

|  |
| --- |
| Business Template  **supermarket sales data** |
| 130+ Thousand Supermarket Logo Royalty-Free Images, Stock Photos & Pictures  | Shutterstock |

Contents

[1 Business Description 3](#_Toc201265571)

[1.1 Business background 3](#_Toc201265572)

[1.2 Problems because of poor data management 3](#_Toc201265573)

[1.3 Benefits from implementing a Data Warehouse 3](#_Toc201265574)

[1.4 DATASETS DESCRIPTION 3](#_Toc201265575)

[1.4.1 Dataset 1: Online Orders 3](#_Toc201265576)

[1.4.2 Dataset 2: Offline Orders 3](#_Toc201265577)

[1.5 GRAIN / DIM / FACT 4](#_Toc201265578)

[2 Business Layer 3NF 4](#_Toc201265579)

[3 Business Layer Dimensional Model 4](#_Toc201265580)

[4 Logical Scheme 4](#_Toc201265581)

[5 Data Flow 4](#_Toc201265582)

[6 Fact Table Partitioning Strategy 4](#_Toc201265583)

# 

# Business Description

## Business background

The business operates in the **retail industry**, specifically selling consumer products across various city branches. The organization handles both **online and offline orders**, with a broad product portfolio classified into categories and sold through multiple store locations. The company employs sales staff, offers occasional discounts, and operates in a competitive environment where timely and accurate reporting of sales data is crucial to drive strategic decisions.

## Problems because of poor data management

The company has faced several challenges due to fragmented and inconsistent data management practices. The lack of a unified reporting system has resulted in discrepancies between online and offline sales data, complicating the process of performance evaluation and financial forecasting. Manual changing of records not only delays strategic decision-making but also increases the risk of human error. Also, the business has struggled to gain insights into key performance metrics such as cost efficiency, revenue generation, and discount impact. Without integrated customer and product-level analytics, it is difficult to tailor offerings, track performance, and stay ahead of the competition.

## Benefits from implementing a Data Warehouse

Implementing a data warehouse will significantly enhance the company's ability to collect, store, and analyze data from both operational systems. With structured access to consistent and cleansed sales and cost data, decision-makers will be able to track profitability, identify high-performing products and branches, and monitor employee performance across locations. The warehouse will support time-based, product-based, and geography-based reporting and allow for the exploration of patterns in customer purchases. Hierarchical structures within the data, such as those linking product categories to individual items and time-based groupings like year, month, and day, will provide valuable multidimensional views. Ultimately, the business will benefit from improved forecasting, targeted marketing strategies, and optimized inventory management.

## DATASETS DESCRIPTION

### ****Dataset 1: Online Orders****

Includes all orders placed through the company's e-commerce platform. The dataset contains:

* **Product Information:**

Item Code, Item Name, Category Code, Category Name, Wholesale Price, Loss Rate

* **Sales Facts:**

Date, Time, Quantity Sold, Unit Selling Price, Discount, Total Sales, Cost, Gross Income

* **Organizational Data:**

Employee ID, Branch, City

* **Customer Information:**

Simulated using Customer IDs

* **Other:**

Transaction ID, Source System = "Online", Date hierarchy (Year, Month, Day)

### ****Dataset 2: Offline Orders****

Includes all orders placed in company’s physical branches

* **Product Information:**

Item Code, Category Code, Wholesale Price, Loss Rate

* **Sales Facts:**

Date, Time, Quantity Sold, Unit Selling Price, Discount, Total Sales, Cost, Gross Income

* **Organizational Data:**

Employee ID, Branch, City

* **Customer Information:**

Customer ID

* **Other:**

Transaction ID, Source System = "Offline", Date hierarchy (Year, Month, Day)

The two datasets differ in several important ways. The online dataset contains complete information for all attributes, while the offline version is limited and that shows the lack of advanced data systems. Additionally, transaction IDs are uniquely prefixed to distinguish between the two sources, and the structure of certain dimensions varies to reflect differing system designs. Despite these differences, both datasets share common dimensions, such as date, customer, and geography, which will allow them to be merged in the data warehouse.

In this dataset we have several entities, such as

1. **Product**

Which represents items available for sale. Includes product codes, names, categories, pricing, and expected loss rates. Used to analyze what is being sold.

1. **Customer**

Which represents the buyer.

1. **Date**

Which represents the transaction date and supports hierarchical analysis by day, month, and year. Enables time-based trend reporting.

1. **Employee**

Which represents staff responsible for handling sales (online processing or in-store). Helps assess staff performance and assignment.

1. **Branch**

Which represents physical store locations and helps us to analyze geographic trends and branch-level performance.

1. **Channel**

That distinguishes the source of the sale — either **Online** or **Offline** — helping compare e-commerce vs. physical store performance.

1. **Sales Facts**

That is the central numeric data of each transaction: quantity sold, unit price, discounts, total revenue, cost, and gross income. Tied to all dimensions for complete analysis.

## GRAIN / DIM / FACT

# Business Layer 3NF

# Business Layer Dimensional Model

# Logical Scheme

# Data Flow

# Fact Table Partitioning Strategy