# **EcoMart Relational Database Design**

#### A. Scenario 2 - EcoMart

**A1: Business Problem**: The EcoMart company faces challenges in managing and analyzing large-scale sales data across various regions, countries, and product categories. The current manual process is time-consuming, prone to errors, and lacks the scalability to accommodate growing data volumes. A database solution is essential to streamline data storage, enable efficient querying, and generate actionable insights.

**A2: Proposed Data Structure:** The proposed data structure includes three core tables—Categories, Products, and Transactions—each designed to store specific attributes. These tables are interlinked through primary and foreign keys, ensuring data consistency and enabling detailed analysis.

**A3: Justification for Database Solution:** A relational database offers scalability, data integrity, and efficient query performance. It eliminates redundancy through normalization, supports large-scale data processing, and provides a robust platform for advanced analytics. By using PostgreSQL, EcoMart can implement a cost-effective, reliable, and scalable solution.

**A4. Usage of Business Data:** The business data will be used to analyze sales trends by region and product category, monitor stock levels, evaluate revenue performance, and optimize order fulfillment. It will also support decision-making through real-time reporting and analytics.

### **B.** Logical Data Model

The logical data model represents the structure of the database, including entities, attributes, and relationships. It ensures data is organized efficiently, adheres to normalization standards, and supports seamless querying and analysis.

Entities in the model include Categories, Products, and Transactions. Each entity contains attributes tailored to its role, such as category\_name, price, and order\_date. Relationships between tables ensure data integration, such as the one-to-many relationship between Categories and Products.

#### **Entities and Attributes:**

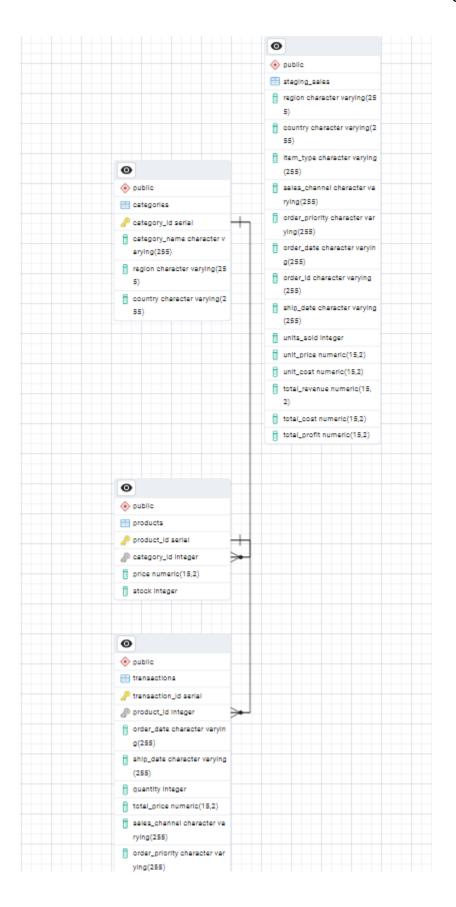
- Categories: category id, category name, region, and country.
- *Products*: product\_id, category\_id, price, and stock.
- <u>Transactions</u>: transaction\_id, product\_id, order\_date, ship\_date, quantity, and total price.

# Relationships:

- Categories → Products: One category can have multiple products.
- Products → Transactions: One product can have multiple transactions.

# Normalization:

The design adheres to 3NF, ensuring no redundant data and maintaining data integrity.



#### C. Database Objects and Storage

The database consists of core objects like tables, indexes, and relationships designed to store and integrate data effectively. Tables represent entities, with attributes optimized for storage efficiency. For instance, "**NUMERIC**" is used for monetary values, while "**VARCHAR**" handles textual data. Indexes are automatically created for primary and foreign keys to optimize query performance.

### D. Scalability

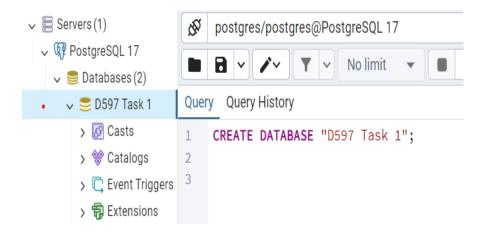
The database design addresses scalability concerns to handle increasing data volumes and user demands. Key strategies include indexing for optimized query performance, particularly on fields like "order\_date" and "region". Normalization minimizes redundancy, enhancing data consistency and storage efficiency. This makes data easier to scale than it was previously, and no need to go through thousands of rows of data.

# E. Privacy and Security

Strong privacy and security measures are built into the database design to ensure compliance with regulations and protect sensitive information. Encryption safeguards data both at rest and in process, preventing unauthorized access. The company can also require audit logging and 2FA as an added layer of security precautions. Additionally, input validation is implemented to protect against SQL attacks, which will assist in maintaining the database's integrity.

## F. Implementation Through pgAdmin

#### F1: Creation of Database:

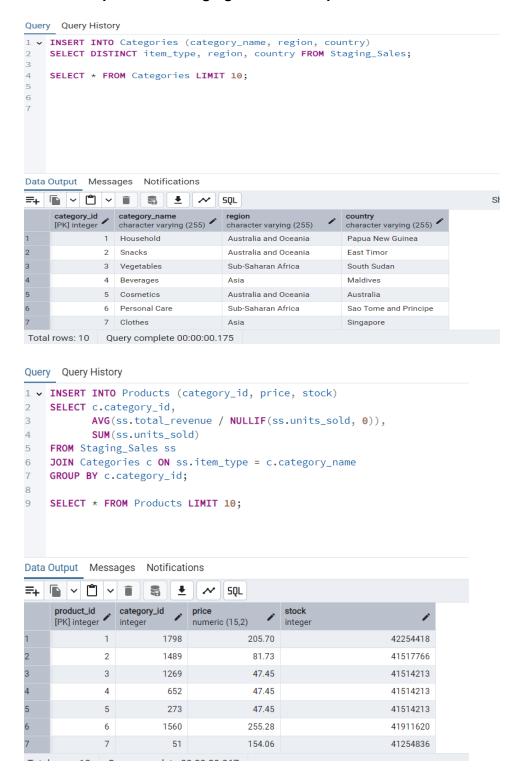


#### F1: Creation of Tables:

```
Query Query History
 1 ➤ CREATE TABLE Staging_Sales (
                  region VARCHAR(100),
                  country VARCHAR(100),
 3
  4
                  item_type VARCHAR(255),
 5
                  sales_channel VARCHAR(255),
                  order_priority VARCHAR(10),
  6
 7
                 order_date DATE,
                 ship_date DATE,
 8
                 units_sold INT,
 9
10
                  total_revenue NUMERIC(15, 2)
11
         );
12
Query Query History
2 V CREATE TABLE Categories (
3 category_id SERIAL PRIMARY KEY,
4 category_name VARCHAR(255),
7 region VARCHAR(255),
           country VARCHAR(255)
9
10 V CREATE TABLE Products (
11 product_id SERIAL PRIMARY KEY,
12 category_id INT REFERENCES Categories(category_id),
13 price NUMERIC(15, 2),
14
15
17

8 V CREATE TABLE Transactions (
19
19 transaction_id SERIAL PRIMARY KEY,
20 product_id INT REFERENCES Products(product_id),
21 order_date VARCHAR(255),
22 ship_date VARCHAR(255),
21
22
23
24
           quantity INT,
total_price NUMERIC(15, 2),
           sales_channel VARCHAR(255)
           order_priority VARCHAR(255)
```

# F2. Data import from "Staging Table" to respective tables:

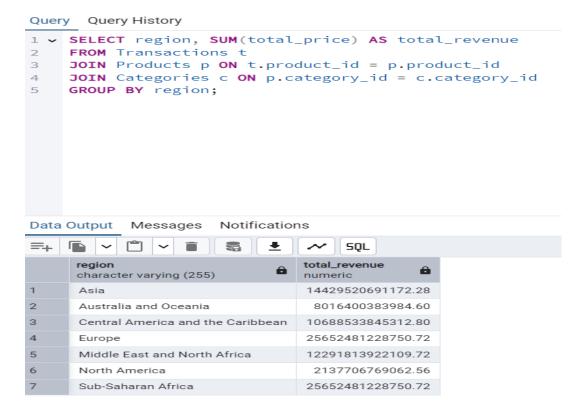


#### Query Query History 1 v INSERT INTO Transactions (product\_id, order\_date, ship\_date, quantity, total\_price, sales\_channel, order\_priority) SELECT p.product\_id, ss.order\_date, ss.ship\_date, 4 ss.units\_sold, 5 ss.total\_revenue, 6 ss.sales\_channel, ss.order\_priority FROM Staging\_Sales ss 10 JOIN Products p ON ss.item\_type = (SELECT category\_name FROM Categories WHERE category\_id = p.category\_id); 11 SELECT \* FROM Transactions LIMIT 10; 12 13 Data Output Messages Notifications =+ 🖺 ∨ 📋 ∨ 🝵 👼 👲 🕢 SQL Showing rows: 1 to 10 Page No: 1 transaction\_id | product\_id | order\_date | character varying (255) | ship\_date character varying (255) variate varying (255) variate varying (255) variate varying character varying (255) 632 12/27/2012 1/30/2013 2910 742864.80 Online М 632 4/26/2010 5/25/2010 587654.56 Online н 2302 3 632 5/5/2011 5/13/2011 6574 1678210.72 Online С 632 7/10/2012 7/22/2012 2213022.32 Online М 5 632 11/24/2015 12/1/2015 1368 349223.04 Offline М 6 632 10/31/2013 11/7/2013 1074473.52 Online С

#### F3: Query Examples:

# F3: Query 1 - Total Revenue by Region:

A query for <u>Total Revenue by Region</u> helps analyze which areas are performing the best or need improvement, so decisions about where to focus resources can be made. It's useful for planning strategies, budgeting, and forecasting future performance, while also identifying potential growth opportunities in certain regions. This type of query can also support compliance, reporting needs, and provide insights for comparing performance against competitors, helping the business stay on track with its goals.



# Before optimization:

```
Query Query History
1 ▼ EXPLAIN ANALYZE
2 SELECT region, SUM(total_price) AS total_revenue
3 FROM Transactions t
4 JOIN Products p ON t.product_id = p.product_id
5    JOIN Categories c ON p.category_id = c.category_id
6 GROUP BY region;
Data Output Messages Notifications
                                                                                                      Showing rows: 1 to 28
                                 . ~
                                            SQL
       QUERY PLAN
      text
21
                      -> Hash (cost=109.60..109.60 rows=6660 width=8) (actual time=1.434..1.435 rows=6660 loops=3)
22
                         Buckets: 8192 Batches: 1 Memory Usage: 325kB
23
                         -> Seq \, Scan \, on \, products \, p \, \, (cost=0.00..109.60 \, rows=6660 \, width=8) \, (actual \, time=0.061..0.749 \, rows=6660 \, loops=3)
                   -> Hash (cost=84.40..84.40 rows=4440 width=20) (actual time=1.124..1.125 rows=4440 loops=3)
24
25
                      Buckets: 8192 Batches: 1 Memory Usage: 301kB
26
                      -> Seq Scan on categories c (cost=0.00..84.40 rows=4440 width=20) (actual time=0.080..0.568 rows=4440 loops=3)
27
       Planning Time: 4.095 ms
       Execution Time: 31408.408 ms
28
 Total rows: 28
                  Query complete 00:00:32.395
```

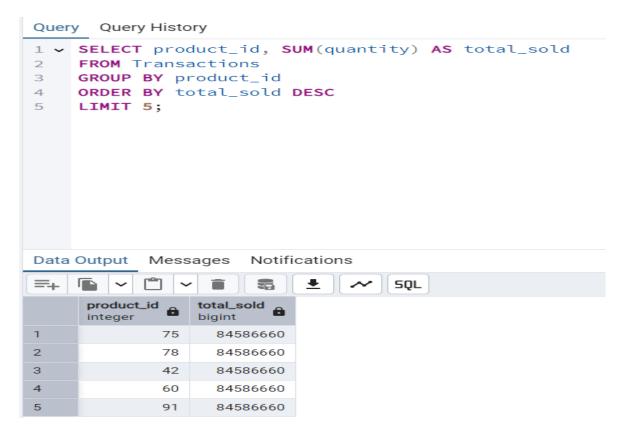
# After optimization:

```
Query Query History
1 - EXPLAIN ANALYZE
2 WITH pre_aggregated AS (
          SELECT product_id, SUM(total_price) AS total_revenue
3
4
          FROM Transactions
5
          GROUP BY product_id
6 )
7 SELECT c.region, SUM(pa.total_revenue) AS total_revenue
8 FROM pre_aggregated pa
9 JOIN Products p ON pa.product_id = p.product_id
JOIN Categories c ON p.category_id = c.category_id
11 GROUP BY c.region;
12
13
Data Output Messages Notifications
    ₹ ~ SQL
                                                                                           Showing rows: 1 to 34
      QUERY PLAN
                                                                                                                â
      text
27
                              Worker 0: Batches: 1 Memory Usage: 1489kB
28
                              Worker 1: Batches: 1 Memory Usage: 1489kB
29
                              -> Parallel Seq Scan on transactions (cost=0.00..1070773.33 rows=30833333 width=12) (actual time=0.504....
30
            -> Hash (cost=84.40..84.40 rows=4440 width=20) (actual time=0.714..0.714 rows=4440 loops=1)
31
              Buckets: 8192 Batches: 1 Memory Usage: 301kB
32
              -> Seq Scan on categories c (cost=0.00..84.40 rows=4440 width=20) (actual time=0.010..0.290 rows=4440 loops=1)
33
      Planning Time: 0.221 ms
      Execution Time: 13983.656 ms
```

Runtime speed changed from 31408.4 ms to 13983.6 ms

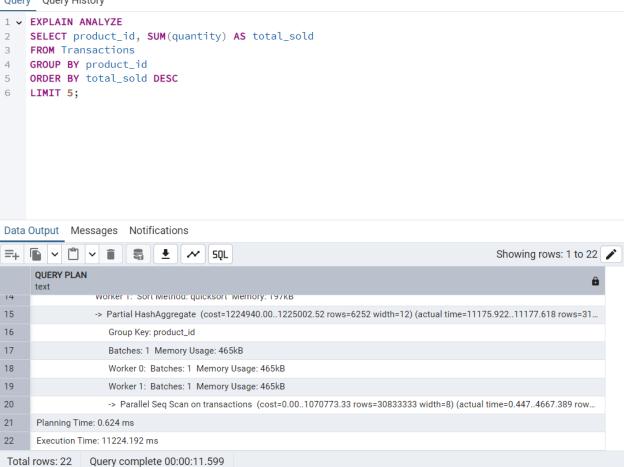
## F3: Query 2 - Top 5 Selling Products

A query for the <u>Top 5 Selling Products</u> helps identify which products are performing the best, giving insight into what customers prefer. This information is useful for making decisions about inventory, marketing, and production, ensuring resources are focused on the most profitable items. It also helps with forecasting future demand, planning promotions, and understanding trends. By knowing your top-performing products, you can drive more sales, increase profitability, and align strategies with what works best for the business.

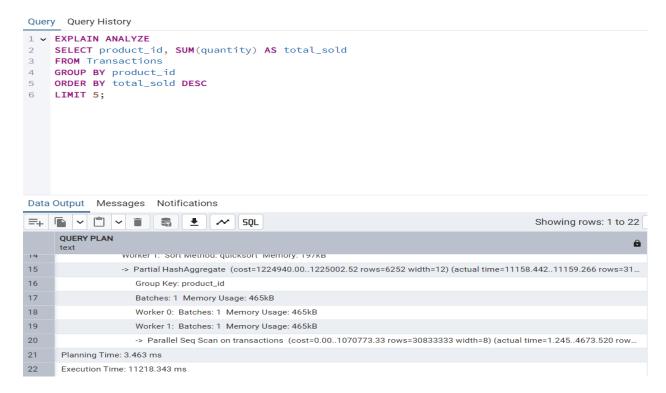


# **Before Optimization:**

Query Query History



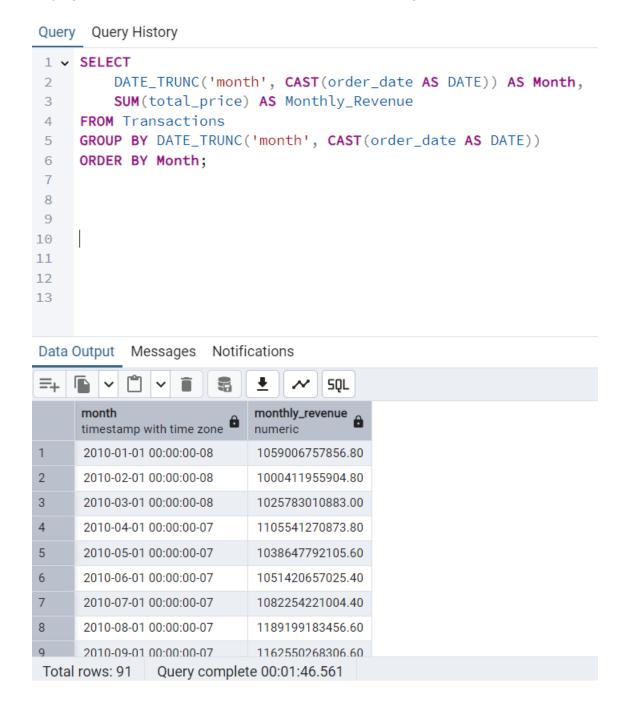
# After optimization:



Runtime speed changed from 11224.1 ms to 11218.3 ms

## F3: Query 3 - Analysis of Monthly Revenues

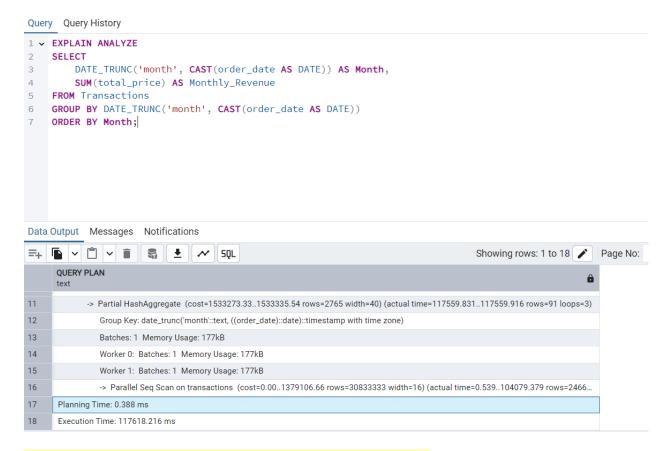
A query for the <u>Analysis of Monthly Revenues</u> helps track how the business is performing over time and spot trends or patterns in revenue. This information is useful for identifying peak sales periods, managing cash flow, and setting realistic goals for growth. It also helps in comparing month-to-month performance, forecasting future revenue, and planning strategies to address any dips or maximize high-performing months. Understanding monthly revenue trends is key to staying on top of the business's financial health and making informed decisions.



# Before optimization:

```
Query Query History
 1 v EXPLAIN ANALYZE
 2
 3
           DATE_TRUNC('month', CAST(order_date AS DATE)) AS Month,
           SUM(total_price) AS Monthly_Revenue
    FROM Transactions
 6 GROUP BY DATE_TRUNC('month', CAST(order_date AS DATE))
 7
      ORDER BY Month;
 8
 9
10
11
12
13
14
Data Output Messages Notifications
=+ | • | • | • | • |
                          $ ± ~
                                          SQL
                                                                                                 Showing rows: 1 to 18 Page 1
      QUERY PLAN
                                                                                                                         ۵
11
             -> Partial \ Hash Aggregate \ (cost=1533273.33..1533335.54 \ rows=2765 \ width=40) \ (actual \ time=120823.803..120823.879 \ rows=91 \ loops=3)
12
                Group Key: date_trunc('month'::text, ((order_date)::date)::timestamp with time zone)
13
               Batches: 1 Memory Usage: 177kB
14
                Worker 0: Batches: 1 Memory Usage: 177kB
15
                Worker 1: Batches: 1 Memory Usage: 177kB
16
               -> Parallel Seq Scan on transactions (cost=0.00..1379106.66 rows=30833333 width=16) (actual time=1.366..106599.275 rows=2466..
17
      Planning Time: 0.104 ms
      Execution Time: 120867.109 ms
Total rows: 18 Query complete 00:02:01.048
```

# After optimization:



Runtime speed changed from 120867.1 ms to 117618.2 ms

# F4: Optimization Techniques:

# Query Query History

```
CREATE INDEX idx_transactions_order_date ON Transactions(order_date);
CREATE INDEX idx_product_id ON Products (product_id);
CREATE INDEX idx_region ON Transactions (region);

Data Output Messages Notifications
```

CREATE INDEX

Query returned successfully in 5 min 49 secs.

# **G**: Panopto Video

### **H: Sources**

No sources, web or otherwise, were used in the creation of the project and its follow up report.