

This member-only story is on us. Upgrade to access all of Medium.

Member-only story

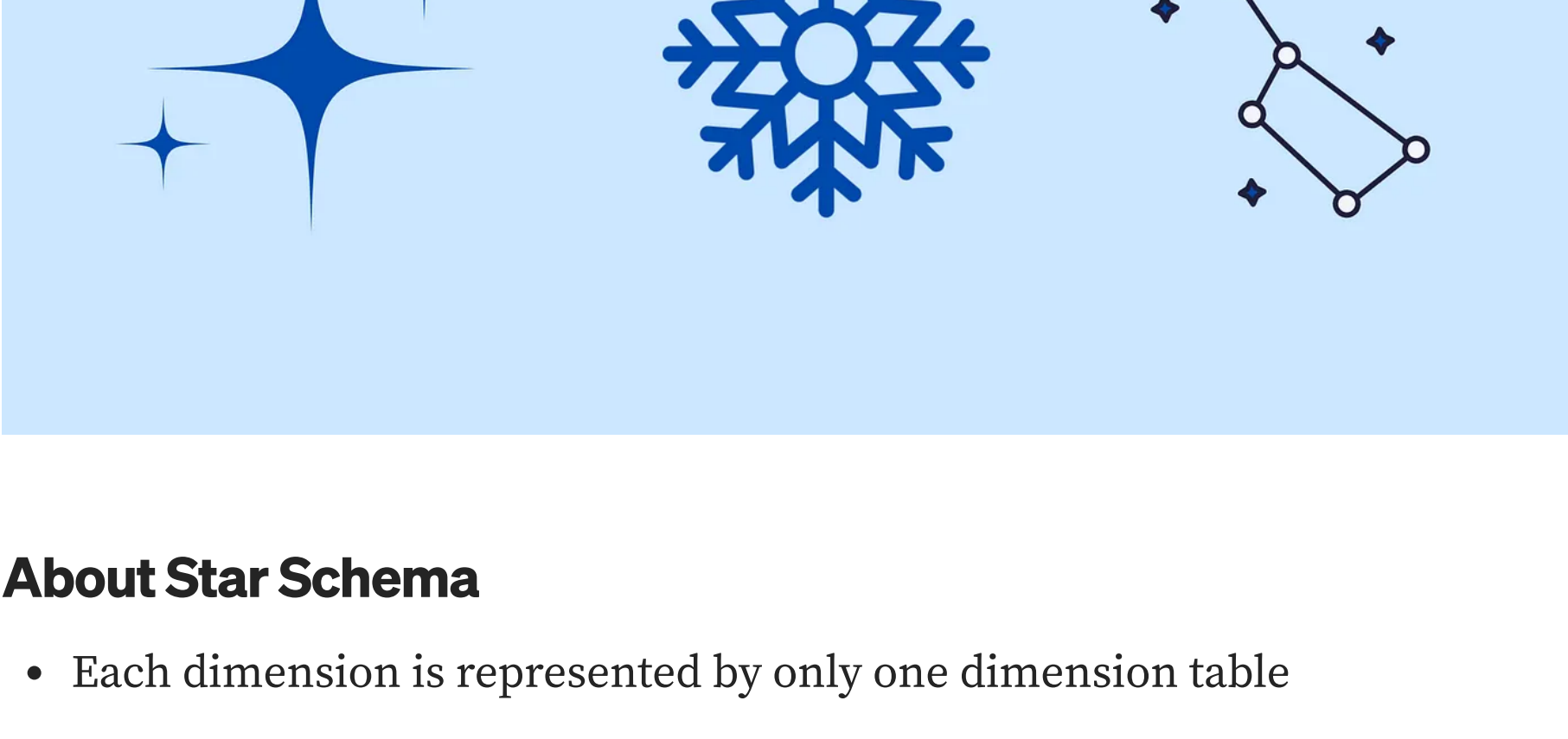
Data Warehousing Schemas

Example: Star Schema, Snowflake Schema, Fact Constellation Schema

Aman Ranjan Verma · Follow
Published in Analytics Vidhya · 3 min read · May 19, 2022

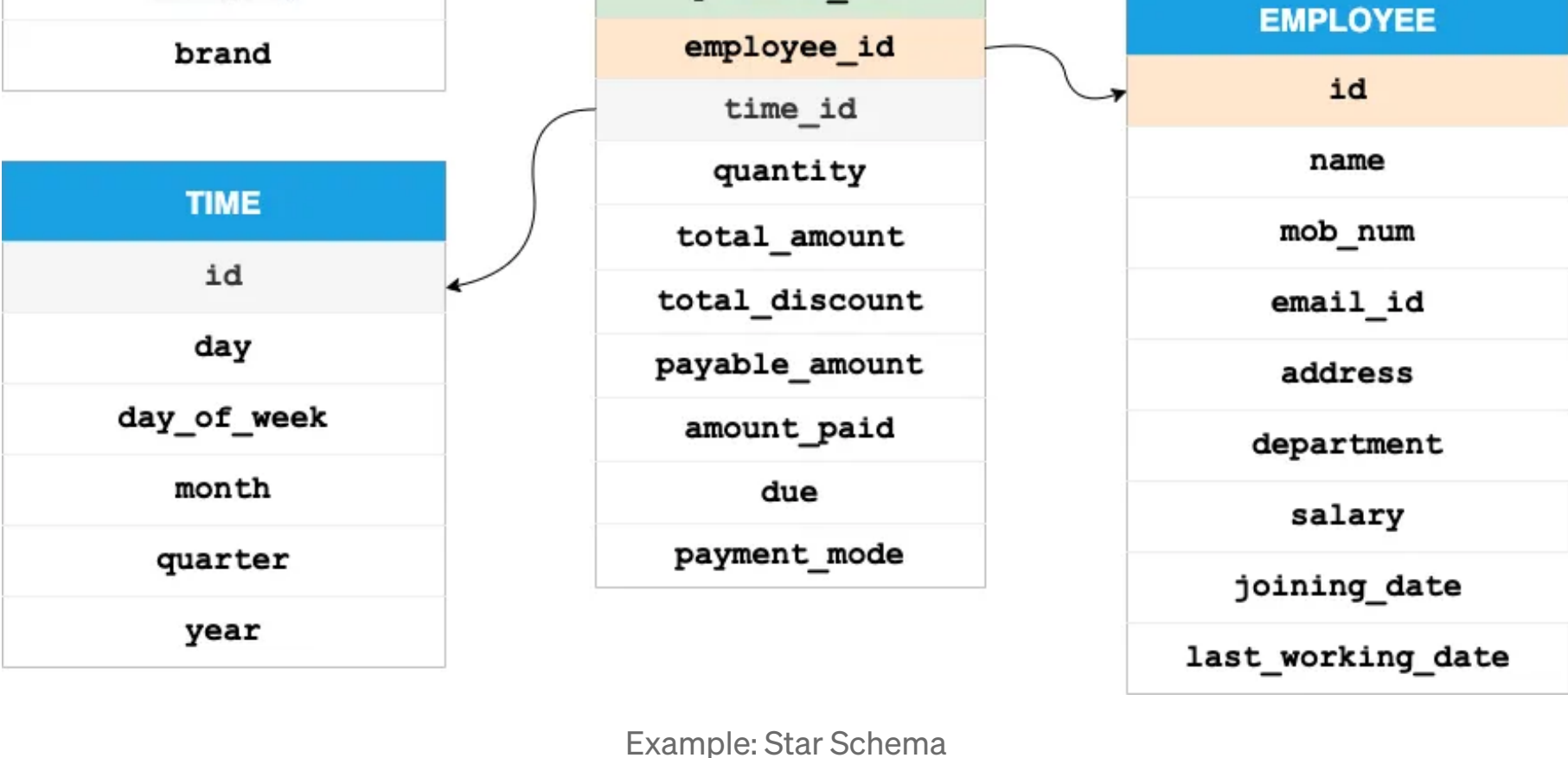
128 1 1 1 1 1

Much like an OLTP system(database), an OLAP system(data warehouse) is also required to maintain a schema. A database uses an entity relational model, while a data warehouse uses a dimensional model. In this blog, we will see the three major schemas that are used to model a data warehouse.



About Star Schema

- Each dimension is represented by only one dimension table
- The dimension tables are directly linked with the fact table
- Queries are not that complex because of straightforward joins.
- Simple queries result in the faster query response time
- Most widely used in the industry

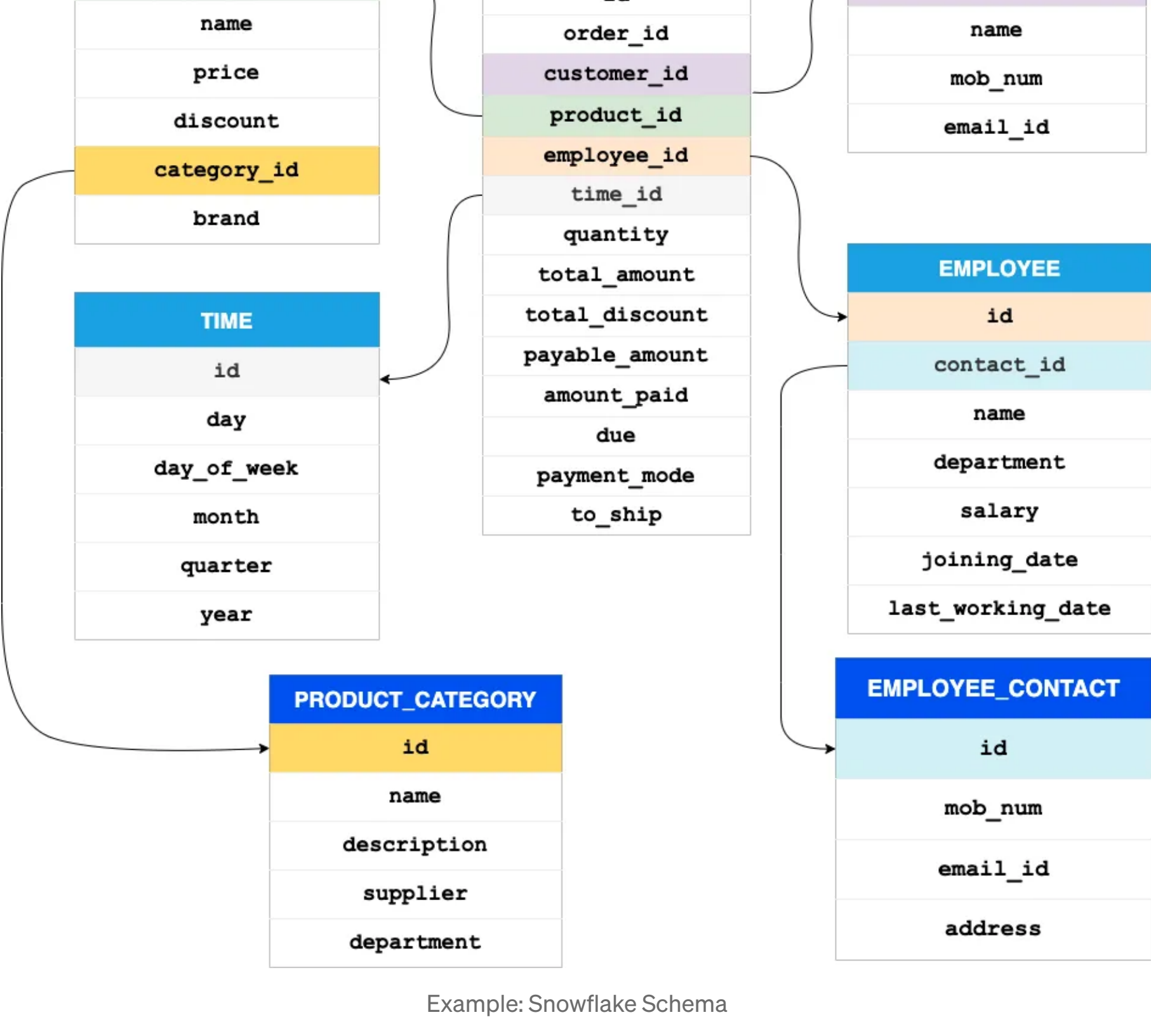


Example: Star Schema

As in this example, we have one fact(Sales) and four dimensions (Product, Time, Customer, Employee). The fact has measures like quantity, amount, discount, and payment.

About Snowflake Schema

- The dimension table is split into sub-dimension tables
- Some dimensions are normalized to avoid redundancy of data
- Used when a dimension table starts growing at a higher rate

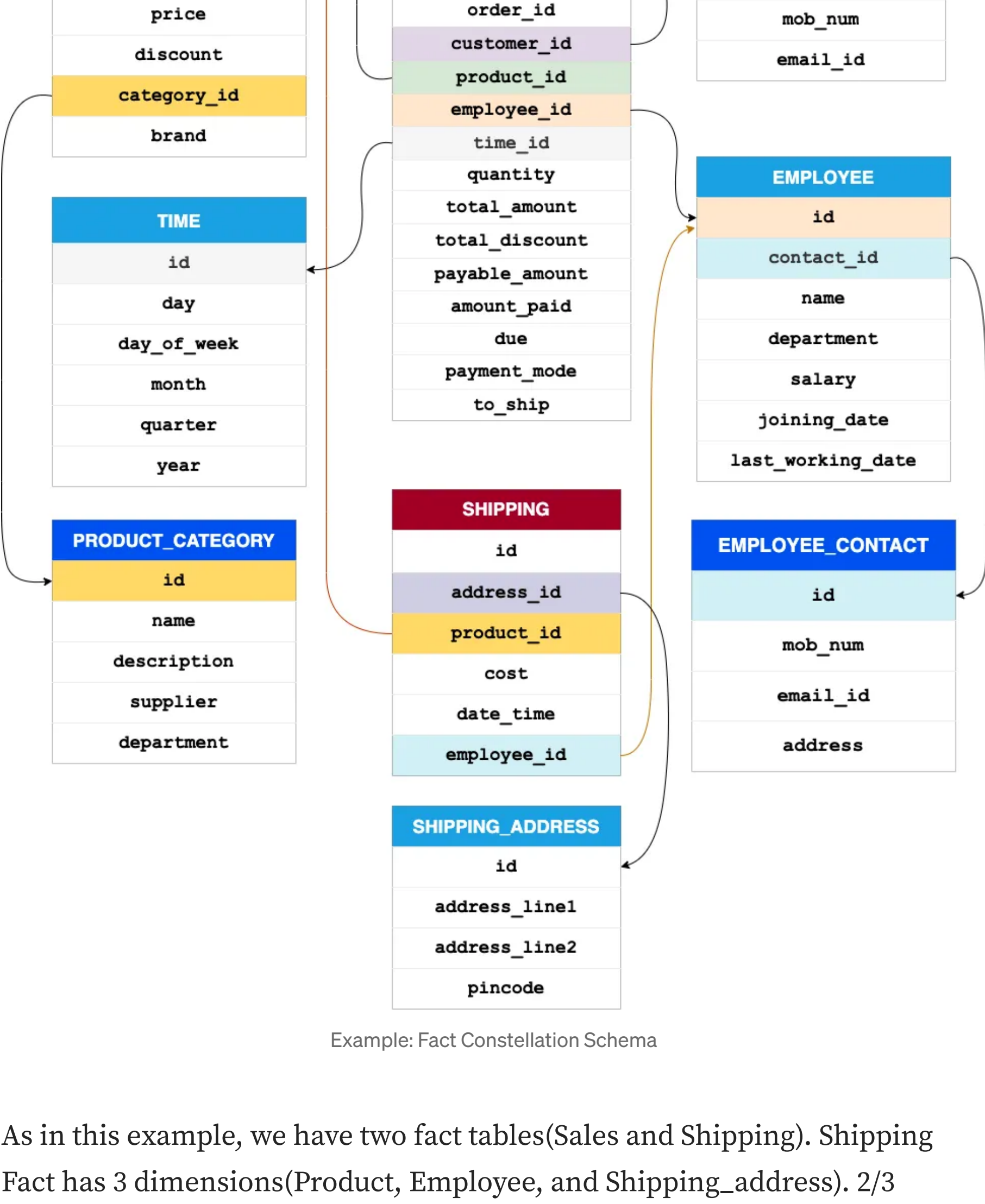


Example: Snowflake Schema

As in this example, the product and employee dimensions have been further divided to hold more data. Businesses are not interested in the employee contact details, so we can segregate them to a different table. Similarly, the product category dimension might be of interest to the inventory management team, so we can keep this information aside. They are no fixed rules, it depends on what is the use case.

About Galaxy Schema

- It is also called Fact Constellation Schema
- It has more than one fact
- It can be an extension of either star schema or snowflake schema
- The same dimension table can be shared between more than one fact table



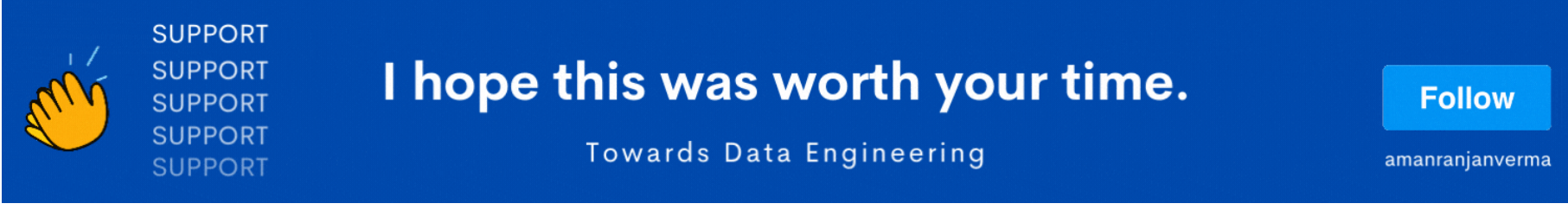
Example: Fact Constellation Schema

As in this example, we have two fact tables(Sales and Shipping). Shipping Fact has 3 dimensions(Product, Employee, and Shipping_address). 2/3 dimensions are shared between the two facts.

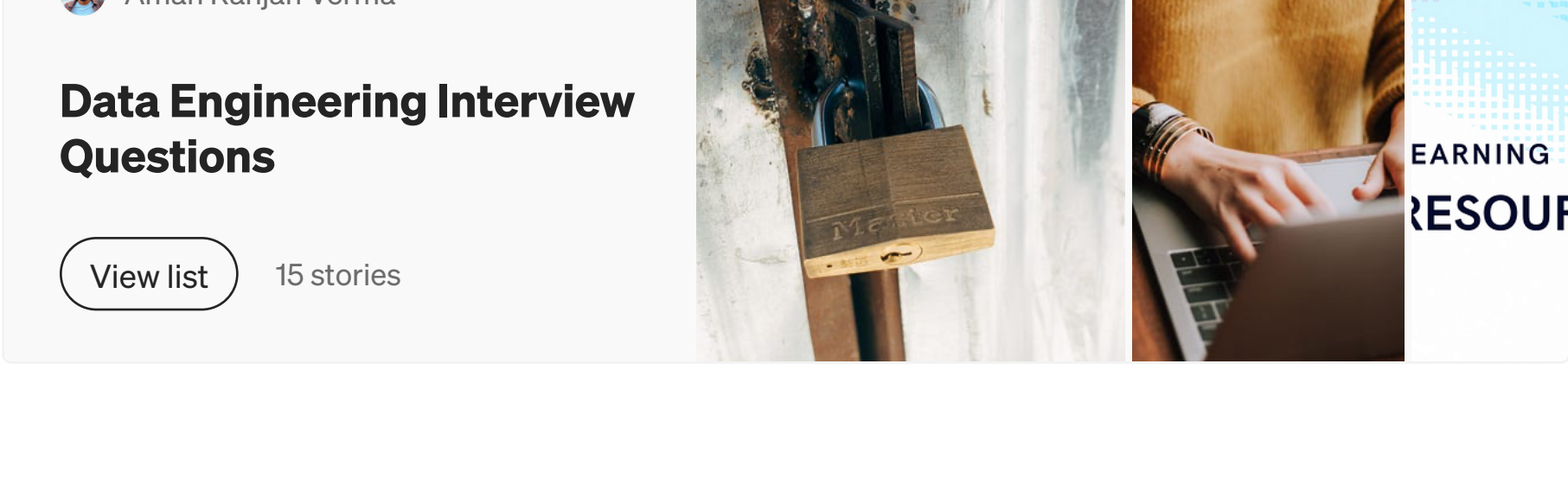
Conclusion

We saw three different data warehouse modeling techniques, the star schema, the snowflake schema, and the fact constellation schema. Each one has its pros and cons. Your use case will let you know how much the cons of one schema impact your need. You will have to keep future requirements in mind before choosing. People start with Star and then they slowly move to the snowflake schema and then to the fact constellation schema.

...



If you are looking to prepare for a Data Engineering interview do check out my interview blog series:



Data Engineering Data Warehouse Database Sql Data Science

128 1 1 1 1 1

Aman Ranjan Verma

Written by Aman Ranjan Verma

955 Followers · Writer for Analytics Vidhya

Senior Data engineer, QuillBot | Ex-Flipkart | Ex-Sigmoid. I publish weekly. Available for 1:1 at [toomate.io/amanverma](#)

Follow

1

2

Slack Webhook URL

slack

Aman Ranjan Verma

How to create a webhook URL for a Slack Channel?

A beginner-friendly, Step-by-Step Guide to create and test Slack Webhook URL for...

3 min read · Jun 29

6 1 1 1 1 1

Bookshelves

Kia Elsinga In Analytics Vidhya

How to create a Python library

Ever wanted to create a Python library, albeit for your team at work or for some open...

7 min read · Jan 26, 2020

2.1K 1 1 1 1 1

Confusion Matrix, Accuracy, Precision, Recall, F1 Score

Binary Classification Metric

Harikrishnan N B In Analytics Vidhya

Confusion Matrix, Accuracy, Precision, Recall, F1 Score

Binary Classification Metric

6 min read · Dec 10, 2019

739 1 1 1 1 1

Coding problems asked in my Data Engineering Interview

List of coding questions that I was asked to solve in my data engineering interviews of...

Aman Ranjan Verma In Towards Data Engineering

Coding problems asked in my Data Engineering Interview

List of coding questions that I was asked to solve in my data engineering interviews of...

3 min read · Dec 6, 2022

13 1 1 1 1 1

See all from Aman Ranjan Verma

See all from Analytics Vidhya

Recommended from Medium

Nuhad Shaabani

Practical Introduction to Data Vault Modeling

The Data Vault modeling is used to model the enterprise Data Warehouse Core layer. The...

10 min read · Jul 15

77 1 1 1 1 1

Chisom Nnamani In Towards Data Engineering

Understanding Data Modeling Concepts as a Data Engineer

Data modeling is the process of creating a data model; a blueprint or plan for organizati...

3 min read · Dec 12

70 1 1 1 1 1

Predictive Modeling w/ Python

20 stories · 701 saves

data science and AI

38 stories · 11 saves

Practical Guides to Machine Learning

10 stories · 803 saves

Coding & Development

11 stories · 332 saves

Sean Coyne

SQL Interview Patterns You Must Know for Data Engineers

SQL interview questions often diverge from typical daily tasks, focusing on deeper...

6 min read · Sep 30

143 1 1 1 1 1

Rahul S

Data Engineering: Data Warehouses

In my previous Data Modeling—Basics blog, we discussed about how an organization...

5 min read · Dec 4

5 1 1 1 1 1

Nishchay Agrawal

Data Engineer Walmart Interview Experience

Subscribe My Youtube Channel For Data Engineering Insights & Interview Experience...

9 min read · Sep 20

252 1 1 1 1 1

SQL Fundamentals in DevOps.dev

Top 10 Advanced SQL Queries

SQL (Structured Query Language) is a versatile tool for managing and querying dat...

3 min read · Nov 17

477 1 1 1 1 1

See more recommendations