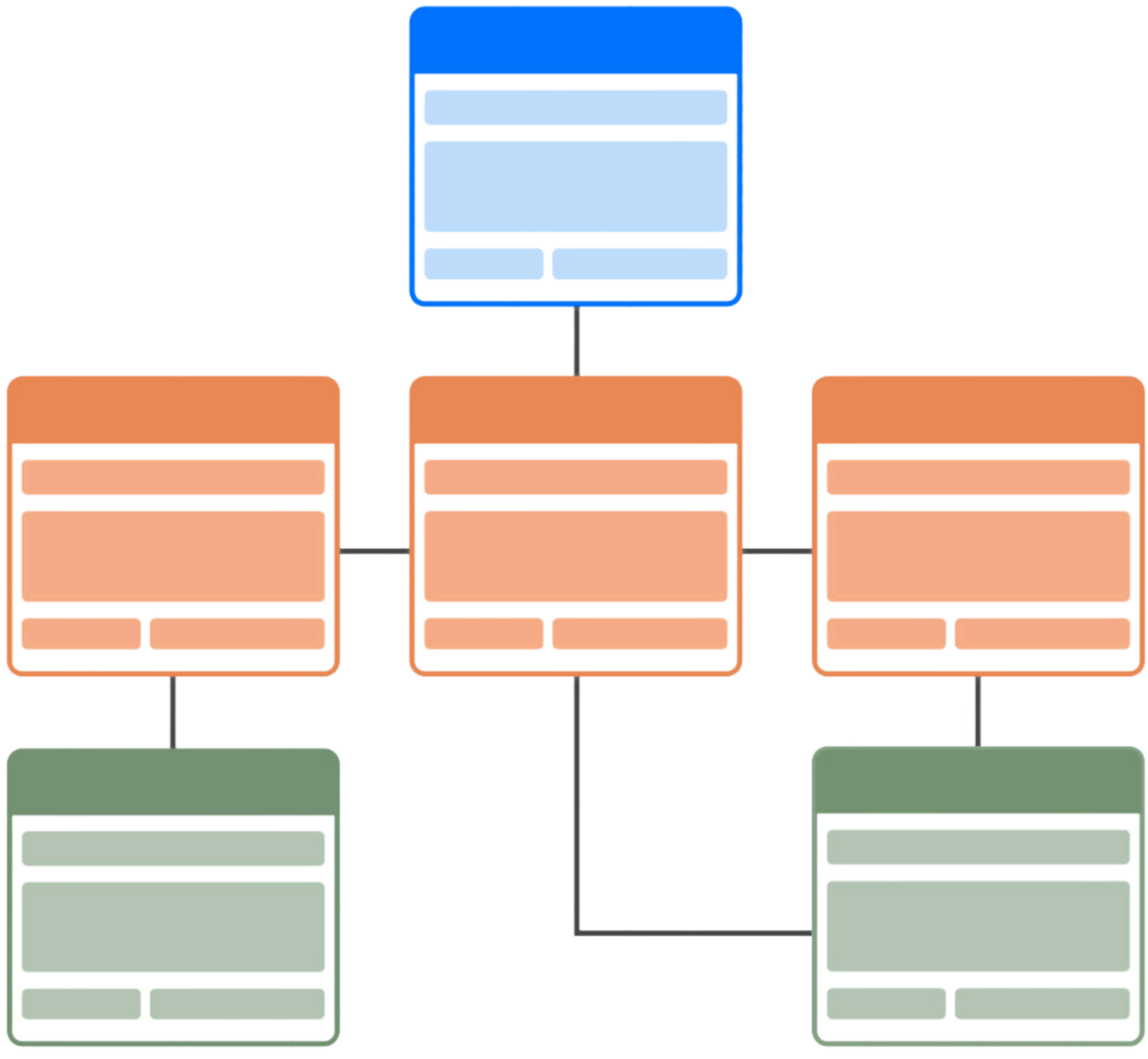
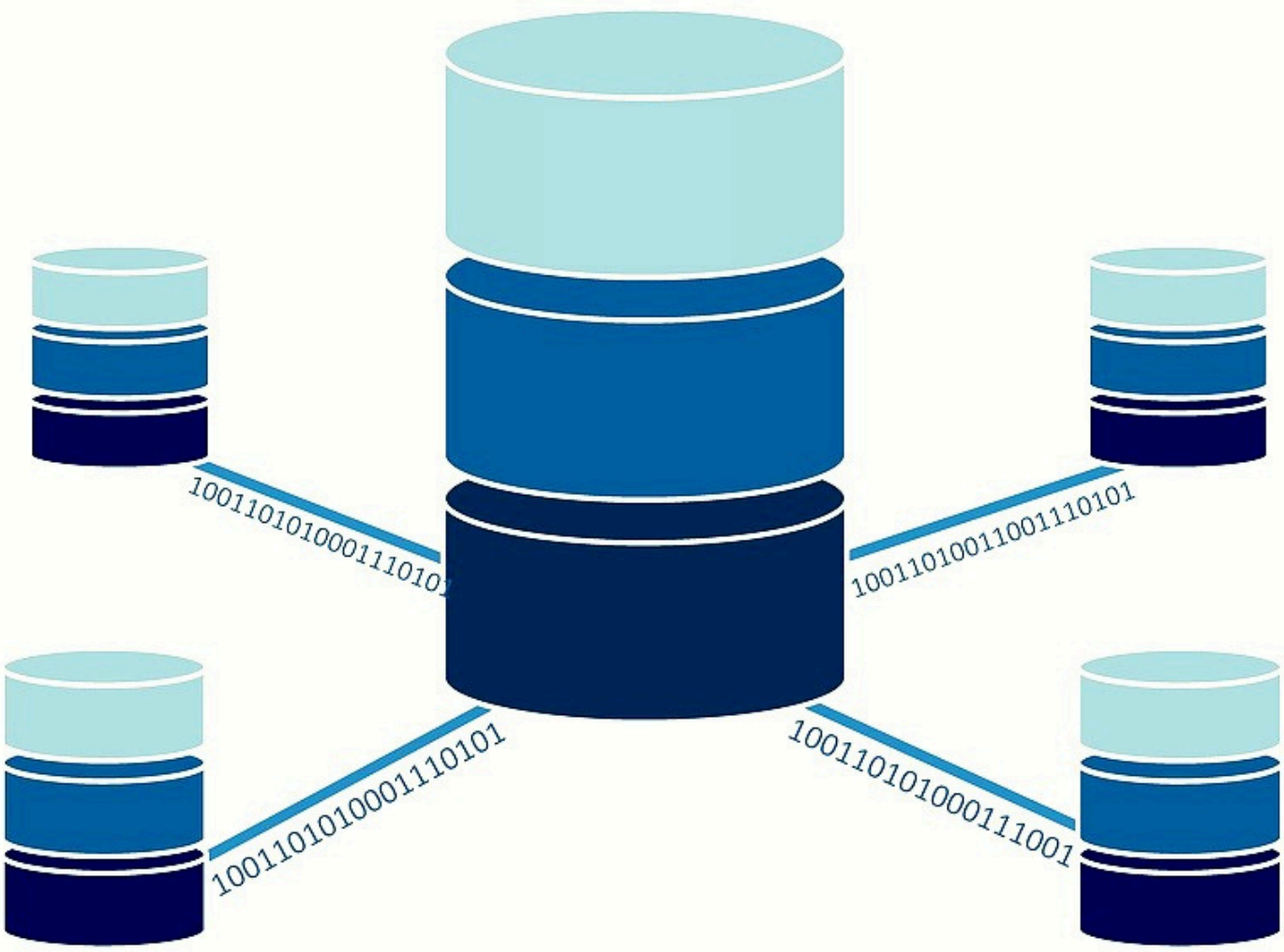
Schemas of Database

A Beginners Guide



What is schema?

It refers to the blueprint that defines how data is organized and stored within the database. It defines the relationships between different tables and the rules for data integrity and constraints. The schema is crucial for ensuring data consistency and facilitating efficient querying and reporting.



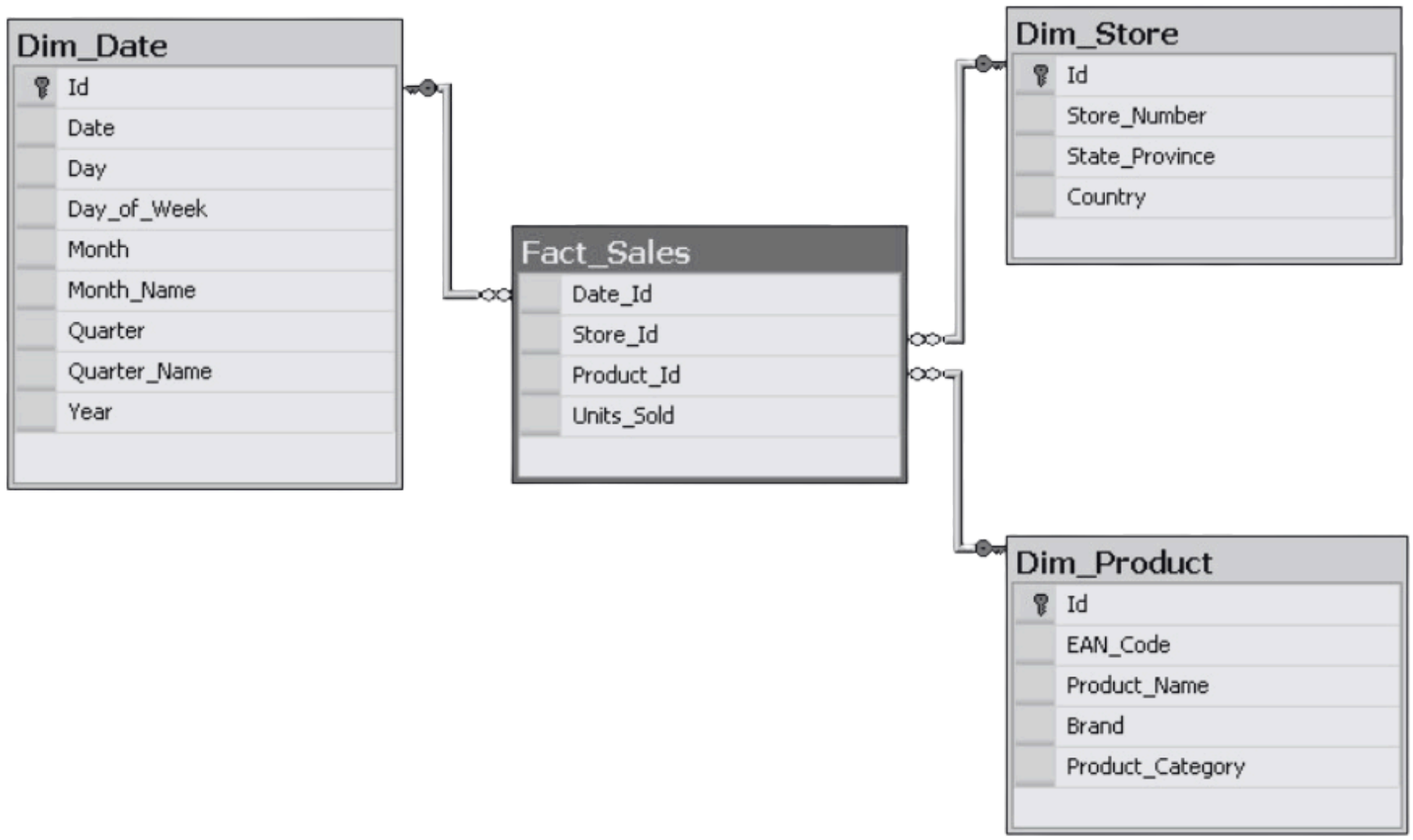
Types of Schemas

3 most commonly used schemas:

- **Star Schema**
- **Snowflake Schema**
- **Galaxy Schema**

Star Schema

- Star schema organizes data around a central fact table.
- It includes dimension tables representing related .
- Simple and suitable for small to medium-sized data warehouses.



ADVANTAGES:

1. Most Suitable for Query Processing: View-only reporting applications show enhanced performance.

2. Simple Queries: Optimized Navigation through the database. It is because the star-join schema logic is much simpler.

3. Simplest and Easiest to design.

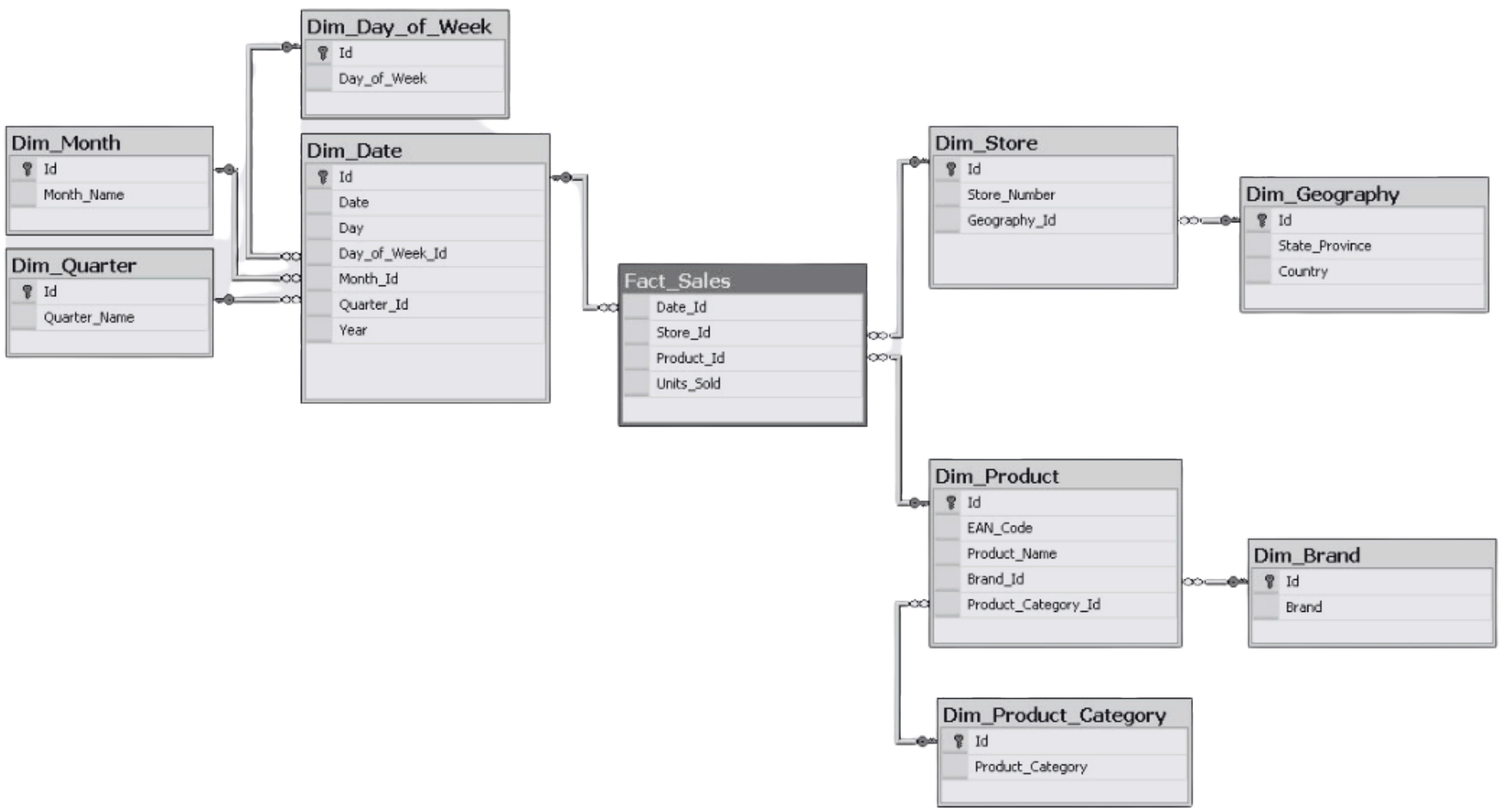
DISADVANTAGES:

1. They don’t support many to many relationships between business entities.

2. More data redundancy: It is a result of each dimension (column) having only one dimension table.

Snowflake Schema

- Snowflake schema extends the star schema with normalized dimension tables.
- Dimension tables are futher split into related tables, reducing redundancy.
- Reduces storage space and improves data integrity.
- Navigation and querying can be more complex than star schema.



ADVANTAGES:

1. Easy to maintain: It is due to reduced data redundancy.

2. Saves Storage space: Dimension tables are easier to update.

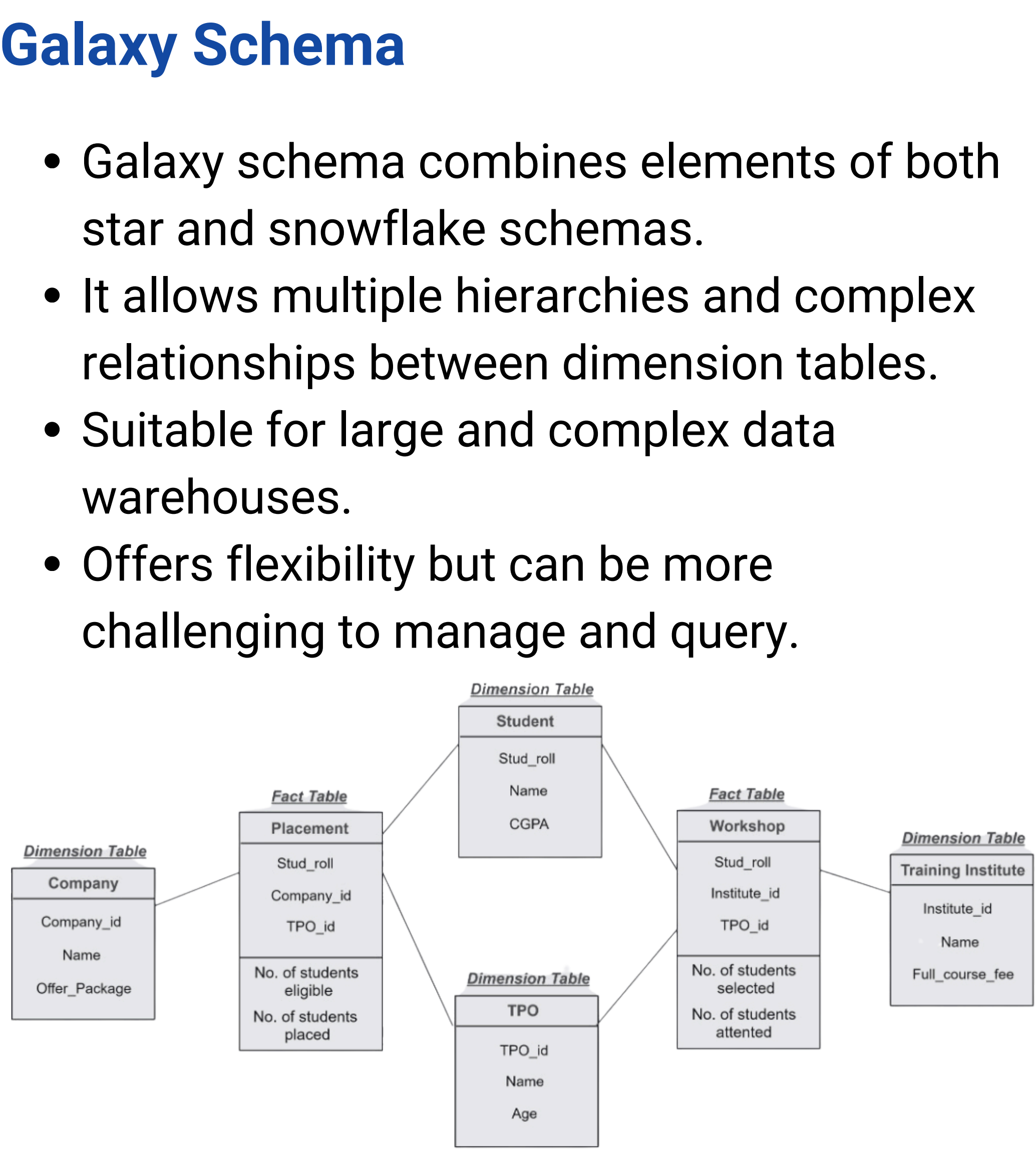
DISADVANTAGES:

1. Complex Schema: Source query joins are complex.

2. Query Performance is not so good: because of the complex queries.

Galaxy Schema

- Galaxy schema combines elements of both star and snowflake schemas.
- It allows multiple hierarchies and complex relationships between dimension tables.
- Suitable for large and complex data warehouses.
- Offers flexibility but can be more challenging to manage and query.



ADVANTAGES:

1. Flexible schema. 2. Effective analysis and reporting.

DISADVANTAGES:

1. Has huge dimension tables hence resulting in difficulty in managing.

2. Hard to maintain: It is because of their complex design and as there are many fact tables.

Follow for more content like this Azure Cloud for Data Engineering

Ganesh R

Azure Data Engineer

https://www.linkedin.com/in/rganesh0203/