

Hands-on Lab: Querying the Data Warehouse (Cubes, Rollups, Grouping Sets and Materialized Views)



Estimated time needed: 30 minutes

Purpose of the Lab:

The purpose of this lab is to provide hands-on experience in advanced SQL query techniques using PostgreSQL in a Cloud IDE environment. The lab focuses on teaching how to create grouping sets, rollups, and cubes for data aggregation and summarization, as well as how to implement and utilize Materialized Query Tables (MQT) for efficient data querying. These skills are essential for managing and analyzing large datasets, particularly in data warehousing and business intelligence contexts.

Benefits of Learning the Lab:

By completing this lab, you will gain valuable insights into the practical application of complex SQL queries and data manipulation techniques. The knowledge of grouping sets, rollups, and cubes will enable learners to effectively summarize and analyze data, which is crucial in making informed business decisions. Understanding and implementing MQTs provides an efficient way to handle large-scale data by reducing the computational load during frequent query executions. These skills are highly beneficial for careers in data analysis, database administration, and business intelligence, enhancing your ability to manage and analyze data in real-world scenarios.

Objectives

In this lab you will learn how to create:

- Grouping sets
- Rollup
- Cube
- Materialized Query Tables (MQT)

Exercise 1 - Launch a PostgreSQL server instance on Cloud IDE and open up the pgAdmin Graphical User Interface.

This lab requires that you complete the previous lab Populate a Data Warehouse.

If you have not finished the Populate a Data Warehouse Lab yet, please finish it before you continue.

GROUPING SETS, CUBE, and ROLLUP allow us to easily create subtotals and grand totals in a variety of ways. All these operators are used along with the GROUP BY operator.

GROUPING SETS operator allows us to group data in a number of different ways in a single SELECT statement.

The **ROLLUP** operator is used to create subtotals and grand totals for a set of columns. The summarized totals are created based on the columns passed to the ROLLUP operator.

The **CUBE** operator produces subtotals and grand totals. In addition, it produces subtotals and grand totals for every permutation of the columns provided to the CUBE operator.

Exercise 2 - Write a query using grouping sets

After you launch a PostgreSQL server instance on Cloud IDE and open up the pgAdmin Graphical User Interface run the below query.

To create a grouping set for three columns labeled year, category, and sum of billedamount, run the sql statement below.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7

1. select year,category, sum(billedamount) as totalbilledamount
2. from "FactBilling"
3. left join "DimCustomer"
4. on "FactBilling".customerid = "DimCustomer".customerid
5. left join "DimMonth"
6. on "FactBilling".monthid="DimMonth".monthid
7. group by grouping sets(year,category);
```

Copied!

The partial output can be seen in the image below.

The screenshot shows the pgAdmin 4 interface with the following components:

- Browser:** A tree view on the left showing the database structure. The 'production1' database is selected, and the 'public' schema is expanded, showing tables like 'FactBilling', 'DimCustomer', and 'DimMonth'.
- Query Editor:** The central pane contains a SQL query:

```
1 select year,category, sum(billedamount) as totalbilledamount
2 from "FactBilling"
3 left join "DimCustomer"
4 on "FactBilling".customerid = "DimCustomer".customerid
5 left join "DimMonth"
6 on "FactBilling".monthid="DimMonth".monthid
7 group by grouping sets(year,category);
8
```
- Data Output:** The bottom pane displays the query results in a table with 4 columns: 'year' (integer), 'category' (character varying (10)), and 'totalbilledamount' (bigint). The results show 13 rows of data for years 2015 through 2018.
- Notifications:** A green message box at the bottom right states: 'Successfully run. Total query runtime: 947 msec. 13 rows affected.'

Exercise 3 - Write a query using rollup

To create a rollup using the three columns year, category and sum of billedamount, run the sql statement below.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8

1. select year,category, sum(billedamount) as totalbilledamount
2. from "FactBilling"
3. left join "DimCustomer"
4. on "FactBilling".customerid = "DimCustomer".customerid
5. left join "DimMonth"
6. on "FactBilling".monthid="DimMonth".monthid
7. group by rollup(year,category)
8. order by year, category;
```

Copied!

The partial output can be seen in the image below.

The screenshot shows the pgAdmin 4 interface with the following components:

- Browser:** The 'postgres' database is selected, and the 'public' schema is expanded, showing the 'FactBilling' table.
- Query Editor:** The central pane contains a SQL query:

```
1 select year,category, sum(billedamount) as totalbilledamount
2 from "FactBilling"
3 left join "DimCustomer"
4 on "FactBilling".customerid = "DimCustomer".customerid
5 left join "DimMonth"
6 on "FactBilling".monthid="DimMonth".monthid
7 group by rollup(year,category)
8 order by year, category;
```
- Data Output:** The bottom pane displays the query results in a table with 4 columns: 'year' (integer), 'category' (character varying (10)), and 'totalbilledamount' (bigint). The results show 9 rows of data for years 2009 through 2011.
- Notifications:** A green message box at the bottom right states: 'Successfully run. Total query runtime: 694 msec. 34 rows affected.'

Exercise 4 - Write a query using cube

To create a cube using the three columns labeled year, category, and sum of billedamount, run the sql statement below.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8

1. select year,category, sum(billedamount) as totalbilledamount
2. from "FactBilling"
3. left join "DimCustomer"
4. on "FactBilling".customerid = "DimCustomer".customerid
5. left join "DimMonth"
6. on "FactBilling".monthid="DimMonth".monthid
7. group by cube(year,category)
8. order by year, category;
```

Copied!

The partial output can be seen in the image below.

The screenshot shows the pgAdmin interface. On the left, the 'Browser' pane displays the database structure, with 'FactBilling' selected under 'public'. The 'Query Editor' pane shows the following SQL query:

```
1 select year,category, sum(billedamount) as totalbilledamount
2 from "FactBilling"
3 left join "DimCustomer"
4 on "FactBilling".customerid = "DimCustomer".customerid
5 left join "DimMonth"
6 on "FactBilling".monthid="DimMonth".monthid
7 group by cube(year,category)
8 order by year, category;
```

Below the query editor, the 'Data Output' tab is active, displaying the following table:

	year	category	totalbilledamount
	integer	character varying (10)	bigint
1	2009	Company	59048255
2	2009	Individual	61215072
3	2009	[null]	120263327
4	2010	Company	58725739
5	2010	Individual	60758919
6	2010	[null]	119484658
7	2011	Company	58559675
8	2011	Individual	60867794
9	2011	[null]	119427469

A green notification box at the bottom right states: 'Successfully run. Total query runtime: 751 msec. 36 rows affected.'

Exercise 5 - Create a Materialized Query Table(MQT)

In pgAdmin we can implement materialized views using Materialized Query Tables.

Step 1: Create the MQT.

Execute the sql statement below to create an MQT named countrystats.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8

1. CREATE MATERIALIZED VIEW countrystats (country, year, totalbilledamount) AS
2. (select country, year, sum(billedamount)
3. from "FactBilling"
4. left join "DimCustomer"
5. on "FactBilling".customerid = "DimCustomer".customerid
6. left join "DimMonth"
7. on "FactBilling".monthid="DimMonth".monthid
8. group by country,year);
```

Copied!

The above command creates an MQT named countrystats that has 3 columns.

- Country
- Year
- totalbilledamount

The MQT is essentially the result of the below query, which gives you the year, quartername and the sum of billed amount grouped by year and quartername.

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7

1. select year, quartername, sum(billedamount) as totalbilledamount
2. from "FactBilling"
3. left join "DimCustomer"
```

```
4. on "FactBilling".customerid = "DimCustomer".customerid
5. left join "DimMonth"
6. on "FactBilling".monthid="DimMonth".monthid
7. group by grouping sets(year, quartername);
```

Copied!

Step 2: Populate/refresh data into the MQT.

Execute the sql statement below to populate the MQT countrystats.

```
1. 1
1. REFRESH MATERIALIZED VIEW countrystats;
```

Copied!

The command above populates the MQT with relevant data.

Step 3: Query the MQT.

Once an MQT is refreshed, you can query it.

Execute the sql statement below to query the MQT countrystats.

```
1. 1
1. select * from countrystats;
```

Copied!

Practice exercises

Problem 1: Create a grouping set for the columns year, quartername, sum(billedamount).

- [Click here for Hint](#)
- ▼ [Click here for Solution](#)

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
1. select year, quartername, sum(billedamount) as totalbilledamount
2. from "FactBilling"
3. left join "DimCustomer"
4. on "FactBilling".customerid = "DimCustomer".customerid
5. left join "DimMonth"
6. on "FactBilling".monthid="DimMonth".monthid
7. group by grouping sets(year, quartername);
```

Copied!

Problem 2: Create a rollup for the columns country, category, sum(billedamount).

- [Click here for Hint](#)
- ▼ [Click here for Solution](#)

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8
1. select year, quartername, sum(billedamount) as totalbilledamount
2. from "FactBilling"
3. left join "DimCustomer"
4. on "FactBilling".customerid = "DimCustomer".customerid
5. left join "DimMonth"
6. on "FactBilling".monthid="DimMonth".monthid
7. group by rollup(year, quartername)
8. order by year, quartername;
```

Copied!

Problem 3: Create a cube for the columns year,country, category, sum(billedamount).

- [Click here for Hint](#)
- ▼ [Click here for Solution](#)

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
7. 7
1. select year, quartername, sum(billedamount) as totalbilledamount
2. from "FactBilling"
3. left join "DimCustomer"
4. on "FactBilling".customerid = "DimCustomer".customerid
5. left join "DimMonth"
6. on "FactBilling".monthid="DimMonth".monthid
7. group by cube(year,quartername);
```

Copied!

Problem 4: Create an MQT named average_billamount with columns year, quarter, category, country, average_bill_amount.

- [Click here for Hint](#)
- ▼ [Click here for Solution](#)

```
1. 1
2. 2
3. 3
4. 4
5. 5
6. 6
```

```
7. 7
8. 8
9. 9

1. CREATE MATERIALIZED VIEW average_billamount (year,quarter,category,country, average_bill_amount) AS
2. (select year,quarter,category,country, avg(billedamount) as average_bill_amount
3. from "FactBilling"
4. left join "DimCustomer"
5. on "FactBilling".customerid = "DimCustomer".customerid
6. left join "DimMonth"
7. on "FactBilling".monthid="DimMonth".monthid
8. group by year,quarter,category,country
9. );
```

Copied!

```
1. 1

1. refresh MATERIALIZED VIEW average_billamount;
```

Copied!

Congratulations! You have successfully finished the Populating a Data Warehouse lab.

Author

Amrutha Rao

Change Log

Date (YYYY-MM-DD)	Version	Changed By	Change Description
2022-04-14	0.2	Amrutha Rao	converted initial version to pgAdmin workaround.
2021-09-29	0.1	Ramesh Sannareddy	Created initial version of the lab

© IBM Corporation 2022. All rights reserved.