

HOW TO

SEPTEMBER 30, 2022

Star Schema vs Snowflake Schema and the 7 Critical Differences

THE 7 MAJOR STAR SCHEMA VS SNOWFLAKE SCHEMA DATABASE DIFFERENCES AND HOW TO CHOOSE THE RIGHT FOR YOUR USE CASE.

Star schemas and snowflake schemas are the two predominant types of data warehouse schemas.

A data warehouse schema refers to the shape your data takes - how you structure your tables and their mutual relationships within a database or data warehouse.

Since the primary purpose of a data warehouse (and other Online Analytical Processing (OLAP) databases) is to provide a centralized view of all the enterprise data for analytics, data warehouse schemas help us achieve superior analytic results.

How do schemas help analytics? And what are the differences and trade-offs between star and snowflake schemas?

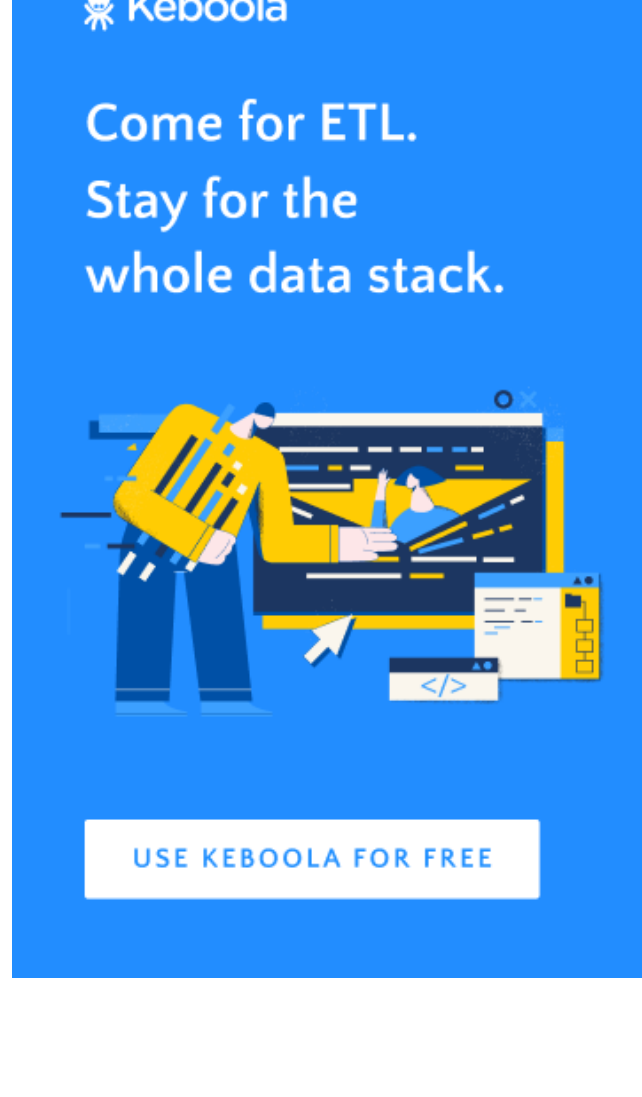
In this article, we compare the two dominant variants of data warehouse schemas and contrast their advantages and disadvantages.

Table of contents:

- What is star schema
- What is snowflake schema
- The 7 critical differences between a star schema and a snowflake schema
- Which one should I pick?

Build data products in days instead of months and focus on what really matters - delivering value to your customers.

Contact us



What is a star schema?

To understand the data modeling behind a star schema, let us look at a retail example. Imagine you are running an international shopping brand and you want to analyze purchases across your physical locations.

You pull out data from your database as an Excel file:

order_id	customer name	customer email	address delivery	purchase date	store	product	# items	total
1	Anthony Hopkins	ah@gmail.com	First road 12, OX12 344, UK	2021-02-27	Abby's	Lamb	14	\$79.10
2	Anthony Hopkins	ah@gmail.com	First road 12, OX12 344, UK	2021-02-28	Abby's	Knife	3	\$72.00
3	Jodie Foster	jf@gmail.com	711-2880 Nulla St. Markato Mississippi 96522	2021-08-03	McFarlen Store	Tape	2	\$6.00
...

But you soon realize there are too many rows, and the data needs to be [cleaned](#) before you can analyze it.

You decide to turn the data into a star schema.

A star schema is a data model that stores information in multiple table types: a single fact table and multiple dimensional tables.

In contrast to the classical database design of normalizing tables, star schemas connect dimensional data with fact data in a shape resembling a star (hence the name), as can be seen from the following diagram:

In the diagram, we see a central fact table (holding all the facts of the sales) and four dimension tables - a separate table describing the customer, date (of purchase), store where the purchase happened, and product purchased.

The fact table is linked via a foreign key relationship to the primary key of each dimension (aka, the id in each dimension table, for example, the customer_id links the customer from the dim_customer table to the fact_sales table).

This type of data modeling allows us to query data faster and with simpler queries than the normalized database design.

What is a snowflake schema?

A snowflake schema is very similar to the simple star schema above. The main difference is that snowflake schemas split dimensional tables into further dimensional tables (also called lookup tables).

For example, the above diagram would show the customer_country field being split into further dimensional tables:

Each dimension is split until it is normalized - aka, there is no redundancy in the dimensional table, no repetition of values (except for identifier values, such as id's).

The 7 critical differences between a star schema and a snowflake schema

1. Normalization of dimension tables

The snowflake schema is a fully normalized data structure. Dimensional hierarchies (such as city > country > region) are stored in separate dimensional tables.

On the other hand, star schema dimensions are denormalized. Denormalization refers to the repeating of the same values within a table.

2. Data redundancy

Star schema stores redundant data in dimension tables, while snowflake schema fully normalizes dimension tables and avoids data redundancy.

For example, a star schema would repeat the values in field *customer_address_country* for each order from the same country.

The redundancy, or duplicated entries, occurs because of the denormalization vs normalization schema design.

3. Query complexity

A simple star schema leads to simple query writing. Because the fact table is joined to only one level of dimensional tables, analysts do not need to write multiple joins.

On the other hand, snowflake schemas require a more complex query design. Because of complex relationships between the fact table and its dimensional tables, more joins are needed to link the additional tables. This causes an additional overhead when writing analytical queries.

4. Query performance

The query execution time is faster in star schemas. Because they require a single join between a fact and its set of attributes in dimensional tables, a star schema acts almost as a single table for query lookups.

In contrast, snowflake schemas require complex joins of dimensional tables with their own sub-dimensional or supra-dimensional tables. This slows down query processing and can affect other OLAP products such as cube processing.

5. Disk space

Star schemas might run queries faster, but they require more storage space than snowflake schemas because of their data redundancy.

6. Data integrity

Data integrity is more at risk in star schemas than snowflake schemas. Because data is stored redundantly, multiple copies of the same data exist in the star schema's dimensional tables. This means new inserts, updates, or deletes can compromise the integrity of data.

In contrast, the snowflake schema is less prone to data integrity issues, because it fully normalizes dimensional tables, storing dimension data only once in the appropriate table.

7. Set up and maintenance

Star schemas are easier to design and set up. Because they are represented by simple relationships, it is easy for a data engineer or data architect to set up an appropriate star schema.

On the other hand, star schemas are harder to maintain than snowflake schemas. As new [data is ingested](#) into the data warehouse, star schemas become harder to maintain and check against data integrity violations.

Build data products in days instead of months and focus on what really matters - delivering value to your customers.

Contact us

Which one should I pick?

Out of the two types of data warehouse schema, which one should you choose?

On one hand, star schemas are simpler, run queries faster, and are easier to set up.

On the other hand, snowflake schemas are less prone to data integrity issues, are easier to maintain, and utilize less space.

Based on the tradeoffs above, it depends on which advantage (or disadvantage) best suits your business use cases.

One thing we know for sure is that with Keboola on top of either data warehouse schema, you will be able to model your data faster.

Design your dream data architecture with Keboola

Keboola is a data platform as a service that helps you automate your data pipelines.

Designing the right data warehouse schema is hard enough.

Lessen the burden on your engineers by automating all the data extraction, transformation, and loading into databases and data warehouses of your choice that follow schema design.

Try Keboola for free.

Keboola offers a [no-questions-asked, always-free tier](#), so you can play around and build your pipelines with a couple of clicks.

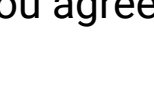
Did you enjoy this content?

Have our newsletter delivered to your inbox.

Business email



I'm not a robot



By subscribing to our newsletter you agree with Keboola Czech s.r.o. [Privacy Policy](#).

Subscribe to Keboola Newsletter



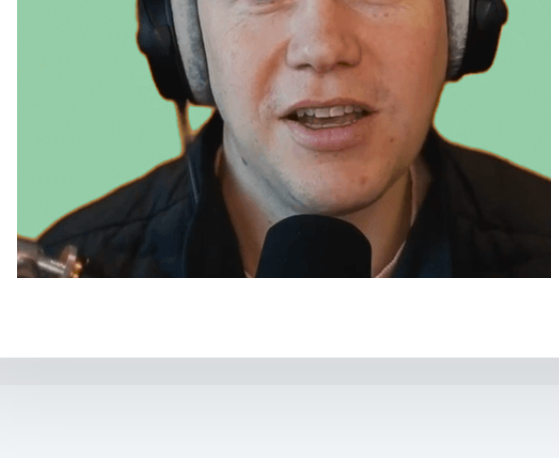
Recommended Articles



NEWSLETTERS

February Newsletter: Discover The Latest From Keboola

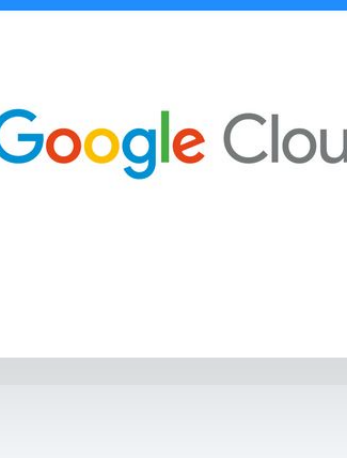
READ MORE →



RESOURCES

Video: Manage All Your Data in One Place with Keboola

READ MORE →



Google Cloud

Ready to get started? Request a free demo, or get in touch today.

Request a Demo

Try Keboola

PRODUCT
Overview
Support

RESOURCES
Customer Stories
Careers

LEGAL
Contact
Terms & Conditions

KEBOOLA NEWSLETTER
Subscribe

We use cookies.

We use cookies to make Keboola's website a better place. Cookies help to provide a more personalized experience and relevant advertising for you, and web analytics for us. By clicking "Accept", you agree to the storing of cookies on your device to enhance site navigation, analyze site usage and assist in our marketing efforts. To learn more about the different cookies we're using, check out our [Cookie Policy](#)

Accept

Cookie Settings