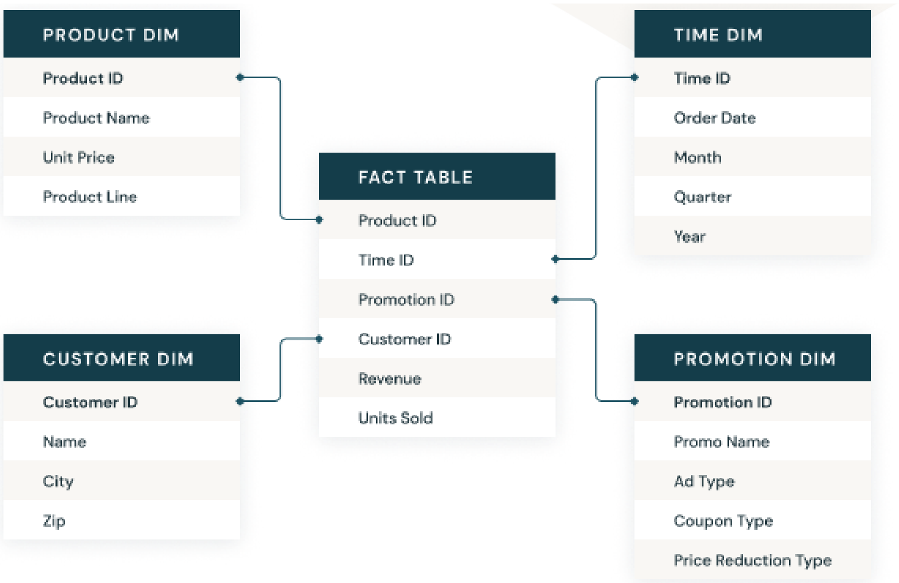
Star Schema

All > Star Schema

What is a star schema?

A **star schema** is a multi-dimensional data model used to organize data in a database so that it is easy to understand and analyze. Star schemas can be applied to data warehouses, databases, data marts, and other tools. The star schema design is optimized for querying large data sets.

Introduced by Ralph Kimball in the 1990s, star schemas are efficient at storing data, maintaining history, and updating data by reducing the duplication of repetitive business definitions, making it fast to aggregate and filter data in the data warehouse.



Fact tables and dimension tables

A star schema is used to denormalize business data into dimensions (like time and product) and facts (like transactions in amounts and quantities).

A star schema has a single fact table in the center, containing business "facts" (like transaction amounts and quantities). The fact table connects to multiple other dimension tables along "dimensions" like time, or product. Star schemas enable users to slice and dice the data however they see fit, typically by joining two or more fact tables and dimension tables together.

Denormalized data

Star schemas **denormalize** the data, which means adding redundant columns to some dimension tables to make querying and working with the data faster and easier. The purpose is to trade some redundancy (duplication of data) in the data model for increased query speed, by avoiding computationally expensive join operations.

In this model, the fact table is normalized but the dimensions tables are not. That is, data from the fact table exists only on the fact table, but dimensional tables may hold redundant data.

- Fact/dimensional models like star schemas are **simple** to understand and implement, and make it easy for end users to find the data they need. They can be applied to data marts and other data resources.
- **Great for simple queries** because of their reduced dependency on joins when accessing the data, as compared to normalized models like snowflake schemas.
- **Adapt well to fit OLAP models.**
- **Improved query performance** as compared to normalized data, because star schemas attempt to avoid computationally expensive joins.

How does a star schema differ from 3NF (Third Normal Form)?

3NF, or Third Normal Form, is a method of reducing data-redundancy through normalization. It is a common standard for databases that are considered fully normalized. It typically has more tables than a star schema due to data normalization. On the flip-side, queries tend to be more complex due to the increased number of joins between large tables.

Resources

- Databricks Blog: Five Simple Steps for Implementing a Star Schema in Databricks with Delta Lake
- Databricks SQL product page
- Databricks glossary - Snowflake schema

Back to Glossary

Learn what’s driving the Lakehouse pattern.

REGISTER NOW

Databricks Inc.
160 Spear Street, 13th Floor
San Francisco, CA 94105
1-866-330-0121