OmniFoods Energy Bar Sales Analysis Using OLAP

Contents

1 Introduction

OmniFoods is a global leader in food products. Now they are focusing on launching a new range of high-energy bars under the brand name OmniPower. These energy bars are becoming popular among other groups of people in recent times, which until now were kept restricted to athletes and sports persons. Using an efficient and robust data analysis solution, the company can check its sales performance, store profitability, and promotional impact on marketing related to energy bars at OmniFoods.

In alignment with these objectives, the company has implemented an **Initial DW integrated with OLAP** and visualization tools to gain actionable insights.

2 Motivation

Having sales and operational data across 34 stores analyzed may prove challenging as the data already exists in transactional systems designed for decision-making. The motivation for this project includes the following:

- **Problems with Current Systems**:
  - Only very basic reports can be generated using transactional databases.
  - Multidimensional queries, such as total revenue by time, store, and product, are complex to achieve manually.
- **Requirements of the New System**:
  - **Decision Making**: The system should enable business managers to project sales trends, regional performance, and the effectiveness of promotions with ease.
  - **Multidimensional Analyses**: Ability to slice and dice data from various perspectives like time, geography, and promotion.
  - **Informed Strategy**: Optimize inventory, improve marketing campaigns, and plan for expansion based on detailed analytics.

3 Theoretical Review

3.1 Business Intelligence (BI)

Business Intelligence refers to the technologies, tools, and processes used to analyze business data and provide actionable insights. It combines data integration, analysis, and visualization to aid decision-making. BI systems often use dashboards and KPIs to present information clearly to stakeholders.

3.2 Data Warehousing

A Data Warehouse (DWH) is a central repository designed to store and organize data for analytical purposes. DWHs often use schemas (Star, Snowflake, or Fact Constellation) to structure data efficiently for Online Analytical Processing (OLAP).

3.3 Key Performance Indicators (KPIs)

KPIs are measurable metrics that evaluate the performance of specific business operations. For instance, metrics like sales by region or product performance can guide strategic decisions.

4 Functional Requirements

4.1 Data Integration

- The system must be able to extract data from multiple sources, including: Sales data, Customer information, Territory data, etc.
- The system should support the transformation of raw data into a consistent format for storage and analysis.
- The system must load the cleaned and formatted data into a centralized Data Warehouse (DWH).

4.2 Data Warehousing and Modeling

- The system must create and maintain a Data Warehouse using a Star Schema, consisting of:
  - A Fact Table
  - Dimension Tables
- The system must support the use of Online Analytical Processing (OLAP) for querying and analyzing the data.

4.3 Data Analysis and Visualization

- The system must generate a BI dashboard to display KPIs, reports, and visualizations.
- The dashboard must allow interactive filtering and drilling down.
- The dashboard must support visual elements such as: Bar charts, line graphs, and pie charts, etc.

5 Non-Functional Requirements

5.1 Performance

- The system must handle large amounts of data without slowing down.
- Reports and dashboards should load quickly.

5.2 Scalability

- The system should be able to grow with the business, handling more data as new stores or products are added.
- The Data Warehouse should be able to store increasing amounts of data over time.

5.3 Reliability

- The system should be available 24/7 with minimal downtime.
- It must back up data regularly and be able to recover from failures quickly.

5.4 Usability

- The system must be easy to use, even for people who aren't tech experts.
- Dashboards should be clear, and users should be able to customize what they see.

5.5 Maintainability

- The system should be easy to update and maintain.
- It should be well-documented, and components should be modular so they can be updated without affecting the whole system.

5.6 Compatibility

- The system should work on common operating systems and web browsers.
- It should be able to import and export data in standard formats like CSV, Excel, and JSON.

6 Background Theory

6.1 Data Warehouse Architecture

OmniFoods' data warehouse uses the **three-tier architecture**:

- **Bottom Tier**: MySQL stores integrated data from multiple sources, including sales transactions, promotional data, and store information.
- **Middle Tier**: An OLAP engine processes queries using ROLAP.
- **Top Tier**: Power BI creates user-level dashboards and visuals.

6.2 Data Gathering

In the first phase of the project, the goal was to gather raw data from disparate sources. The data for this project was primarily extracted from Excel and CSV files. These files contained important business data such as sales information, customer details, product information, and sales territories. These data sources represented the "original" form of the data, before any processing or transformation took place.

This phase was critical for establishing a foundation of raw data that could be further processed and analyzed in the subsequent phases of the project.

6.3 Data Preparation / ETL or ELT Process

Once the data was gathered in its raw form, the next step involved transforming it into a structured format suitable for analysis. This phase is known as the ETL (Extract, Transform, Load) process.

6.3.1 Data Extraction (E)

The raw data was extracted from the CSV files. This step involved reading the files and storing them in a DataFrame for easy manipulation and analysis.

6.3.2 Data Transformation (T)

- The data was cleansed by removing duplicates, handling missing values, and correcting inconsistencies.
- Data types were adjusted to ensure proper compatibility and accuracy.
- In some cases, certain fields were transformed to match the required structure for the Data Warehouse (DWH). For example, date formats were standardized, and certain numerical fields were converted for consistency.

6.3.3 Data Loading (L)

- After cleaning and transforming the data, the next step was to load the data into the Data Warehouse (DWH).
- The DWH was structured with a star schema, and data was placed into fact tables (e.g., sales data) and dimension tables (e.g., customer and product data).
- This step involved mapping the transformed data into the correct tables within the DWH to ensure that it was ready for querying and analysis.

By the end of this phase, the data was in a structured format and stored in the Data Warehouse, ready to be used for further analysis and reporting.

6.4 OLAP Architecture

For OLAP (Online Analytical Processing), we will use **ROLAP** (Relational OLAP). This method uses relational databases (SQL-based queries) to perform the analysis. It allows for flexible querying and faster query processing, as the data is stored in relational tables rather than a pre-aggregated OLAP cube.

- Dimension tables: These hold the descriptive attributes of store, time, and promotion.

The ROLAP approach will be useful for this project because we are querying from relational tables in a Data Warehouse using SQL. The proposed architecture facilitates OmniFood in performing multi-dimensional analysis.

7 Proposed System

7.1 Overview

The proposed system provides OmniFood with the analytical capability to do sales forecasting against Budget/Plan data and:

- Analyze sales trends over time (monthly, yearly).
- Identify best and worst-performing regions.
- Quantify revenue impact of promotions.
- Calculate ROI of marketing campaigns.
- Benchmark stores on traffic and revenue measures.

7.2 System Flow

1. **Extract** data from CSV files into staging tables.
2. **Transform** data to resolve inconsistencies, enhance information, and calculate derived measures.
3. **Load** transformed data into the star schema.
4. **Analyze** data by OLAP and display results in Power BI.

sectionStar Schema

7.3 Schema Design

The star schema for the OmniFoods Data Warehouse includes a central fact table and four associated dimension tables. Below is the diagram representing the star schema:

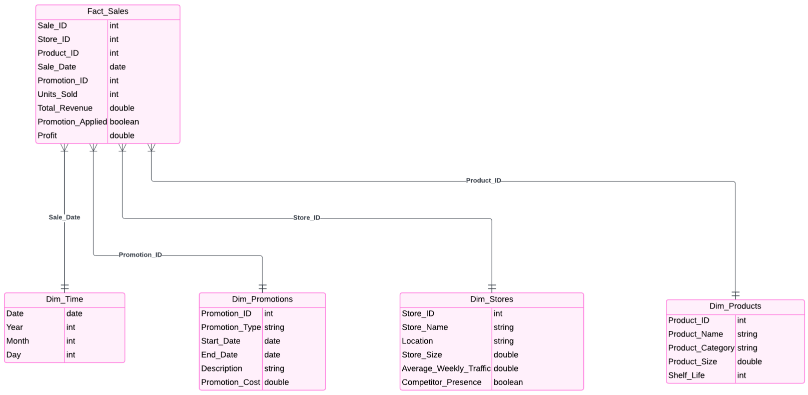

Figure 1: Star Schema for OmniFoods Data Warehouse

7.4 Explanation

- **Fact_Sales**: Contains transactional data such as `Units_Sold`, `Total_Revenue`, and `Promotion_Applied`.
- **Dim_Stores**: Includes store-related attributes such as `Location`, `Store_Size`, and `Traffic`.
- **Dim_Products**: Describes products, including `Product_Name`, `Category`, and `Shelf_Life`.
- **Dim_Time**: Breaks down sales dates into `Year`, `Month`, and `Day`.
- **Dim_Promotions**: Tracks promotional campaigns with fields like `Promotion_Type`, `Start_Date`, and `Cost`.

8 Implementation

8.1 ETL Process

Step 1: Extract

Data from various CSV files, such as `sales.csv`, `stores.csv`, and `promotions.csv`, is extracted to staging tables in MySQL.

Step 2: Transform

- Clean data:
  - Replace `NULL` values in `Total_Revenue` with calculated values (`Units_Sold * Price_Per_Unit`).

- Standardize text fields (e.g., all location names are in uppercase).

- Enrich data by adding derived fields such as `Profit` and `Promotion_Effectiveness`.
- Normalize data by separating transactional data into a fact table and descriptive data into dimension tables.

Step 3: Load The cleaned and transformed data is loaded into the star schema.

8.2 OLAP Queries

Query 1: Total Revenue by Store

```
SELECT Store_Name, SUM(Total_Revenue) AS Total_Revenue
FROM fact_sales
JOIN dim_stores ON fact_sales.Store_ID = dim_stores.Store_ID
GROUP BY Store_Name
ORDER BY Total_Revenue DESC;
```

Query 2: Revenue Trends by Time

```
SELECT Year, Month, SUM(Total_Revenue) AS Total_Revenue
FROM fact_sales
JOIN dim_time ON fact_sales.Sale_Date = dim_time.Date
GROUP BY Year, Month
ORDER BY Year, Month;
```

Query 3: Promotion Effectiveness

```
SELECT Promotion_Type, SUM(Total_Revenue) AS Total_Revenue
FROM fact_sales
JOIN dim_promotions ON fact_sales.Promotion_Applied = 'Yes'
GROUP BY Promotion_Type
ORDER BY Total_Revenue DESC;
```

8.3 Results

Monthly Sales Analysis

Month	**Units Sold**	**Total Revenue**

January	1,200	$6,000
February	1,400	$7,000
March	1,800	$9,500

Store Performance Analysis

Store Name	Region	Total Revenue
Store 12	East	$12,500
Store 8	West	$11,200

9 Evaluation

In this section, we evaluate the overall success of the project, focusing on how well it met the objectives, the effectiveness of the solution, challenges encountered, and areas for improvement.

9.1 Success in Meeting Objectives

The primary goal of this project was to integrate multiple data sources, transform and store them in a Data Warehouse (DWH), and analyze the data using OLAP queries to support strategic decision-making. This objective was successfully achieved through the following:

- Data was extracted from Excel and CSV files, cleaned, transformed, and loaded into the DWH.
- A star schema was designed, with fact and dimension tables enabling effective and efficient querying.
- OLAP queries were performed to derive actionable insights, presented through an interactive Power BI dashboard for easy interpretation.

9.2 Effectiveness of the Solution

User Experience: The Power BI dashboard provided a user-friendly interface, enabling users to explore data through intuitive visualizations and easily view KPIs.

Data Accessibility: The use of SQL queries to interact with the Data Warehouse allowed flexible and fast access to the required data, making analysis seamless and efficient.

Business Value: The visualizations and reports helped identify patterns, trends, and areas needing attention. These insights empower management to make more informed decisions, such as optimizing product offerings, targeting new markets, and refining marketing strategies.

9.3 Challenges Encountered

Data Quality Issues: During the data extraction and cleaning process, inconsistencies and missing values in the raw data posed challenges. Resolving these issues required extra effort to ensure data completeness and accuracy.

System Integration: Integrating data from multiple sources required careful mapping and transformation to align with the Data Warehouse schema. Some sources needed additional formatting to match the expected structure.

Performance of OLAP Queries: Some OLAP queries experienced slower-than-expected execution times, particularly for complex aggregations. Optimizing query performance in the Data Warehouse environment proved challenging.

9.4 Areas for Improvement

Data Security: The project did not implement advanced data security measures, such as encryption or user access control. In a real-world application, these measures would be crucial to protect sensitive company data.

Real-Time Data Integration: Currently, the system is based on static data extracted from CSV and Excel files. Incorporating real-time data integration, such as connecting directly to live databases, would enhance the system's dynamism and provide up-to-the-minute insights.

Scalability: While the current setup is sufficient for the project's scope, a scalable solution is necessary for larger datasets and increased query loads, particularly in a production environment.

In summary, this project effectively met its objectives and provided a robust solution for data-driven decision-making. However, addressing the areas for improvement would ensure greater reliability, efficiency, and long-term scalability of the system.

10 Conclusion

This Business Intelligence (BI) project has been an insightful and rewarding experience, allowing us to work through all stages of a BI solution: from data extraction to final visualization. The overall goal was to provide a decision-support system that would empower OmniFoods to make data-driven strategic decisions regarding inventory, store locations, and marketing initiatives.

The key strengths of this project lie in:

- Identifying sales trends across time, regions, and stores.
- Measuring the effectiveness of marketing promotions.
- Improving overall sales strategy and profitability.
- Successfully integrating multiple data sources into a structured and efficient Data Warehouse.
- Providing a user-friendly interface through Power BI to deliver actionable insights.

Challenges encountered during the data preparation phase, particularly with data quality and system integration, were resolved through careful cleaning and transformation of the data.

Looking ahead, future extensions could include:

- Enhancing data security to ensure robust protection of sensitive information.
- Integrating real-time data feeds for more dynamic and up-to-date analytics.
- Improving scalability to handle larger datasets and support live business environments.
- Incorporating predictive analytics to anticipate trends and improve decision-making further.

In conclusion, this BI project successfully delivered a comprehensive solution tailored to OmniFoods' needs and demonstrated the value of data-driven decision-making. While the project achieved its goals, there are opportunities to refine and enhance the system to meet future challenges and business growth.