

# Analysis and Summary of Product Reviews Project

---

## Report

### Designing a Fundamental Data Warehouse Schema for E-commerce

---

Instructor: Nguyễn Ngọc Thiện  
Implementer: Nguyễn Phúc Nhân

Sep 2024, Ho Chi Minh City



## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Objectives . . . . .	2
1.2	Project Scope . . . . .	2
<b>2</b>	<b>Data Warehouse Architecture and Implementation</b>	<b>2</b>
2.1	Dimension Modeling . . . . .	2
2.1.1	Fact Table: . . . . .	2
2.1.2	Dimension Table: . . . . .	2
2.1.3	Star Schema vs. Snowflake Schema in Dimension Modeling . . . . .	3
2.1.4	Benefits of Dimension Modeling: . . . . .	3
2.2	Slowly Changing - Dim Table . . . . .	3
2.3	Event Record - Fact Table . . . . .	5
2.4	Model Diagram . . . . .	7



# 1 Introduction

## 1.1 Objectives

The objectives of project are outlined below:

- Design and implement data warehouse to centralize and streamline data from an e-commerce platform
- Analyze product reviews to understand customer sentiment and product performance, driving business strategies and improving customer satisfaction
- Use data to train language models, analyze customer reviews for deeper insights into user feedback.

## 1.2 Project Scope

- The scope of the project encompasses a case study focused on analyzing and summarizing user feedback on products from an e-commerce platform.
- This will help users save time by providing concise summaries of multiple reviews, allowing them to quickly assess product quality and make informed purchasing decisions.

# 2 Data Warehouse Architecture and Implementation

## 2.1 Dimension Modeling

**Dimension modeling** is a database design methodology used in data warehouses to organize and structure data for easy querying and analysis. It is widely adopted in decision support systems, especially for reporting and data analysis.

Dimension modeling primarily consists of two types of tables:

### 2.1.1 Fact Table:

- Contains quantitative data related to business events, such as sales, revenue, and product quantities sold.
- Each record in the fact table typically contains foreign keys that link to dimension tables.
- The data in fact tables is usually numerical, like sales amounts, revenue, or customer ratings.

### 2.1.2 Dimension Table:

- Contains descriptive attributes that provide context for the data in the fact table. These attributes are relatively static and provide qualitative information.



- Examples include product information (name, category, brand), customer details (name, region, registration date), or time (day, month, quarter, year).

### 2.1.3 Star Schema vs. Snowflake Schema in Dimension Modeling

Dimension modeling is usually organized in either a **star schema** or a **snowflake schema**:

- **Star Schema**: This is the simplest model, where the fact table is at the center, and the dimension tables surround it. The fact table is linked directly to each dimension table through foreign keys.

- **Snowflake Schema**: This is a more complex version of the star schema, where the dimension tables are further normalized into smaller tables, creating a structure that resembles a snowflake.

### 2.1.4 Benefits of Dimension Modeling:

- **Easy Querying**: Dimension modeling is optimized for querying and reporting, making it easier to run complex queries.
- **High Performance**: The structure of dimension modeling improves performance in analytical systems.
- **Intuitive and Understandable**: Designed for end users, dimension modeling is user-friendly, making it easier to navigate without deep technical knowledge of the database.

Dimension modeling is a powerful tool for organizing and analyzing large datasets in data warehouses, making it particularly useful for businesses aiming to derive insights from complex data.

## 2.2 Slowly Changing - Dim Table

Name of Table	Attributes	Purpose
---------------	------------	---------



Name of Table	Attributes	Purpose
dim_Product	<ul style="list-style-type: none"><li>• product_id</li><li>• name</li><li>• description</li><li>• category_id</li><li>• brand_id</li><li>• price</li><li>• discount</li><li>• rating_average</li><li>• review_count</li></ul>	This table stores comprehensive details about each product, including its category and brand affiliations. By analyzing this data, businesses can gain insights into product performance, customer preferences, and the impact of pricing strategies on sales and reviews.
dim_Customer	<ul style="list-style-type: none"><li>• customer_id</li><li>• name</li><li>• full_name</li><li>• region</li><li>• created_time</li><li>• joined_time</li><li>• total_reviews</li><li>• total_thanks</li></ul>	Contains customer information, allowing segmentation and analysis of customer behavior and demographics.
dim_Seller	<ul style="list-style-type: none"><li>• seller_id</li><li>• name</li><li>• is_best_store</li></ul>	Holds seller details, facilitating the evaluation of seller performance and reputation.



Name of Table	Attributes	Purpose
dim_Category	<ul style="list-style-type: none"><li>• category_id</li><li>• name</li></ul>	Categorizes products, aiding in the analysis of product categories.
dim_Brand	<ul style="list-style-type: none"><li>• brand_id</li><li>• name</li></ul>	Stores brand information, enabling brand performance analysis.
dim_Date	<ul style="list-style-type: none"><li>• date_id</li><li>• date</li><li>• day</li><li>• month</li><li>• year</li><li>• quarter</li><li>• is_weekend</li><li>• is_holiday</li></ul>	Provides date-related attributes for time-based analysis.

### 2.3 Event Record - Fact Table

Name of Table	Attributes	Purpose
---------------	------------	---------



Name of Table	Attributes	Purpose
fact_Comment	<ul style="list-style-type: none"><li>• comment_id</li><li>• review_id</li><li>• customer_id</li><li>• content</li><li>• score</li><li>• is_reported</li></ul>	Stores comments on reviews, allowing sentiment analysis and quality control of reviews.
fact_Sentiment	<ul style="list-style-type: none"><li>• sentiment_id</li><li>• review_id</li><li>• sentiment_type</li></ul>	Captures sentiment analysis results, aiding in understanding customer sentiment.
fact_Sales	<ul style="list-style-type: none"><li>• sale_id</li><li>• product_id</li><li>• customer_id</li><li>• seller_id</li><li>• date_id</li><li>• quantity</li><li>• price</li><li>• discount</li><li>• total_amount</li></ul>	Records sales transactions, enabling sales performance analysis.



Name of Table	Attributes	Purpose
fact_Reviews	<ul style="list-style-type: none"><li>• review_id</li><li>• product_id</li><li>• customer_id</li><li>• seller_id</li><li>• date_id</li><li>• rating</li><li>• content</li><li>• helpful_votes</li><li>• helpful_count</li><li>• summary_id</li></ul>	Stores product reviews, facilitating detailed review analysis.
fact_ReviewSummary	<ul style="list-style-type: none"><li>• summary_id</li><li>• product_id</li><li>• summary_text</li><li>• positive_points</li><li>• negative_points</li><li>• average_rating</li><li>• total_reviews</li><li>• last_updated</li></ul>	Summarizes reviews, providing a consolidated view of product feedback.

## 2.4 Model Diagram

The diagram below illustrates the star schema used in the product review analysis data mart for an e-commerce platform.

The fact tables capture key transactions and review data, while the dimension tables provide descriptive information necessary for analysis, such as product details, customer data, and dates.



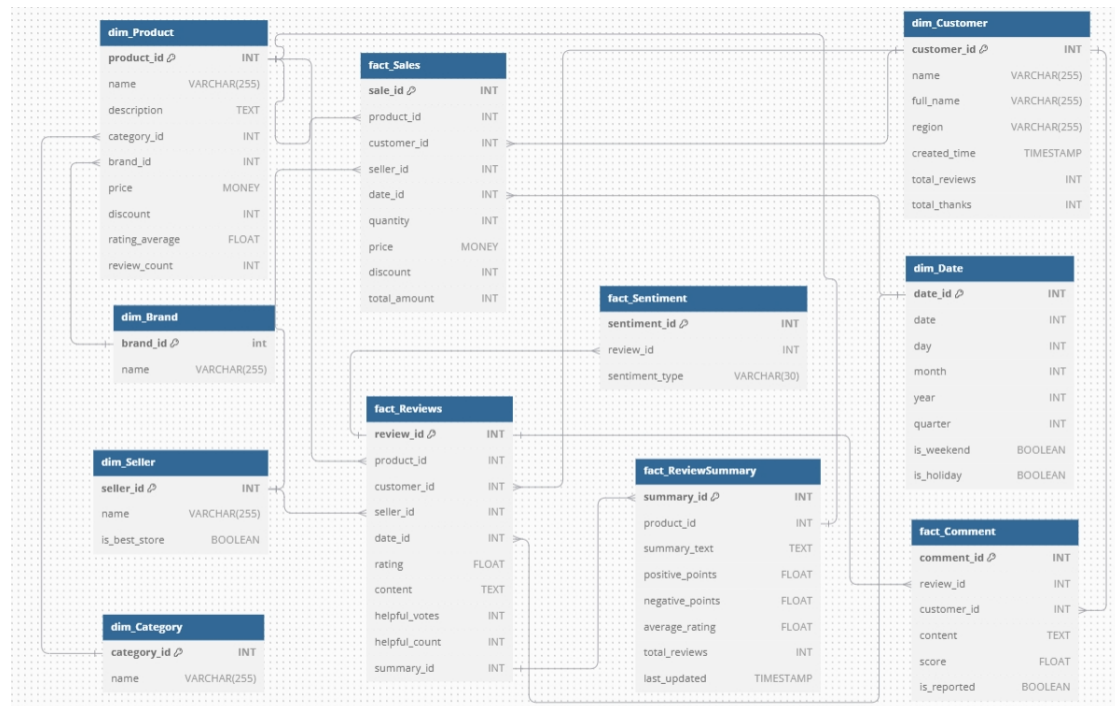


Figure 1: Star Schema Model for E-commerce Product Review Analysis

## References

- [1] [Data warehousing and business intelligence project report](#)
- [2] [Data Warehouse Project - Sunny U Okoro](#)