



Introduction to Dimensional Data Modeling

Seckin Dinc · [Follow](#)

7 min read · Jul 21



31



1

Photo by [Alexander Schimmeck](#) on [Unsplash](#)

Dimensional data modeling is a popular approach to designing databases that are optimized for reporting and analysis. This technique focuses on organizing data into a structure that is intuitive and easy to understand,

allowing users to quickly and efficiently analyze large amounts of information.

This article will explore the basics of dimensional data modeling, including its key concepts, advantages, and applications, providing a foundation for understanding this essential data management technique.

What is Dimensional Data Modeling?

Dimensional data modeling is a technique used to design databases that are optimized for reporting and analysis. It involves organizing data into a structure that is easy to understand and use, allowing users to quickly and efficiently analyze large amounts of information. The dimensional model is designed to contain the same data as a normalized model. However, it presents the information in a user-friendly format that ensures efficient querying and flexibility to adapt to changes.

By organizing data in this way, dimensional models can provide users with a powerful tool for analyzing complex data sets and gaining valuable insights into their business operations. This approach is widely used in business intelligence and data warehousing applications and has become increasingly popular in recent years due to the explosion of data available for analysis in many industries.

Some of the advantages of the dimensional data model include;

- 1. Simplified Querying:** Dimensional data modeling enables faster and simpler querying of data by providing an intuitive and easy-to-understand schema. This allows business users to analyze and understand the data better, without having to involve IT specialists.

- 2. Faster Performance:** Since dimensional data models are optimized for querying and analysis, they can provide much faster performance compared to traditional normalized models. This is because data is organized in a denormalized structure, which eliminates the need for complex joins.
- 3. Flexibility:** The dimensional data model is highly flexible and can easily adapt to changes in business requirements. It can also incorporate data from multiple sources, making it easier to consolidate and analyze data from across the organization.
- 4. Improved Data Quality:** By simplifying the data structure and reducing the need for complex joins, the dimensional data model can improve data quality by reducing errors and inconsistencies.
- 5. Better Business Insights:** By providing a clear and concise view of the data, the dimensional data model can help organizations gain better business insights and make more informed decisions. It can also enable better tracking of key performance indicators (KPIs) and help identify trends and patterns in the data.

Star Schema

The Star Schema is a popular dimensional modeling technique used in data warehousing to represent data in a simple, easy-to-understand manner. It is named after its resemblance to a star, with the fact table at the center and the dimension tables surrounding it, like the rays of a star.

In a Star Schema, the fact table contains the primary measures or metrics of interest, while the dimension tables provide the context for these measures. Each dimension table represents a specific aspect or attribute of the data, such as time, geography, or product.

The Star Schema is denormalized, meaning that data is duplicated across the dimension tables to eliminate the need for complex joins, which can slow down queries. This makes it much easier and faster to query and analyze data.

One of the key benefits of the Star Schema is its simplicity and ease of use. It provides a clear, intuitive view of the data that is easy for business users to understand and analyze, without requiring extensive technical knowledge or expertise. This makes it a popular choice for data warehousing projects, particularly those focused on business intelligence and reporting.

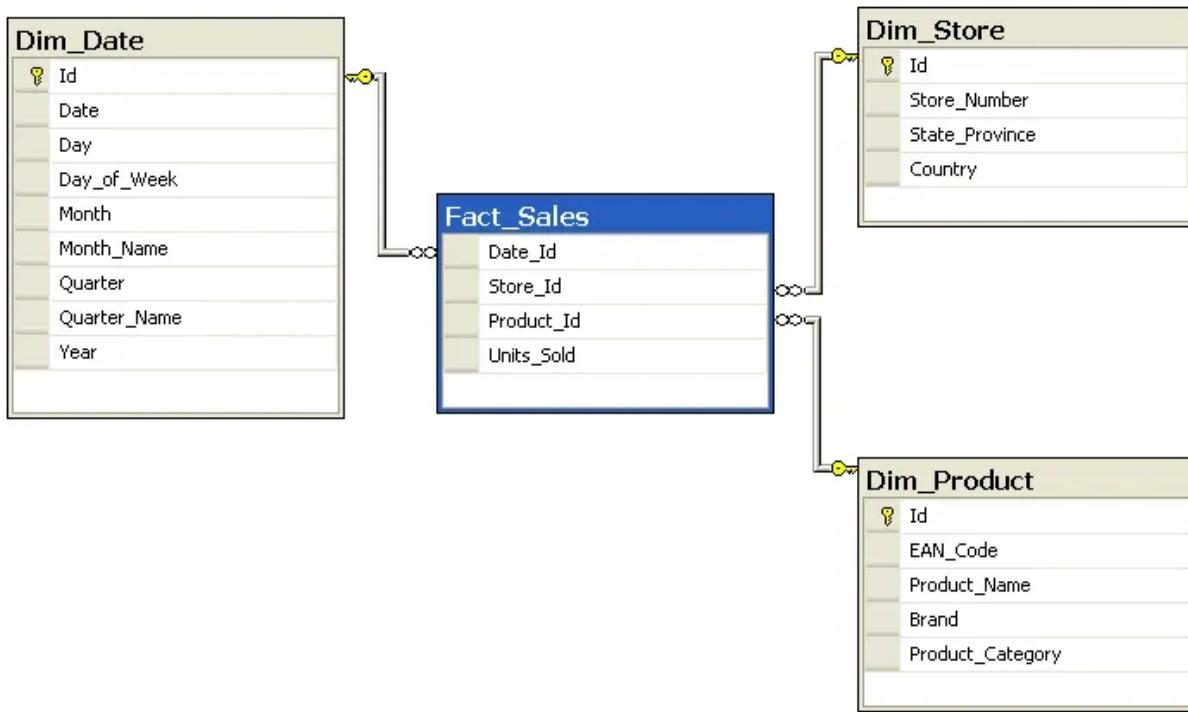


Image courtesy https://en.wikipedia.org/wiki/Star_schema

Fact and Dimension Tables



Photo by [Yulia Matvienko](#) on [Unsplash](#)

Fact tables and dimension tables are two key components of Kimball's dimensional data model. You can easily think of them as Batman and Superman in a DC movie. Without the other, they can't exist in this modeling approach.

A fact table contains the measures or metrics of interest that represent a specific business process, such as sales revenue, units sold, or customer orders. Each row in the fact table represents a single event or transaction that has occurred, and each column represents a different attribute or dimension of the event, such as date, product, or customer.

Dimension tables, on the other hand, provide context for the measures in the fact table. They contain the descriptive attributes or dimensions that provide additional information about the event or transaction recorded in the fact table. Examples of dimension tables include the date, geography, product, and customer.

The relationship between the fact table and dimension tables is established through a set of foreign keys, which are used to join the tables and retrieve the relevant information.

One Big Table (OBT)



Photo by [Sigmund](#) on [Unsplash](#)

One Big Table is a design approach where all the data from various source systems is combined into a single large table within the data warehouse. This table contains all the relevant information needed for analytical reporting and decision-making.

The One Big Table design is in contrast to the traditional approach of creating multiple normalized tables with complex relationships between

them, as seen in a typical relational database. In the One Big Table approach, data is denormalized, meaning that redundant data is stored within the table to avoid the need for joins and complex queries.

Key characteristics of the One Big Table approach

Denormalization: The data is stored in a flattened format, which means that instead of splitting it into multiple tables with relationships, it's all combined into one table. This can lead to data duplication, but it simplifies querying.

Simplified querying: Since all the required data is in a single table, there is no need for complex joins across multiple tables. This can make queries faster and more straightforward.

Performance: In some cases, the One Big Table design can lead to improved query performance, especially for analytical workloads, as it reduces the overhead of joining multiple tables.

Data transformation: Before loading the data into the One Big Table, some data transformation and cleansing might be necessary to ensure data consistency and accuracy.

Easier ETL (Extract, Transform, Load) processes: The ETL processes are often simpler and more efficient since there is no need to manage multiple tables and their relationships.

Start Schema vs One Big Table

1. Star Schema:

Pros:

- **Organized and Tidy:** The Start Schema brings order to the data universe! It divides data into different tables, each with its own role and relationships. So clean, so neat!

- **Reduces Data Duplication:** By breaking data into separate tables, you avoid duplicating information and save precious storage space. Efficiency at its finest!
- **Easy to Maintain:** Need to make changes? No sweat! Updating specific tables is a piece of cake, without disturbing the whole warehouse.

Cons:

- **Complex Joins:** Brace yourself for the SQL acrobatics! To get the full picture, you'll have to perform daring joins between tables. Sometimes, it feels like a wild circus act!
- **Performance Hit:** With all those joins, the Start Schema can slow down like a snail on a hot day. Analytical queries might take longer to execute.
- **Schema Evolution Hassles:** As your data evolves, altering the schema can be a bit of a headache. Changes may ripple through multiple tables, like a domino disaster!

2. One Big Table:

Pros:

- **Simplicity Rules:** Welcome to the land of simplicity! One Big Table keeps all data in a single place. No complicated joins or relational drama.
- **Faster Queries:** Time for a speed boost! With data in one place, queries become lightning-fast, and your analytics can race ahead.
- **Easy Peasy ETL:** Loading data is a breeze! No need to juggle multiple tables; you just plop everything into One Big Table.

Cons:

- **Data Duplication Galore:** Watch out for those duplicates! Storing everything in one place might lead to redundant data. Dups are like

party crashers at a data warehouse party!

- **Lack of Flexibility:** What's that sound? Oh, it's the flexibility train leaving the station! Changing specific data becomes a bit trickier when everything is bundled up.
- **Not Suitable for Every Party:** One Big Table might steal the show in some scenarios, but for complex, ever-changing data, it might be a buzzkill.

The Conclusion

In the epic battle between Start Schema and One Big Table, the winner depends on your data warehouse needs! If you're a fan of tidy organization and optimized storage, Start Schema might be your champion. But if you crave simplicity, speed, and straightforwardness, One Big Table could be the hero you need!

Remember, folks, every data warehouse is unique, and what works for one might not be the best fit for another. Choose wisely, and may your data adventures be ever thrilling and fruitful!

Data

Data Modeling

Sql

Analytics

Data Analytics



Written by Seckin Dinc

Follow

