Redshift Lab - 6

06:12 PM

Perform Analytical Queries on the Redshift Tables

The online store is going to run a campaign and will need some information from their redshift data warehouse. Your manager has asked you to extract the following useful information from the redshift cluster:

- Total items and customer information regarding items sold on 12/31/2020
- Top ten buyers by quantity and items sold.
- Top ten buyers by total price of items purchased.

You will also store the information for customers who have purchased something.

1. Overwrite the contents of the **Query 1** editor with the following, and click **Run**. This will query the orders table and show the total amount of items sold on 12/31/2020.
   SELECT sum(quantity)
   FROM ps_store_schema.orders
   WHERE purchase_date = '12/31/2020';
   You should see a table underneath showing the sum of the items sold on the particular date.
2. Overwrite the contents of the **Query 1** editor with the following, and click **Run** to show customer information for items sold on 12/31/2020.
   select first_name, last_name, email, product_name, quantity, quantity * price as total_price
   from ps_store_schema.customers As c
   join ps_store_schema.orders AS s on c.customer_number = s.customer_id
   join ps_store_schema.products AS p on s.product_id = p.product_id
   WHERE purchase_date = '12/31/2020';
   You should also see a table underneath showing four customers.
3. Overwrite the contents of the **Query 1** editor with the following, and click **Run** to query show the top ten buyers by the quantity.
   SELECT first_name, last_name, email, total_quantity
   FROM (SELECT customer_id, sum(quantity) total_quantity
   FROM ps_store_schema.orders
   GROUP BY customer_id
   ORDER BY total_quantity desc limit 10) S, ps_store_schema.customers
   WHERE S.customer_id = customer_number
   ORDER BY S.total_quantity desc;
   You should see a resulting table showing the top 10 buyers.
   Note: Above the table you can click **Export** to export the resulting data in various formats.
4. Overwrite the contents of the **Query 1** editor with the following, and click **Run** to show the top ten items sold by quantity.
   SELECT product_name, total_quantity
   FROM (SELECT product_id, sum(quantity) total_quantity

```
FROM ps_store_schema.orders
GROUP BY product_id
ORDER BY total_quantity desc limit 10) AS S, ps_store_schema.products
WHERE S.product_id = products.product_id
ORDER BY total_quantity desc;
```

You should see a table underneath showing the top ten items.

5. Overwrite the contents of the **Query 1** editor with the following, and click **Run** to show the top ten buyers by total price of items purchased.

```
with sales as (select customer_number, first_name, last_name, email,
product_name, quantity, quantity * price as total_price
from ps_store_schema.customers As c
join ps_store_schema.orders AS s on c.customer_number = s.customer_id
join ps_store_schema.products AS p on s.product_id = p.product_id
order by first_name, last_name)
select customer_number, first_name, last_name, email, SUM(total_price) from
sales
group by customer_number, first_name, last_name, email
order by sum DESC;
```

You should see a table underneath showing the top ten buyers.

6. Overwrite the contents of the **Query 1** editor with the following command after replacing <S3-bucket-name> and <Role-arn>, then click **Run**. This stores in the S3 bucket all the customers who have purchased something.

```
unload ('select customer_number, first_name, last_name, email, product_name,
quantity, quantity * price as total_price
from ps_store_schema.customers As c
join ps_store_schema.orders AS s on c.customer_number = s.customer_id
join ps_store_schema.products AS p on s.product_id = p.product_id
order by first_name, last_name')
to 's3://<S3-bucket-name>/unload/info-data-'
iam_role '<Role-arn>'
CSV
header
allowoverwrite
parallel off;
```

7. In the top search box, type in and click on **S3**. Click on the **ps-lab-resources-** link, and then click the **unload/** folder link.

You will see the CSV file whose name starts with info-data and contains unloaded data in the CSV format.