06	10	Small Box	34	Regular Air
2007-01-31	10	Small Box	517	Regular Air
2007-03-14	10	Small Box	1082	Regular Air
2007-03-03	8	Small Box	351	Regular Air
2007-03-09	47	Small Pack	561	Express Air
2007-01-30	30	Small Pack	611	Regular Air
2007-03-02	25	Small Pack	709	Regular Air
2007-03-03	23	Small Pack	214	Express Air
2007-02-09	16	Small Pack	142	Regular Air
2007-03-12	1	Wrap Bag	14	Regular Air

```
Time taken: 19.761 seconds, Fetched: 30 row(s)
hive>
```

3. Add a column named as “Remarks”. The remarks should be “Very High” if the quantity is greater than 35, “High” if it is greater than 30 and less than 35, “Average” if it is greater than 25 and less than 30, “Low” if it is less than 25 and greater than 15 else “Very Low”. Display all the fields and records.

Code:

```
Select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode,
CASE WHEN Order_Quantity >=35 THEN 'Very High'
      WHEN Order_Quantity >=30 and Order_Quantity<35 THEN 'High'
      WHEN Order_Quantity >=25 and Order_Quantity<30 THEN 'Average'
      WHEN Order_Quantity >=15 and Order_Quantity<25 THEN 'Low'
      WHEN Order_Quantity >=35 THEN 'Very Low'
From LogisticsOperations;
```

Output:

```
hive> select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode, CASE
Display all 634 possibilities? (y or n)
hive> select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode, CASE WHEN Order_
Quantity >= 35 THEN 'Very High' WHEN (Order_Quantity >= 30 and Order_Quantity < 35) THEN 'High
' WHEN (Order_Quantity >= 25 and Order_Quantity < 30) THEN 'Average' WHEN (Order_Quantity >= 1
5 and Order_Quantity < 25) THEN 'Low' ELSE 'Very Low' END as Remarks FROM LogisticsOperations;

Query ID = rakshandahedawoo_20240216054106_bdb68d6d-1780-4276-8a81-116ea8ac9478
Total jobs = 1
Launching Job 1 out of 1
Tez session was closed. Reopening...
Session re-established.
Session re-established.
Status: Running (Executing on YARN cluster with App id application_1708056013283_0012)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01 [=====>>] 100% ELAPSED TIME: 7.83 s
-----
OK
2007-01-27      31      Medium Box      6567      Express Air      High
2007-01-28      39      Large Box      1780      Regular Air      Very High
2007-01-29      15      Jumbo Drum      578      Delivery Truck      Low
2007-01-30      30      Small Pack      611      Regular Air      High
2007-01-31      10      Small Box      517      Regular Air      Very Low
2007-02-01      41      Small Box      283      Regular Air      Very High
2007-02-02      42      Jumbo Drum      7373      Delivery Truck      Very High
2007-02-04      15      Jumbo Box      9572      Delivery Truck      Low
2007-02-06      6      Medium Box      463      Regular Air      Very Low
2007-02-09      16      Small Pack      142      Regular Air      Low
2007-03-01      29      Small Box      3212      Regular Air      Average
2007-03-02      46      Small Box      1603      Express Air      Very High
2007-03-03      23      Small Pack      214      Express Air      Low
2007-03-04      18      Small Box      2253      Regular Air      Low
2007-03-05      38      Small Box      1498      Express Air      Very High
2007-03-02      25      Small Pack      709      Regular Air      Average
2007-03-03      8      Small Box      351      Regular Air      Very Low
2007-03-04      37      Small Box      512      Regular Air      Very High
2007-03-05      15      Jumbo Drum      1044      Delivery Truck      Low
2007-03-06      10      Small Box      34      Regular Air      Very Low
```

```

2007-01-27      31      Medium Box      6567      Express Air      High
2007-01-28      39      Large Box      1780      Regular Air      Very High
2007-01-29      15      Jumbo Drum      578      Delivery Truck      Low
2007-01-30      30      Small Pack      611      Regular Air      High
2007-01-31      10      Small Box      517      Regular Air      Very Low
2007-02-01      41      Small Box      283      Regular Air      Very High
2007-02-02      42      Jumbo Drum      7373      Delivery Truck      Very High
2007-02-04      15      Jumbo Box      9572      Delivery Truck      Low
2007-02-06      6      Medium Box      463      Regular Air      Very Low
2007-02-09      16      Small Pack      142      Regular Air      Low
2007-03-01      29      Small Box      3212      Regular Air      Average
2007-03-02      46      Small Box      1603      Express Air      Very High
2007-03-03      23      Small Pack      214      Express Air      Low
2007-03-04      18      Small Box      2253      Regular Air      Low
2007-03-05      38      Small Box      1498      Express Air      Very High
2007-03-02      25      Small Pack      709      Regular Air      Average
2007-03-03      8      Small Box      351      Regular Air      Very Low
2007-03-04      37      Small Box      512      Regular Air      Very High
2007-03-05      15      Jumbo Drum      1044      Delivery Truck      Low
2007-03-06      10      Small Box      34      Regular Air      Very Low
2007-03-07      1      Large Box      4663      Express Air      Very Low
2007-03-08      29      Small Box      891      Regular Air      Average
2007-03-09      47      Small Pack      561      Express Air      Very High
2007-03-10      1      Jumbo Box      226      Delivery Truck      Very Low
2007-03-11      34      Small Box      88      Regular Air      High
2007-03-12      1      Wrap Bag      14      Regular Air      Very Low
2007-03-13      14      Jumbo Drum      4566      Delivery Truck      Very Low
2007-03-14      10      Small Box      1082      Regular Air      Very Low
2007-03-15      36      Small Box      2554      Express Air      Very High
2007-03-16      29      Jumbo Drum      3212      Regular Air      Average
Time taken: 18.862 seconds, Fetched: 30 row(s)
hive> 

```

4. Determine the minimum and maximum sales of all different product containers till each order date. Display all the fields and records.

Code:

Select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode,
MIN(Sales) OVER (PARTITION BY Product_Container Order By Order_Date)
MAX(Sales) OVER (PARTITION BY Product_Container Order By Order_Date)
From LogisticsOperations;

Output:

```

hive> select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode, MIN(Sales) OVER
(PARTITION BY Product_Container ORDER BY Order_Date), MAX(Sales) OVER (PARTITION BY Product_Container ORDER BY Order_Date) FROM LogisticsOperations;
Query ID = rakshandahedawoo_20240216054445_546ed518-7567-4166-9646-fb578ed9e781
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708056013283_0012)
-----
VERTICES      MODE      STATUS      TOTAL      COMPLETED      RUNNING      PENDING      FAILED      KILLED
-----
Map 1 ..... container      SUCCEEDED      1      1      0      0      0      0
Reducer 2 ..... container      SUCCEEDED      1      1      0      0      0      0
-----
VERTICES: 02/02 [=====>>] 100% ELAPSED TIME: 8.24 s
-----
OK
2007-02-04      15      Jumbo Box      9572      Delivery Truck      9572      9572
2007-03-10      1      Jumbo Box      226      Delivery Truck      226      9572
2007-01-29      15      Jumbo Drum      578      Delivery Truck      578      578
2007-02-02      42      Jumbo Drum      7373      Delivery Truck      578      7373
2007-03-05      15      Jumbo Drum      1044      Delivery Truck      578      7373
2007-03-13      14      Jumbo Drum      4566      Delivery Truck      578      7373
2007-03-16      29      Jumbo Drum      3212      Regular Air      578      7373
2007-01-28      39      Large Box      1780      Regular Air      1780      1780
2007-03-07      1      Large Box      4663      Express Air      1780      4663
2007-01-27      31      Medium Box      6567      Express Air      6567      6567
2007-02-06      6      Medium Box      463      Regular Air      463      6567
2007-01-31      10      Small Box      517      Regular Air      517      517
2007-02-01      41      Small Box      283      Regular Air      283      517
2007-03-01      29      Small Box      3212      Regular Air      283      3212
2007-03-02      46      Small Box      1603      Express Air      283      3212
2007-03-03      8      Small Box      351      Regular Air      283      3212
2007-03-04      37      Small Box      512      Regular Air      283      3212
2007-03-04      18      Small Box      2253      Regular Air      283      3212
2007-03-05      38      Small Box      1498      Express Air      283      3212
2007-03-06      10      Small Box      34      Regular Air      34      3212
2007-03-08      29      Small Box      891      Regular Air      34      3212
2007-03-11      34      Small Box      88      Regular Air      34      3212
2007-03-14      10      Small Box      1082      Regular Air      34      3212
2007-03-15      36      Small Box      2554      Express Air      34      3212
2007-01-30      30      Small Pack      611      Regular Air      611      611
2007-02-09      16      Small Pack      142      Regular Air      142      611

```

```

OK
2007-02-04      15      Jumbo Box      9572      Delivery Truck      9572      9572
2007-03-10       1      Jumbo Box      226      Delivery Truck      226      9572
2007-01-29      15      Jumbo Drum      578      Delivery Truck      578      578
2007-02-02      42      Jumbo Drum      7373      Delivery Truck      578      7373
2007-03-05      15      Jumbo Drum      1044      Delivery Truck      578      7373
2007-03-13      14      Jumbo Drum      4566      Delivery Truck      578      7373
2007-03-16      29      Jumbo Drum      3212      Regular Air         578      7373
2007-01-28      39      Large Box       1780      Regular Air         1780     1780
2007-03-07       1      Large Box       4663      Express Air         1780     4663
2007-01-27      31      Medium Box      6567      Express Air         6567     6567
2007-02-06       6      Medium Box      463      Regular Air         463      6567
2007-01-31      10      Small Box       517      Regular Air         517      517
2007-02-01      41      Small Box       283      Regular Air         283      517
2007-03-01      29      Small Box       3212      Regular Air         283      3212
2007-03-02      46      Small Box       1603      Express Air         283      3212
2007-03-03       8      Small Box       351      Regular Air         283      3212
2007-03-04      37      Small Box       512      Regular Air         283      3212
2007-03-04      18      Small Box       2253      Regular Air         283      3212
2007-03-05      38      Small Box       1498      Express Air         283      3212
2007-03-06      10      Small Box       34      Regular Air         34      3212
2007-03-08      29      Small Box       891      Regular Air         34      3212
2007-03-11      34      Small Box       88      Regular Air         34      3212
2007-03-14      10      Small Box       1082      Regular Air         34      3212
2007-03-15      36      Small Box       2554      Express Air         34      3212
2007-01-30      30      Small Pack      611      Regular Air         611      611
2007-02-09      16      Small Pack      142      Regular Air         142      611
2007-03-02      25      Small Pack      709      Regular Air         142      709
2007-03-03      23      Small Pack      214      Express Air         142      709
2007-03-09      47      Small Pack      561      Express Air         142      709
2007-03-12       1      Wrap Bag        14      Regular Air         14      14
Time taken: 8.737 seconds, Fetched: 30 row(s)
hive> 

```

5. Determine the previous and next quantity sold for different product containers for each order date. Display all the fields and records.

Code:

Select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode,
Lag(Order_Quantity) OVER (PARTITION BY Product_Container Order By Order_Date)
Lead(Order_Quantity) OVER (PARTITION BY Product_Container Order By Order_Date)
From LogisticsOperations;

Output:

```

hive> select Order_Date, Order_Quantity, Product_Container, Sales, Ship_Mode, Lag(Order_Quantity) OVER (PARTITION BY Product_Container ORDER BY Order_Date), Lead(Order_Quantity) OVER (PARTITION BY Product_Container ORDER BY Order_Date) FROM LogisticsOperations;
Query ID = rakshandahedawoo_20240216054723_346bada2-0291-4e87-aa8f-807952c071d7
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708056013283_0012)

-----
VERTICES      MODE      STATUS      TOTAL      COMPLETED      RUNNING      PENDING      FAILED      KILLED
-----
Map 1 ..... container      SUCCEEDED      1           1           0           0           0           0
Reducer 2 ..... container      SUCCEEDED      1           1           0           0           0           0
-----
VERTICES: 02/02 [=====>>>] 100% ELAPSED TIME: 7.49 s
-----
OK
2007-02-04      15      Jumbo Box      9572      Delivery Truck      NULL      1
2007-03-10       1      Jumbo Box      226      Delivery Truck      15      NULL
2007-01-29      15      Jumbo Drum      578      Delivery Truck      NULL      42
2007-02-02      42      Jumbo Drum      7373      Delivery Truck      15      15
2007-03-05      15      Jumbo Drum      1044      Delivery Truck      42      14
2007-03-13      14      Jumbo Drum      4566      Delivery Truck      15      29
2007-03-16      29      Jumbo Drum      3212      Regular Air         14      NULL
2007-01-28      39      Large Box       1780      Regular Air         NULL      1
2007-03-07       1      Large Box       4663      Express Air         39      NULL
2007-01-27      31      Medium Box      6567      Express Air         NULL      6
2007-02-06       6      Medium Box      463      Regular Air         31      NULL
2007-01-31      10      Small Box       517      Regular Air         NULL      41
2007-02-01      41      Small Box       283      Regular Air         10      29
2007-03-01      29      Small Box       3212      Regular Air         41      46
2007-03-02      46      Small Box       1603      Express Air         29      8
2007-03-03       8      Small Box       351      Regular Air         46      37
2007-03-04      37      Small Box       512      Regular Air         8      18
2007-03-04      18      Small Box       2253      Regular Air         37      38
2007-03-05      38      Small Box       1498      Express Air         18      10
2007-03-06      10      Small Box       34      Regular Air         38      29
2007-03-08      29      Small Box       891      Regular Air         10      34
2007-03-11      34      Small Box       88      Regular Air         29      10
2007-03-14      10      Small Box       1082      Regular Air         34      36
2007-03-15      36      Small Box       2554      Express Air         10      NULL
2007-01-30      30      Small Pack      611      Regular Air         NULL      16
2007-02-09      16      Small Pack      142      Regular Air         30      25

```

```

OK
2007-02-04      15      Jumbo Box      9572      Delivery Truck      NULL      1
2007-03-10      1      Jumbo Box      226      Delivery Truck      15      NULL
2007-01-29      15      Jumbo Drum      578      Delivery Truck      NULL      42
2007-02-02      42      Jumbo Drum      7373     Delivery Truck      15      15
2007-03-05      15      Jumbo Drum      1044     Delivery Truck      42      14
2007-03-13      14      Jumbo Drum      4566     Delivery Truck      15      29
2007-03-16      29      Jumbo Drum      3212     Regular Air         14      NULL
2007-01-28      39      Large Box       1780     Regular Air         NULL     1
2007-03-07      1      Large Box       4663     Express Air         39      NULL
2007-01-27      31      Medium Box      6567     Express Air         NULL     6
2007-02-06      6      Medium Box      463      Regular Air         31      NULL
2007-01-31      10     Small Box       517      Regular Air         NULL     41
2007-02-01      41     Small Box       283      Regular Air         10      29
2007-03-01      29     Small Box       3212     Regular Air         41      46
2007-03-02      46     Small Box       1603     Express Air         29      8
2007-03-03      8      Small Box       351      Regular Air         46      37
2007-03-04      37     Small Box       512      Regular Air         8       18
2007-03-04      18     Small Box       2253     Regular Air         37      38
2007-03-05      38     Small Box       1498     Express Air         18      10
2007-03-06      10     Small Box       34       Regular Air         38      29
2007-03-08      29     Small Box       891      Regular Air         10      34
2007-03-11      34     Small Box       88       Regular Air         29      10
2007-03-14      10     Small Box       1082     Regular Air         34      36
2007-03-15      36     Small Box       2554     Express Air         10      NULL
2007-01-30      30     Small Pack      611      Regular Air         NULL     16
2007-02-09      16     Small Pack      142      Regular Air         30      25
2007-03-02      25     Small Pack      709      Regular Air         16      23
2007-03-03      23     Small Pack      214      Express Air         25      47
2007-03-09      47     Small Pack      561      Express Air         23      NULL
2007-03-12      1      Wrap Bag       14       Regular Air         NULL     NULL
Time taken: 7.922 seconds, Fetched: 30 row(s)
hive>

```

Section C-

Assist the manager of the HR Department by executing the following queries considering the attached file “Departmentdata.json”. The file contains information of 12 departments. The information related to department name, manager, budget allocated (in millions) and the information of 2 employees from each department is included in the json file. Create the structure of table and load data from the file in HIVE.

Code:

```
ADD JAR hdfs:///user/rakshandahedawoo/json-serde-1.3.8-jar-with-dependencies.jar;
```

```
create table Department (deptname string, manager String, employees
ARRAY<struct<employee_id:String, name:
struct<fname:String,lname:String>,position:String, salary:INT >>, budget INT) ROW
FORMAT SERDE 'org.openx.data.jsonserde.JsonSerDe' stored as textfile;
```

```
load data local inpath 'Departmentdata.json' into table Department;
```

```

hive> ADD JAR hdfs://user/rakshandahedawoo/json-serde-1.3.8-jar-with-dependencies.jar;
Added [/tmp/f6475a31-a9a9-4e03-9dda-25768630fa0f_resources/json-serde-1.3.8-jar-with-dependencies.jar] to class
path
Added resources: [hdfs://user/rakshandahedawoo/json-serde-1.3.8-jar-with-dependencies.jar]
hive> create table Department (deptname string, manager String, employees ARRAY<struct<employee_id:String, name
: struct<fname:String,lname:String>,position:String, salary:INT >>, budget INT) ROW FORMAT SERDE 'org.openx.dat
a.jsonserde.JsonSerDe' stored as textfile;
OK
Time taken: 0.589 seconds
hive> load data local inpath 'Departmentdata.json' into table Department;
Loading data to table walmart.department
OK
Time taken: 0.499 seconds
hive> select * from Department;
OK
Engineering      JohnDoe [{"employee_id":"1","name":{"fname":"Alice","lname":"Smith"},"position":"SoftwareEngine
er","salary":80000}, {"employee_id":"2","name":{"fname":"Bob","lname":"Johnson"},"position":"SeniorSoftwareEngin
eer","salary":100000}] 9
Marketing         JaneSmith [{"employee_id":"3","name":{"fname":"Charlie","lname":"Brown"},"position":"Mark
etingCoordinator","salary":60000}, {"employee_id":"4","name":{"fname":"Diana","lname":"Miller"},"position":"Mark
etingManager","salary":90000}] 7
Finance           MichaelJohnson [{"employee_id":"5","name":{"fname":"Emma","lname":"Davis"},"position":"FinancialAnalys
t","salary":75000}, {"employee_id":"6","name":{"fname":"Frank","lname":"Wilson"},"position":"SeniorFinancialAnal
yst","salary":95000}] 6
Sales             EmilyBrown [{"employee_id":"7","name":{"fname":"Grace","lname":"Lee"},"position":"SalesRepresentat
ive","salary":70000}, {"employee_id":"8","name":{"fname":"Henry","lname":"Clark"},"position":"SalesManager","sal
ary":100000}] 12
Research          DavidJohnson [{"employee_id":"9","name":{"fname":"Isabella","lname":"Martinez"},"position":"
ResearchScientist","salary":85000}, {"employee_id":"10","name":{"fname":"Jack","lname":"Thompson"},"position":"S
eniorResearchScientist","salary":110000}] 13
HumanResources    SarahAdams [{"employee_id":"11","name":{"fname":"Kelly","lname":"White"},"position":"HRSpe
cialist","salary":65000}, {"employee_id":"12","name":{"fname":"Liam","lname":"Harris"},"position":"HRManager","s
alary":95000}] 8
Operations         OliviaWilson [{"employee_id":"13","name":{"fname":"Mia","lname":"Thomas"},"position":"Operat
ionsCoordinator","salary":60000}, {"employee_id":"14","name":{"fname":"Nathan","lname":"Rodriguez"},"position":"

```

Queries:

1. Display the department name along with the last name of the first 2 employees of each department.

Code:

```

select deptname, employees[0].name.lname, employees[1].name.lname
from Department;

```

```

hive> select deptname,employees[0].name.lname,employees[1].name.lname from Department;
Query ID = rakshandahedawoo_20240217000244_acbee063-605f-4d8f-9cbe-3f87fbd7020b
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708125033784_0005)

-----
VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01  [=====>>] 100%  ELAPSED TIME: 7.78 s
-----
OK
Engineering      Smith  Johnson
Marketing         Brown  Miller
Finance Davis    Wilson
Sales Lee       Clark
Research          Martinez Thompson
HumanResources    White  Harris
Operations         Thomas Rodriguez
CustomerService  Martinez Garcia
QualityAssurance  Perez  Scott
Legal Green      Martinez
IT Lewis        Anderson
Production        Turner  Mitchell
Time taken: 10.352 seconds, Fetched: 12 row(s)
hive> 

```


2. Display the name of the department and manager if either of the first two employees has salary > 100000.

Code:

Select deptname, manager

From Department

Where (employees[0].salary>100000 or employees[1].salary>100000);

```
hive> Select deptname,manager From Department Where (employees[0].salary>100000 or employees[1].salary>100000);

Query ID = rakshandahedawoo_20240217000535_17d0798b-86a4-4a0b-8ca6-12c2f2a098f8
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708125033784_0005)

-----
      VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01  [=====>>>] 100%  ELAPSED TIME: 8.64 s
-----

OK
Research      DavidJohnson
Legal  VictoriaKing
IT      WilliamBrown
Time taken: 9.797 seconds, Fetched: 3 row(s)
hive> 
```

3. Display the name of those departments that has a budget of more than 12 Million and salary of the first 2 employees belonging to that department.

Code:

Select deptname, employees[0].salary, employees[1].salary

From Department Where budget>12;

```
hive> Select deptname,employees[0].salary,employees[1].salary From Department Where budget>12;

Query ID = rakshandahedawoo_20240217000837_39cb014d-6c6c-4fc7-aa8e-cc81c0bc169c
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708125033784_0005)

-----
      VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01  [=====>>>] 100%  ELAPSED TIME: 7.30 s
-----

OK
Research      85000    110000
CustomerService 55000    85000
IT      65000    105000
Production 50000    95000
Time taken: 7.738 seconds, Fetched: 4 row(s)
hive> 
```

4. Display the first name and position of the first 2 employees whose manager name is starting with “D”.

Code:

```
Select employees[0].name.fname,employees[0].position,
employees[1].name.fname,employees[1].position
From Department Where manager Like 'D%';
```

```
hive> Select employees[0].name.fname,employees[0].position,employees[1].name.fname,employees[1]
].position From Department Where manager Like 'D%';
Query ID = rakshandahedawoo_20240217001124_9ded1c7a-605d-4daa-9e9e-61e7b1967f06
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708125033784_0005)

-----
      VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01  [=====>>] 100%  ELAPSED TIME: 7.69 s
-----
OK
Isabella      ResearchScientist      Jack      SeniorResearchScientist
Time taken: 8.007 seconds, Fetched: 1 row(s)
hive> 
```

5. Display the budget and position of first two employees associated with Finance, Marketing, Sales and Legal department.

Code:

```
Select budget, employees[0].position, employees[1].position
From Department Where deptname in ('Finance','Marketing','Legal','Sales');
```

```
hive> Select budget, employees[0].position, employees[1].position From Department Where deptna
me in ('Finance','Marketing','Legal','Sales');
Query ID = rakshandahedawoo_20240217001351_68b924ce-80d4-416c-aaec-5d922e938385
Total jobs = 1
Launching Job 1 out of 1
Status: Running (Executing on YARN cluster with App id application_1708125033784_0005)

-----
      VERTICES      MODE      STATUS  TOTAL  COMPLETED  RUNNING  PENDING  FAILED  KILLED
-----
Map 1 ..... container      SUCCEEDED      1          1          0          0          0          0
-----
VERTICES: 01/01  [=====>>] 100%  ELAPSED TIME: 7.91 s
-----
OK
7      MarketingCoordinator      MarketingManager
6      FinancialAnalyst      SeniorFinancialAnalyst
12     SalesRepresentative      SalesManager
11     LegalAssistant      LegalCounsel
Time taken: 8.286 seconds, Fetched: 4 row(s)
hive> 
```