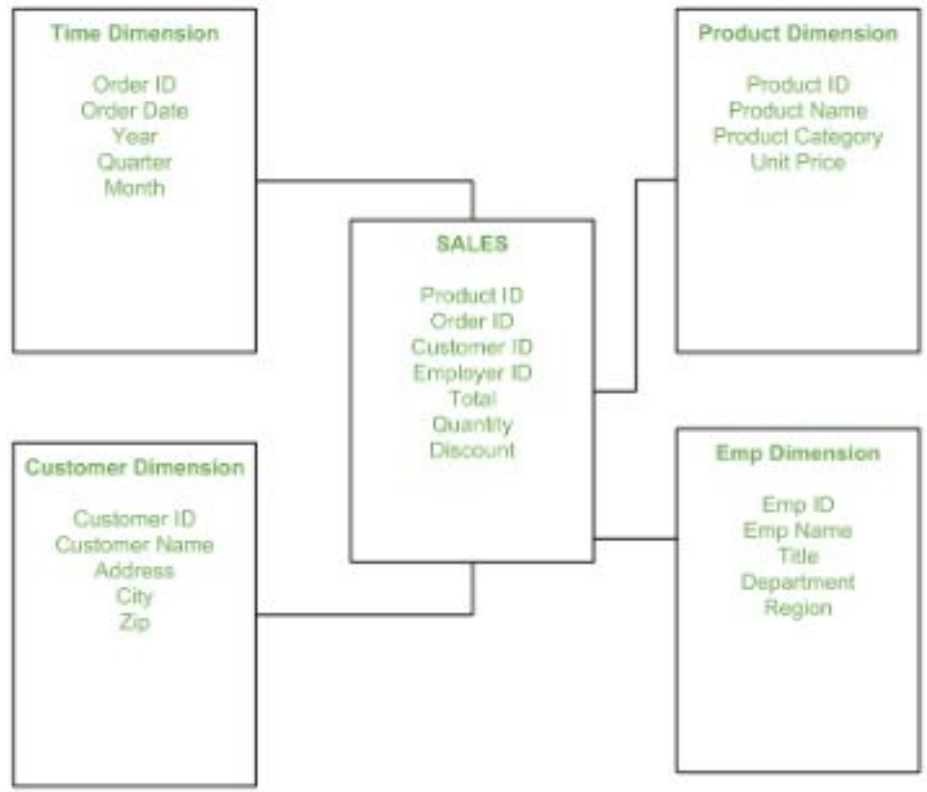
Star Schema

- **Star schema** is the fundamental schema among the data mart schema and it is simplest**. This schema is widely used to develop or build a data warehouse and dimensional data marts**. It includes one or more fact tables indexing any number of dimensional tables. The star schema is a necessary cause of the snowflake schema. It is also efficient for handling basic queries.
- It is said to be star as its **physical model resembles to the star shape** having a **fact table at its center and the dimension tables** at its peripheral representing the star's points. Below is an example to demonstrate the Star Schema:


Time Dimension
Order ID
Order Date
Year
Quarter
Month
Product Dimension
Product ID
Product Name
Product Category
Unit Price
SALES
Product ID
Order ID
Customer ID
Employer ID
Total
Quantity
Discount
Customer Dimension
Customer ID
Customer Name
Address
City
Zip
Emp Dimension
Emp ID
Emp Name
Title
Department
Region

- In the above demonstration, SALES is a fact table having attributes i.e. (Product ID, Order ID, Customer ID, Employer ID, Total, Quantity, Discount) which references to the dimension tables. **Employee dimension table** contains the attributes: Emp ID, Emp Name, Title, Department and Region. *Product dimension table* contains the attributes: Product ID, Product Name, Product Category, Unit Price. *Customer dimension table* contains the attributes: Customer ID, Customer Name, Address, City, Zip. *Time dimension table* contains the attributes: Order ID, Order Date, Year, Quarter, Month.

Model of Star Schema :

- In Star Schema, Business process data, that holds the quantitative data about a business is distributed in fact tables, and dimensions which are descriptive characteristics related to fact data. Sales price, sale quantity, distant, speed, weight, and weight measurements are few examples of fact data in star schema.
  Often, A Star Schema having multiple dimensions is termed as Centipede Schema. It is easy to handle a star schema which have dimensions of few attributes.

Advantages of Star Schema :

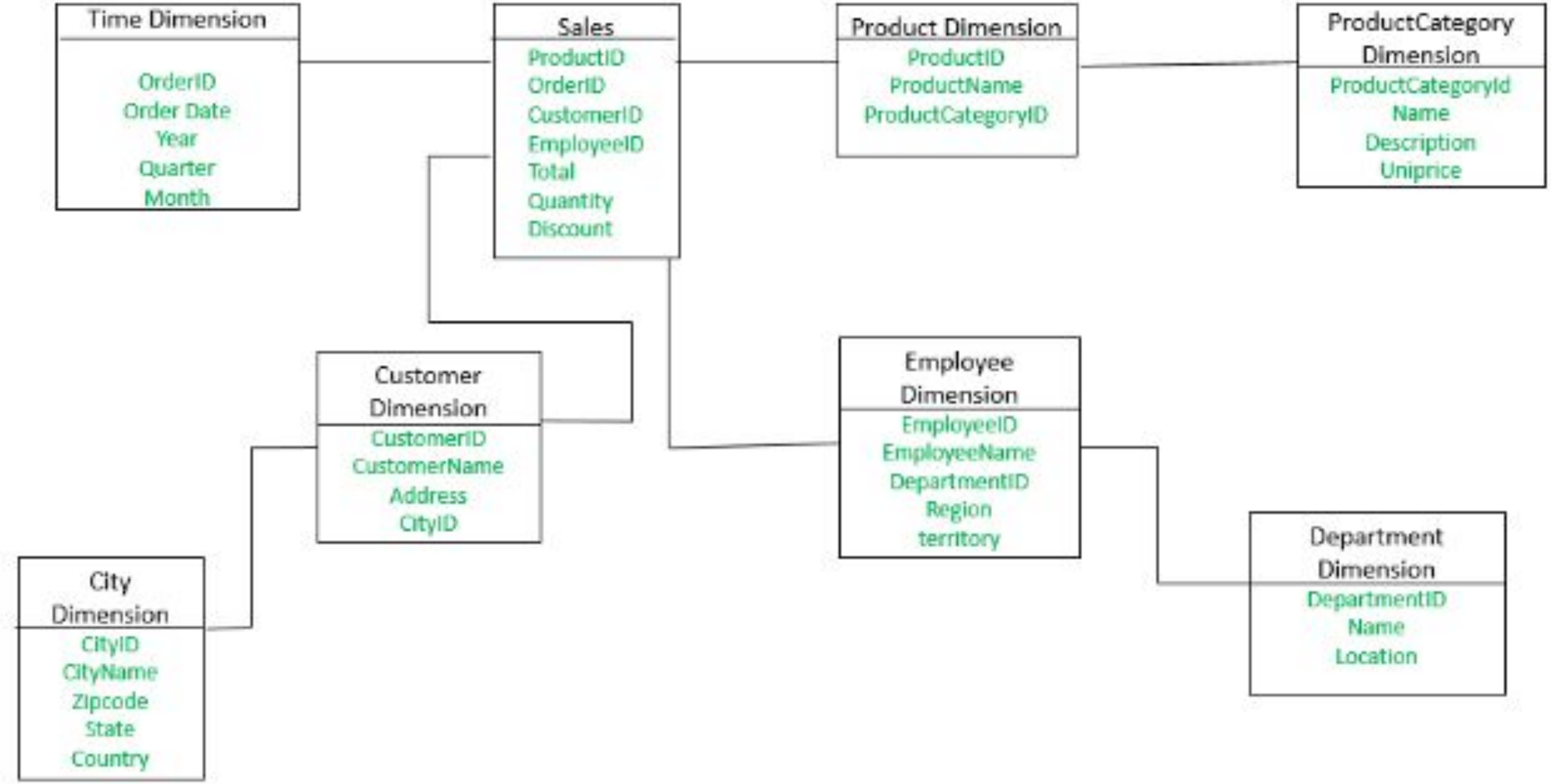
1. **Simpler Queries –**
   Join logic of star schema is quite cinch in comparison to other join logic which are needed to fetch data from a transactional schema that is highly normalized.
2. **Simplified Business Reporting Logic –**
   In comparison to a transactional schema that is highly normalized, the star schema makes simpler common business reporting logic, such as of reporting and period-over-period.
3. **Feeding Cubes –**
   Star schema is widely used by all OLAP systems to design OLAP cubes efficiently. In fact, major OLAP systems deliver a ROLAP mode of operation which can use a star schema as a source without designing a cube structure.

Disadvantages of Star Schema –

1. Data integrity is not enforced well since in a highly de-normalized schema state.
2. Not flexible in terms if analytical needs as a normalized data model.
3. Star schemas don’t reinforce many-to-many relationships within business entities – at least not frequently.

Snowflake Schema

- **Introduction:** The snowflake schema is a **variant of the star schema**. Here, the **centralized fact table is connected to multiple dimensions**. In the **snowflake schema, dimensions are present in a normalized form in multiple related tables**. **The snowflake structure materialized when the dimensions of a star schema are detailed and highly structured**, having several levels of relationship, and the child tables have multiple parent tables. The snowflake effect affects only the dimension tables and does not affect the fact tables.


Time Dimension
OrderID
Order Date
Year
Quarter
Month
Sales
ProductID
OrderID
CustomerID
EmployeeID
Total
Quantity
Discount
Product Dimension
ProductID
ProductName
ProductCategoryID
ProductCategory Dimension
ProductCategoryId
Name
Description
Uniprice
Customer Dimension
CustomerID
CustomerName
Address
CityID
Employee Dimension
EmployeeID
EmployeeName
DepartmentID
Region
territory
City Dimension
CityID
CityName
Zipcode
State
Country
Department Dimension
DepartmentID
Name
Location

- The **Employee** dimension table now contains the attributes: EmployeeID, EmployeeName, DepartmentID, Region, Territory.
- The DepartmentID attribute links with the **Employee** table with the **Department** dimension table. The **Department** dimension is used to provide detail about each department, such as the Name and Location of the department.
- The **Customer** dimension table now contains the attributes: CustomerID, CustomerName, Address, CityID. The CityID attributes link the **Customer** dimension table with the **City** dimension table. The **City** dimension table has details about each city such as CityName, Zipcode, State, and Country.
- The main difference between star schema and snowflake schema is that the **dimension table of the snowflake schema is maintained in the normalized form to reduce redundancy**. The advantage here is that such tables (normalized) are easy to maintain and save storage space. However, it also means that more joins will be needed to execute the query. This will adversely impact system performance.

What is snowflaking?

- The snowflake design is the result of further expansion and **normalized of the dimension table**. In other words, a dimension table is said to be snowflaked if the low-cardinality attribute of the dimensions has been divided into separate normalized tables. These tables are then joined to the original dimension table with referential constraints (foreign key constrain).
  Generally, snowflaking is not recommended in the dimension table, as it hampers the understandability and performance of the dimension model as more tables would be required to be joined to satisfy the queries.

Characteristics of snowflake schema:

- The dimension model of a snowflake under the following conditions:
- The **snowflake schema uses small disk space**.
- It is **easy to implement dimension** that is added to the schema.
- There are **multiple tables, so performance is reduced**.
- The dimension table consists of **two or more sets of attributes that define information at different grains**.
- The sets of **attributes of the same dimension table are being populated by different source systems**.

Advantages:

- There are two main advantages of snowflake schema given below:
- It **provides structured data which reduces the problem of data integrity.**
- It **uses small disk space because data are highly structured**.

Disadvantages:

- Snowflaking **reduces space consumed by dimension tables but compared with the entire data warehouse the saving is usually insignificant.**
- **Avoid snowflaking or normalization of a dimension table, unless required and appropriate**.
- Do not snowflake hierarchies of one dimension table into separate tables. **Hierarchies should belong to the dimension table only and should never be snowflakes.**
- **Multiple hierarchies that can belong to the same dimension have been designed at the lowest possible detail.**