Last updated on Jan 18, 2024

What are the steps for designing a star schema in a data warehouse?

Powered by AI and the LinkedIn community

A star schema is a common way of organizing data in a data warehouse, where a central fact table contains the measures of interest and several dimension tables store the attributes that describe the facts. A star schema can improve the performance and simplicity of analytical queries, but it requires careful design to ensure data quality and consistency. In this article, you will learn the steps for designing a star schema in a data warehouse.

Top experts in this article

Selected by the community from 12 contributions. Learn more

Earn a Community Top Voice badge
Add to collaborative articles to get recognized for your expertise on your profile. Learn more
Start a contribution

ABDUL QADEER
Transforming Complexity into Clarity 🚀 | BI DEVELOPER | POWER BI | SQL | SSIS | Python
View contribution · 2

See what others are saying

1 Identify the business requirements

The first step is to understand the business requirements and goals of the data warehouse. What are the key performance indicators (KPIs) that the business wants to measure and analyze? What are the dimensions and hierarchies that the business uses to slice and dice the data? What are the sources and formats of the data that need to be integrated and transformed? By answering these questions, you can define the scope and purpose of the star schema.

Add your perspective

Ashish Kondare
Started Internship in Data Science at SCode Technologies, Activel...

Identify Business Requirements:

Understand the business needs and reporting requirements to determine what data will be stored in the data warehouse.

Like ···

Jon F.
Founder, Phantom Commerce LLC & BudgetBreeze | 1000+ Mon...

Requirements. Requirements. Requirements. High level it is about how your solution aligns with the data requirements that the organization has for data warehousing efforts. Also, keep in mind for where the data is being sourced from and in what ways does the data need to be transformed and wrangled.

Like ···

Load more contributions

2 Define the fact table

The next step is to define the fact table, which is the core of the star schema. The fact table should contain the numeric measures that are relevant to the business requirements, such as sales, revenue, profit, or customer satisfaction. The fact table should also have foreign keys that link to the dimension tables, which provide the context for the measures. The fact table should have a high level of granularity, meaning that it should store the most detailed and atomic data possible.

Add your perspective

ABDUL QADEER
Transforming Complexity into Clarity 🚀 | BI DEVELOPER | POWER...

When defining the fact table, prioritize indexing(clustered and non-clustered) and partitioning strategies for optimal query performance, particularly with large datasets. Efficient extraction of data is crucial to any environment, and these optimization techniques significantly contribute to ti...see more

Like · 2

Ashish Kondare
Started Internship in Data Science at SCode Technologies, Activel...

Select Dimensions and Facts:

Identify dimensions (descriptive attributes) and facts (measurable metrics) that are essential for reporting and analysis. ...

...see more

Like

Load more contributions

3 Define the dimension tables

The third step is to define the dimension tables, which are the spokes of the star schema. The dimension tables should contain the descriptive attributes that characterize the facts, such as product, customer, location, time, or channel. The dimension tables should have primary keys that match the foreign keys in the fact table, and they should have descriptive names and labels for the attributes. The dimension tables should also have a low level of granularity, meaning that they should store the most aggregated and summarized data possible.

Add your perspective

Ashish Kondare
Started Internship in Data Science at SCode Technologies, Activel...

Select Dimensions and Facts:

Identify dimensions (descriptive attributes) and facts (measurable metrics) that are essential for reporting and analysis. ...

...see more

Like

4 Apply the design principles

The fourth step is to apply the design principles that can optimize the star schema for performance and usability. Surrogate keys should be used instead of natural keys for primary and foreign keys to avoid data inconsistencies and improve query speed. Additionally, null values should be avoided and default values or flags should be used for missing or unknown data to ensure data quality and accuracy. Consistent naming conventions and data types should be used for tables and columns to facilitate data integration and documentation. Furthermore, dimension tables should be normalized and the fact table should be denormalized to reduce data redundancy and improve query efficiency. Lastly, indexes, partitions, and compression techniques should be implemented to enhance data access and storage.

Add your perspective

Abhishek Patil
5 HackerRank || Google Cloud || React Js || Tech Enthusiast || P...

Surrogate keys are a game-changer in star schema design, as they prevent the pitfalls of using natural keys that can change over time, leading to data discrepancies. By using surrogate keys, we ensure a stable and consistent key structure that significantly speeds up query performance, especia...see more

Like

5 Validate and test the star schema

The final step is to validate and test the star schema, to ensure that it meets the business requirements and expectations. You can use various methods to do so, such as loading sample data from the source systems and verifying that the data is correctly transformed and loaded into the star schema. Additionally, you should perform data quality checks and resolve any data issues or errors. Sample queries and reports should be run and compared with the expected outcomes, while feedback from end users and stakeholders should be sought out and incorporated into the star schema.

Add your perspective

Abhishek Patil
5 HackerRank || Google Cloud || React Js || Tech Enthusiast || P...

Validation and testing of a star schema are critical to ensure its performance and accuracy.I've seen firsthand how loading representative datasets and performing rigorous data quality assessments can uncover issues that might not be apparent at the design stage. Moreover, engaging with end-use...see more

Like

6 Here's what else to consider

This is a space to share examples, stories, or insights that don't fit into any of the previous sections. What else would you like to add?

Add your perspective

Abhishek Chaudhary

Creating a successful star schema isn't just about how it looks; it's about how it adapts and grows over time. Knowing the quirks of where your data comes from is crucial; anomalies in the data can provide valuable insights, but they can also trip you up if you're not careful. ...

...see more

Like (edited)

Database Development + Follow

Rate this article
We created this article with the help of AI. What do you think of it?
It's great | It's not so great

Report this article

More articles on Database Development

Your database replication process is lagging. What steps can you take to speed it up?

You need to store data but the best database software for you?

See all

Explore Other Skills

Web Development | Programming | Agile Methodologies | Machine Learning | Software Development | Computer Science | Show more

More relevant reading

OLAP
How do you balance storage efficiency and query speed in star schema and snowflake schema?

Data Warehousing
How do you update and maintain a star schema over time?

Dimensional Modeling
How do you optimize query performance and data loading in a partitioned data warehouse?

Data Warehouse Architecture
How do you incorporate hierarchies and aggregates in a star schema?