

Big Data Processing Lab-6

Name: Sumit Vishwakarma

Roll No: 202201320

ColabLink:

https://colab.research.google.com/drive/1cM5b-p2byG6z2JuKlsey6AtCv7xh_MBt#scrollTo=AWcj40rYPDR1

Q1: In this lab also, we use the same dataset of the Lab05 and perform the following tasks.

Compute full data cube lattice as discussed in lectures.

<https://moodle.daiict.ac.in/mod/resource/view.php?id=6242> and materialize it in parquet data format. For efficient querying, store every lattice node in a separate data file.

Perform the following operations SalesCube.

1. List total sales (amount) of all items of category 'Tshirts' Project: item,
sum(amount)

```
# Query to get total sales of all items of category 'Tshirts'
query = """
SELECT i.itemID AS item, SUM(sa.price) AS total_sales
FROM Sales sa
JOIN Item i ON sa.itemID = i.itemID
WHERE i.category = 'Tshirt'
GROUP BY i.itemID
"""

# Execute query and save result
tshirt_sales_df = spark.sql(query)
tshirt_sales_df.write.mode('overwrite').parquet('/content/gdrive/My Drive/iris/datacube1.parquet')
df = spark.read.parquet('/content/gdrive/My Drive/iris/datacube1')
df.show()
```

item	total_sales
item1	135
item3	780

Q2: List total sales (amount) of all items of category 'Tshirts' at 'store1' Project:

item, sum(amount)

```
[31] # Query to get total sales of 'Tshirts' at 'store1'
query = """
SELECT i.itemID AS item, SUM(sa.price) AS total_sales
FROM Sales sa
JOIN Item i ON sa.itemID = i.itemID
WHERE i.category = 'Tshirt' AND sa.storeID = 'store1'
GROUP BY i.itemID
"""

# Execute query and save result
store1_tshirt_sales_df = spark.sql(query)
store1_tshirt_sales_df.write.mode('overwrite').parquet('/content/store1_tshirt_sales.parquet')
store1_tshirt_sales_df.show()
```

item	total_sales
item1	70
item3	105

Q3. Give total sales (amount) of customer 'cust1'. Project: sum(amount)

```
[32] # Query to get total sales of customer 'cust1'
query = """
SELECT SUM(sa.price) AS total_sales
FROM Sales sa
WHERE sa.custID = 'cust1'
"""

# Execute query and save result
customer_sales_df = spark.sql(query)
customer_sales_df.write.mode('overwrite').parquet('/content/customer_sales.parquet')
customer_sales_df.show()
```

total_sales
670

Q4: Give total sales (amount) of customer 'cust1' from items of category 'Tshirts' Project: item, sum(amount)

```
[33] # Query to get total sales of customer 'cust1' from 'Tshirts'
query = """
SELECT i.itemID AS item, SUM(sa.price) AS total_sales
FROM Sales sa
JOIN Item i ON sa.itemID = i.itemID
WHERE sa.custID = 'cust1' AND i.category = 'Tshirt'
GROUP BY i.itemID
"""

# Execute query and save result
customer_tshirt_sales_df = spark.sql(query)
customer_tshirt_sales_df.write.mode('overwrite').parquet('/content/customer_tshirt_sales.parquet')
customer_tshirt_sales_df.show()
```

item	total_sales
item1	10
item3	170

Q5. Give total sales (amount) of all stores of 'WA' state for items of category 'Tshirts' Project: store,item, sum(amount)

```
[34] # Query to get total sales of 'Tshirts' in 'WA'
query = """
SELECT s.storeID AS store, i.itemID AS item, SUM(sa.price) AS total_sales
FROM Sales sa
JOIN Store s ON sa.storeID = s.storeID
JOIN Item i ON sa.itemID = i.itemID
WHERE i.category = 'Tshirt' AND s.state = 'WA'
GROUP BY s.storeID, i.itemID
"""

# Execute query and save result
wa_tshirt_sales_df = spark.sql(query)
wa_tshirt_sales_df.write.mode('overwrite').parquet('/content/wa_tshirt_sales.parquet')
wa_tshirt_sales_df.show()
```

store	item	total_sales
store6	item3	165

Q6. Give gender-wise total sales (amount) of items in the 'Jacket' category.

Project: gender,item, sum (amount)

```
[35] # Query to get gender-wise total sales of 'Jacket'
query = """
SELECT c.gender AS gender, i.itemID AS item, SUM(sa.price) AS total_sales
FROM Sales sa
JOIN Item i ON sa.itemID = i.itemID
JOIN Customer c ON sa.custID = c.custID
WHERE i.category = 'Jacket'
GROUP BY c.gender, i.itemID
"""

# Execute query and save result
gender_jacket_sales_df = spark.sql(query)
gender_jacket_sales_df.write.mode('overwrite').parquet('/content/gender_jacket_sales.parquet')
gender_jacket_sales_df.show()
```

```
↻
+-----+-----+
|gender| item|total_sales|
+-----+-----+
| M|item5|      330|
| F|item4|      495|
| F|item2|      565|
| M|item2|      760|
| M|item4|      160|
| F|item5|      125|
+-----+-----+
```