OLAP Operations (using Redshift or PostgreSQL)

Develop the queries to retrieve information from the OLAP operations performed and to gain a deeper understanding of the sales data through different dimensions, aggregations, and filters.

Objective: Perform OLAP operations (Drill Down, Rollup, Cube, Slice, and Dice) on the "sales_sample" table to analyze sales data. The project will include the following tasks: 1.Database Creation Create a database to store the sales data (Redshift or PostgreSQL). Create a table named "sales_sample" with the specified columns: Product_Id (Integer) Region (varchar(50))-like East ,West etc Date (Date) Sales_Amount (int/numeric)

Query:

CREATE DATABASE sales_olap;

\c sales_olap

CREATE TABLE sales_sample (Product_Id INTEGER, Region VARCHAR(50), Date DATE, Sales_Amount NUMERIC);

2.Data Creation

Query:

Insert 10 sample records into the "sales_sample" table, representing sales data.

INSERT INTO sales_sample (Product_Id, Region, Date, Sales_Amount) VALUES (101, 'East', '2024-01-15', 5000), (102, 'West', '2024-01-15', 6000), (101, 'North', '2024-01-16', 4500), (103, 'South', '2024-01-16', 5500), (102, 'East', '2024-01-17', 7000), (101, 'West', '2024-01-17', 4800), (103, 'North', '2024-01-18', 6200), (102, 'South', '2024-01-18', 5300), (101, 'East', '2024-01-19', 5800), (103, 'West', '2024-01-19', 6500);

Output:

OLAP Operations:

a) Drill Down:

Query:

SELECT Region, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY Region ORDER BY Region;

Output:

```
sales_olap=#
sales_olap=# SELECT Region, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY Region ORDER BY Region;
 region | total_sales
--------+-------------
 East   |       17800
 North  |       10700
 South  |       10800
```

Query:

SELECT Region, Product_Id, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY Region, Product_Id ORDER BY Region, Product_Id;

Output:

```
 South  |       10800
 West   |       17300
(4 rows)


sales_olap=# SELECT Region, Product_Id, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY Region, Product_Id ORDER BY Region, Product_Id;
 region | product_id | total_sales
--------+------------+-------------
 East   |        101 |       10800
 East   |        102 |        7000
 North  |        101 |        4500
 North  |        103 |        6200
 South  |        102 |        5300
 South  |        103 |        5500
 West   |        101 |        4800
 West   |        102 |        6000
 West   |        103 |        6500
(9 rows)
```

Query:

SELECT Region, Product_Id, Date, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY Region, Product_Id, Date ORDER BY Region, Product_Id, Date;

Output:

```
sales_olap=# SELECT Region, Product_Id, Date, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY Region, Product_Id, Date ORDER BY Region, Product_Id, Date;
 region | product_id |    date    | total_sales
--------+------------+------------+-------------
 East   |        101 | 2024-01-15 |        5000
 East   |        101 | 2024-01-19 |        5800
 East   |        102 | 2024-01-17 |        7000
 North  |        101 | 2024-01-16 |        4500
 North  |        103 | 2024-01-18 |        6200
 South  |        102 | 2024-01-18 |        5300
 South  |        103 | 2024-01-16 |        5500
 West   |        101 | 2024-01-17 |        4800
 West   |        102 | 2024-01-15 |        6000
 West   |        103 | 2024-01-19 |        6500
(10 rows)
```

b) Rollup:

Query:

SELECT Region, Product_Id, SUM(Sales_Amount) as Total_Sales FROM sales_sample
GROUP BY ROLLUP(Region, Product_Id) ORDER BY Region, Product_Id;

Output:

```
sales_olap=# SELECT Region, Product_Id, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY ROLLUP(Region, Product_Id) ORDER BY Region, Product_Id;
 region | product_id | total_sales
--------+------------+-------------
 East   |        101 |       10800
 East   |        102 |        7000
 East   |            |       17800
 North  |        101 |        4500
 North  |        103 |        6200
 North  |            |       10700
 South  |        102 |        5300
 South  |        103 |        5500
 South  |            |       10800
 West   |        101 |        4800
 West   |        102 |        6000
 West   |        103 |        6500
 West   |            |       17300
        |            |       56600
(14 rows)
```

c) Cube:

Query:

SELECT COALESCE(Region, 'All Regions') as Region, COALESCE(TO_CHAR(Date, 'YYYY-MM'), 'All Dates') as Month, COALESCE(Product_Id::TEXT, 'All Products') as Product, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY CUBE(Region, TO_CHAR(Date, 'YYYY-MM'), Product_Id) ORDER BY Region, Month, Product;

Output:

```
sales_olap=# SELECT COALESCE(Region, 'All Regions') as Region, COALESCE(TO_CHAR(Date, 'YYYY-MM'), 'All Dates') as Month, COALESCE(Product_Id::TEXT, 'All Products') as Pro
duct, SUM(Sales_Amount) as Total_Sales FROM sales_sample GROUP BY CUBE(Region, TO_CHAR(Date, 'YYYY-MM'), Product_Id) ORDER BY Region, Month, Product;
   region    |   month   |   product    | total_sales
-------------+-----------+--------------+-------------
 All Regions | 2024-01   | 101          |       20100
 All Regions | 2024-01   | 102          |       18300
 All Regions | 2024-01   | 103          |       18200
 All Regions | 2024-01   | All Products |       56600
 All Regions | All Dates | 101          |       20100
 All Regions | All Dates | 102          |       18300
 All Regions | All Dates | 103          |       18200
 All Regions | All Dates | All Products |       56600
 East        | 2024-01   | 101          |       10800
 East        | 2024-01   | 102          |        7000
 East        | 2024-01   | All Products |       17800
 East        | All Dates | 101          |       10800
 East        | All Dates | 102          |        7000
 East        | All Dates | All Products |       17800
 North       | 2024-01   | 101          |        4500
 North       | 2024-01   | 103          |        6200
 North       | 2024-01   | All Products |       10700
 North       | All Dates | 101          |        4500
 North       | All Dates | 103          |        6200
 North       | All Dates | All Products |       10700
 South       | 2024-01   | 102          |        5300
 South       | 2024-01   | 103          |        5500
 South       | 2024-01   | All Products |       10800
 South       | All Dates | 102          |        5300
 South       | All Dates | 103          |        5500
 South       | All Dates | All Products |       10800
 West        | 2024-01   | 101          |        4800
 West        | 2024-01   | 102          |        6000
 West        | 2024-01   | 103          |        6500
 West        | 2024-01   | All Products |       17300
 West        | All Dates | 101          |        4800
 West        | All Dates | 102          |        6000
 West        | All Dates | 103          |        6500
 West        | All Dates | All Products |       17300
(34 rows)
```

d) Slice:

Query:

SELECT Product_Id, Date, Sales_Amount FROM sales_sample WHERE Region = 'East';

SELECT Product_Id, Region, Sales_Amount FROM sales_sample WHERE Date BETWEEN '2024-01-15' AND '2024-01-17';

Output:

```
sales_olap=# SELECT Product_Id, Date, Sales_Amount FROM sales_sample WHERE Region = 'East';
 product_id |    date    | sales_amount
------------+------------+--------------
        101 | 2024-01-15 |         5000
        102 | 2024-01-17 |         7000
        101 | 2024-01-19 |         5800
(3 rows)

sales_olap=# SELECT Product_Id, Region, Sales_Amount FROM sales_sample WHERE Date BETWEEN '2024-01-15' AND '2024-01-17';
 product_id | region | sales_amount
------------+--------+--------------
        101 | East   |         5000
        102 | West   |         6000
        101 | North  |         4500
        103 | South  |         5500
        102 | East   |         7000
        101 | West   |         4800
(6 rows)
```

e) Dice:

Query:

SELECT * FROM sales_sample WHERE Region IN ('East', 'West') AND Date BETWEEN '2024-01-15' AND '2024-01-17' AND Product_Id IN (101, 102);

Output:

```
sales_olap=# SELECT * FROM sales_sample WHERE Region IN ('East', 'West') AND Date BETWEEN '2024-01-15' AND '2024-01-17' AND Product_Id IN (101, 102);
 product_id | region |    date    | sales_amount
------------+--------+------------+--------------
        101 | East   | 2024-01-15 |         5000
        102 | West   | 2024-01-15 |         6000
        102 | East   | 2024-01-17 |         7000
        101 | West   | 2024-01-17 |         4800
(4 rows)
```