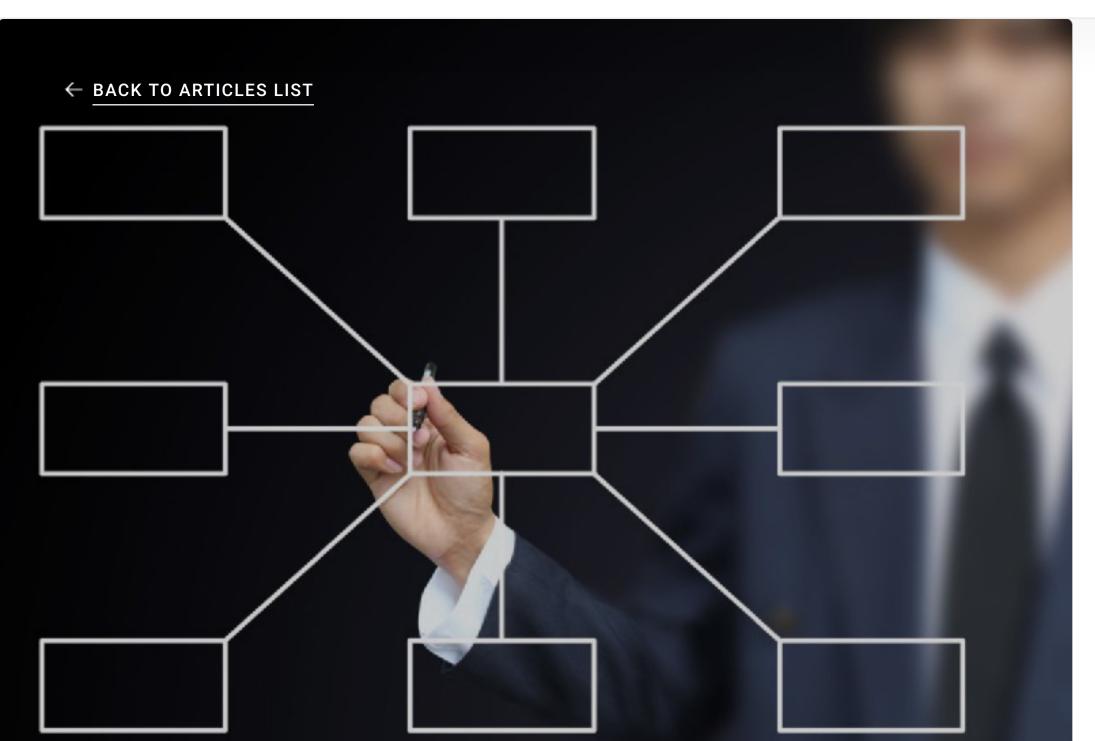
May 23, 2023

- 9 minutes read



TECHNICAL What Is the Star Schema Data Model? An Explanation with 3 Examples

Radu works as a full-time data engineer. He loves SQL and anything data-related. In

his spare time, he likes to put his data to work by building data pipelines. He even trains machine learning models to predict the best time of the day to drink his coffee. He also enjoys helping others and sharing his knowledge on our blog and Stack Overflow. Tags: data modeling star schema fact table

Radu Gheorghiu

to structure your data? What are its advantages over other data schemas? We'll answer these questions and more in this article. What Is a Star Schema?

The star schema is a popular approach for designing **DATA WAREHOUSING** and business intelligence data models. It uses dimensional modeling; in other words, it utilizes

In this article, we will talk about the **STAR SCHEMA DATA MODEL**. When should you use this

dimension tables with descriptive attributes that provide context for the measures in the fact table. What Is Dimensional Modeling?

In dimensional modeling, a data warehouse is designed to store data in a way that

improves its retrieval times. This is important for many data analytics applications.

There are two types of tables in dimensional modeling. The fact table contains the quantitative data or measures of the business, while the dimension tables provide

context or descriptive attributes by which the measures are analyzed. For example, a sales data model could have a sales fact table with dimensions like time, product, customer, and geography.

What Is a Fact Table? In the star schema data model, the fact table is the most important element of the entire schema. It is at the center of the model and generally contains only numerical data, or data that can be measured.

measures to be analyzed by different dimensions. Each row in the fact table represents

WAREHOUSING.

a unique combination of the dimension attributes and their corresponding measure. For example, a row in a fact table containing sales data might have a column representing the quantity of a specific product sold on a particular date to a specific customer in a specific region.

The fact table is connected to each dimension by the foreign keys, allowing the

The purpose of a fact table is to enable fast and efficient querying of the data, making it possible to analyze large datasets quickly and efficiently. The fact table is usually very large and can contain many millions or billions of rows. This is why it is essential to have a well-designed data model for any application.

In short, the fact table in the star schema data model is where the numerical data or measures of the business are stored. It has foreign keys for each dimension and enables fast and efficient querying of the data. If you want to go in depth about the intricacies of fact tables, be sure to check out our article on FACT TABLES IN DATA

What Is a Dimension Table? A dimension table is the other important component of the star schema data model. It contains descriptive attributes that add context to the quantitative data in the fact table. Each dimension represents a specific attribute or viewpoint from which the data can be analyzed. Examples of dimensions for a sales data model can include time, product, customer, and geography.

Dimension tables are generally much smaller than fact tables, making them easier to

to be analyzed by different dimensions. For example, in a sales data model, the fact

Similarly to the fact table, in the star schema data model, each dimension table is connected to the fact table through a foreign key. This relationship allows the fact table

query and analyze.

table could be analyzed by the product, customer, and time dimensions. This can be used to gain insights into sales trends by product, customer demographics, and sales performance over time. In conclusion, a dimension table is the place where descriptive attributes are stored, attributes which add context to the quantitative data in the fact tables. If you want to go more in-depth, we have a great article on the **TYPES OF DIMENSIONS IN A DATA WAREHOUSE**.

understand for business users with little technical knowledge. For example, in a sales data model, some potential analytical queries that can be performed using the star schema data model include:

• Sales by product: Analyzing the sales of different products over a time period, such

• Sales by region: Analyzing the sales made in different regions over a specific time

The star schema data model's design optimizes query performance for analytical

queries rather than transactional queries. This makes it easy to navigate and

• Sales by customer: Analyzing the sales made to different customers over a specific time period, such as monthly or quarterly sales by customer.

as monthly or quarterly sales.

How Do You Build a Star Schema?

analyzed from different perspectives.

be tracked.

schema.

Sales Data Model

this scenario.

Product

Integer

Varchar(100) M

Varchar(100) M

Varchar(100) M

Account

AccountDescript Varchar(100) M

scenario.

Integer

Varchar(100) M

Varchar(100) M

M PI

AccountID

AccountName

AccountType

M PI

ProductID

Category Subcategory

ProductName

Why Use a Star Schema in Data Modeling?

period, such as monthly or quarterly sales by region. By analyzing data in this way, businesses can make informed decisions and optimize their strategies for growth and success.

To build a star schema, there are generally 5 steps you need to go through. The

following is a brief summary of each step: 1. Identify business processes: The first step in building a star schema is to identify the key business processes that need to be modeled. This involves understanding

the organization's data requirements and identifying the key metrics that need to

step is to select the dimensions that will be used in the model. Dimensions are the

descriptive attributes that provide context for the measures in the fact table. For example, in a sales data model, dimensions might include time, product, customer, and geography. 3. Create the fact table: The fact table is the centerpiece of the star schema and

contains the quantitative data or measures of the business. The fact table is

connected to the dimension tables through foreign keys, allowing the data to be

2. Select dimensions: Once the business processes have been identified, the next

4. Define the relationships: Once the fact table and dimension tables have been created, the next step is to define the relationships between them. This involves creating foreign keys in the fact table that reference the primary keys in the dimension tables. 5. Populate the tables: Once the schema has been designed and the relationships

have been defined, the final step is to populate the tables with data. This involves

importing data from various sources and transforming it to conform to the

Examples of Star Schema Data Models We're now going to go over three example use cases where the star schema data model could be deployed.

as revenue, quantity sold, and profit. The fact table would contain the sales data, while the dimension tables would provide context, such as product, customer, time, and geography dimensions. Below you can see an example of what a star schema data model would look like for

A sales data model could be designed using a star schema to track sales metrics, such

Sales SaleID M PI M ProductID Integer

Decimal(10, 2) M

Decimal(10, 2) M

TimeID

DayOfWeek

Date

Month

Year

Quarter

CustomerID

GeographyID QuantitySold

Revenue Profit

Customer

Integer

Varchar(50) M

Varchar(50) M

Varchar(50) M

Varchar(50) M

Varchar(50) M

CustomerID

FirstName

LastName

State

Country

M

M M

Fact Table

Integer M PI

Date M

Integer M

Integer M

Integer M

Integer M

Fact Table

Time

Integer M PI

Date M

Integer M

Integer M

Integer M

Integer M

TimeID

DayOfWeek

Date

Month

Quarter Year

Geography

Integer

Varchar(50) M

Varchar(50) M

Varchar(50) M

Dimension Tables

M PI

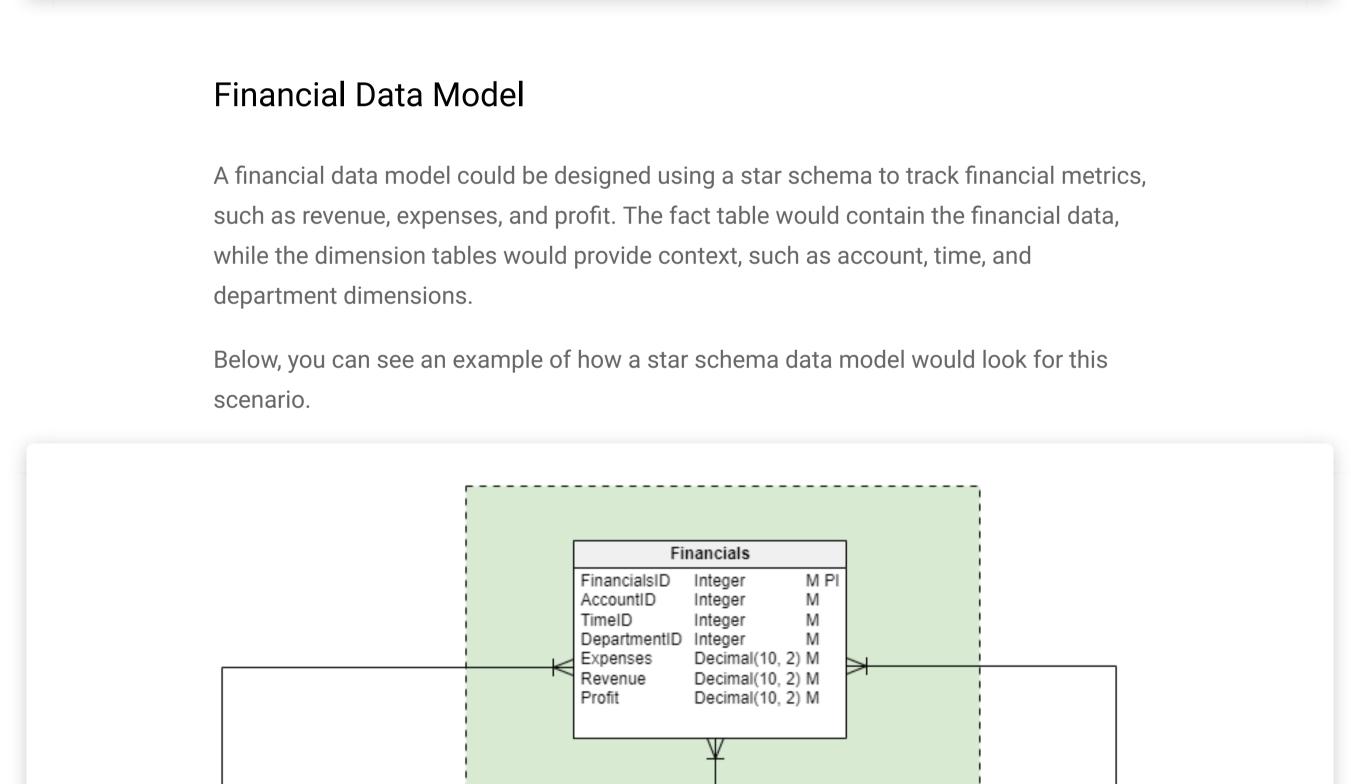
GeographyID

City

State

Country

Time



Dimension Tables

DepartmentID

Country

Department

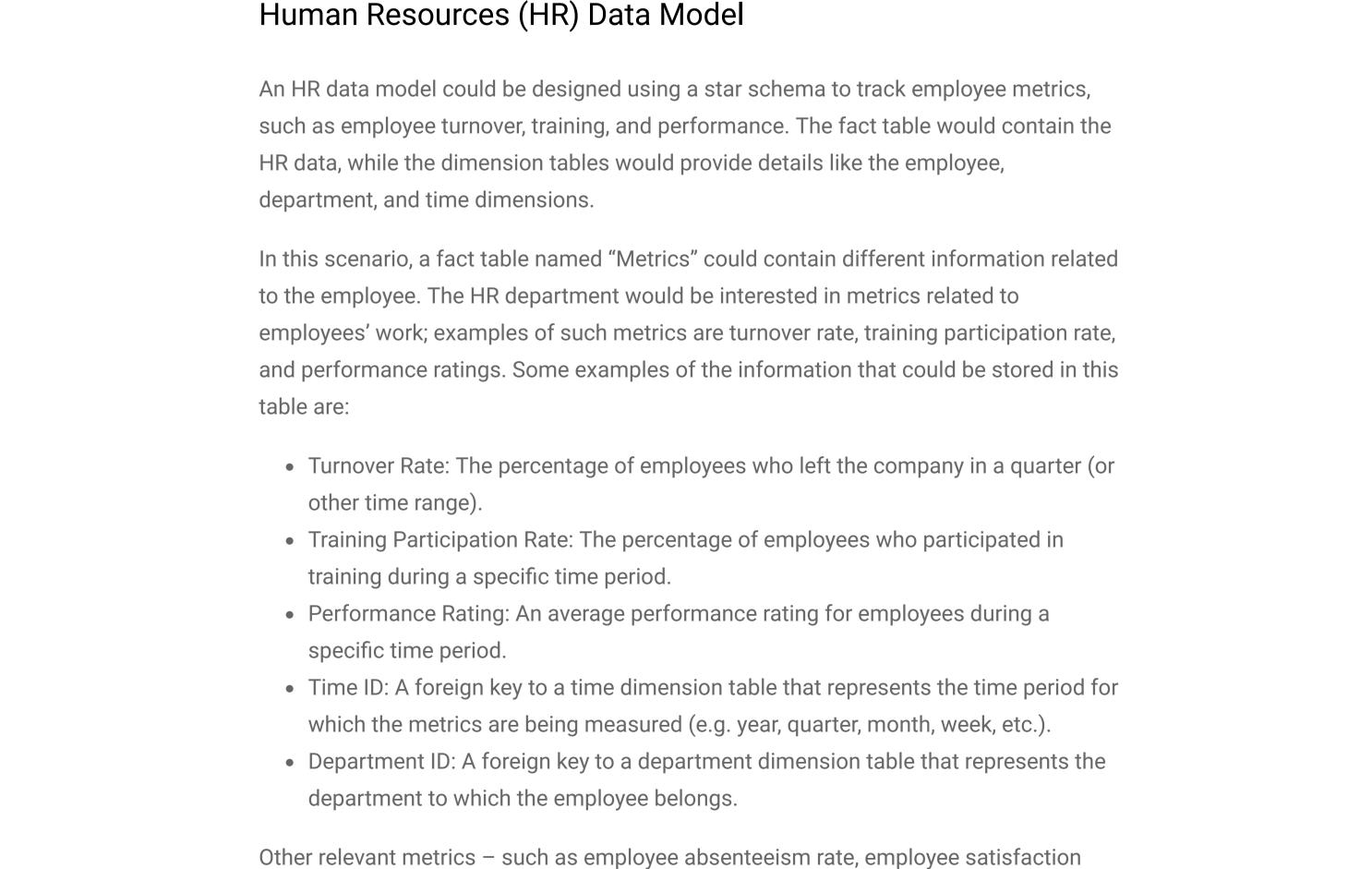
DepartmentName Varchar(50) M

DepartmentDescr Varchar(50) M

Integer

Varchar(50) M

Varchar(50) M



rate, employee engagement score, etc. - could also be included.

Below, you can see an example of how a star schema data model would look for this

DepartmentID Integer

MetricID EmployeeID Turnover

Metrics

Decimal(10, 2) M Decimal(10, 2) M

M

Fact Table

Month

Quarter

Integer M

Integer M

Integer M

Dimension Tables

Department Employee DepartmentID Integer EmployeeID Integer M P FirstName DepartmentNam Varchar(100) M DayOfWeek LastName Integer M

Varchar(50) M

HireDate

RESOURCE ON DATABASE SCHEMAS.

Advantages of a Star Schema

downtime of the overall system.

make better decisions based on data.

Tags: data modeling star schema fact table

SUBSCRIBE

drive growth and success.

TerminationDate

In summary, here are some advantages of using a star schema in data modeling: • Improved performance: Queries will run faster on a star schema data model, which will enable analysts and business users to get answers to their questions quickly and efficiently.

trends and patterns that might otherwise be missed.

be analyzed quickly and combined in dynamic ways.

When a Star Schema Is the Right Choice

• Simplified analysis: Because a star schema is intuitive and clear, the organization

• Scalability: The star schema is able to handle large amounts of data, making it a

great choice for businesses that deal with massive amounts of data that need to

• Easy maintenance: Because the star schema has a simple design, it makes it easy

to maintain and update the model, all while reducing the risk of errors and

of data will make analysis easy for business users, which means they can identify

If you want more examples of database schemas, feel free to check out our free

Now that you've made it to the end of the article, I hope that you have a better understanding of where a star schema can be used and what advantages it provides for data modeling and analysis. By organizing data into dimensions and measures, the star schema provides an easy way to analyze large datasets, allowing business users to

Additionally, the star schema's scalable design makes it an ideal choice for situations

maintenance ensures that it can be updated and modified as needed.

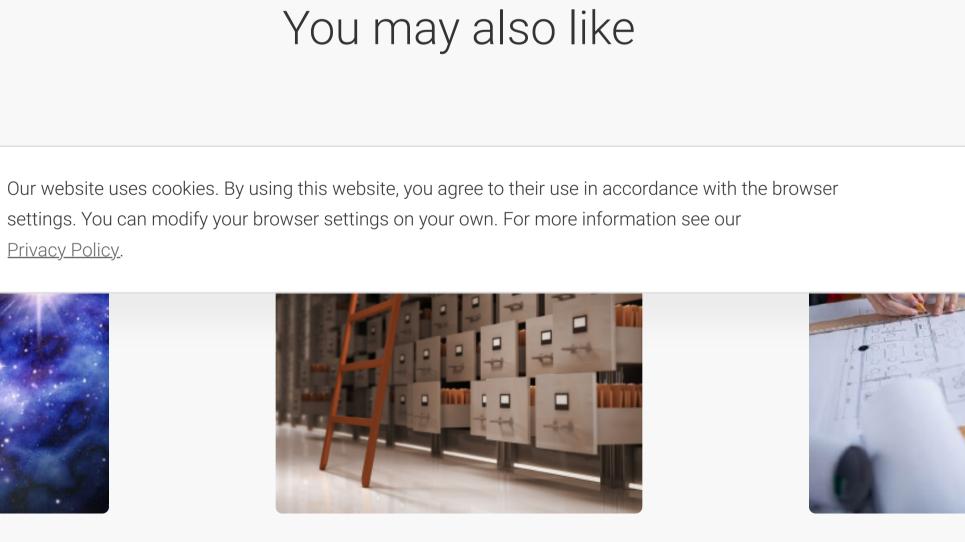
where businesses need to store and analyze massive amounts of data. Plus, its ease of

Overall, the star schema is a powerful tool for any businesses that are looking to gain a

deeper understanding of their operations and make decisions based on data. This can

- There are however scenarios where a star schema is not sufficient and you might need an extended version of this data model. In that case, be sure to check out the snowflake schema. You can learn more in our article that compares a **STAR SCHEMA AND A SNOWFLAKE SCHEMA IN DATA WAREHOUSING.**
- Join our weekly newsletter to be notified about the latest posts. Email address

Subscribe to our newsletter



READ MORE Follow Us

OK

READ MORE

FAQ

DESIGN PATTERNS

Quick links Vertabelo Blog Features

Support If you have any questions or you need our help, you can contact us through our Pricing Academy SUPPORT SITE

Most Common Dimensional Table Types **READ MORE** •

Documentation

TECHNICAL DESIGN PATTERNS The Star Schema What Is a Data Dimensions of Warehouse? Dimensions: A Look at Data Warehousing's