**Experiment 2:** One case study on building Data warehouse/Data Mart

**Aim:** Write Detailed Problem statement and design dimensional modelling (creation of star and snowflake schema)

**Theory:**

Case study:

The case study revolves around the application and significance of data warehousing and mining techniques within the context of a library management system. In an era where libraries are adapting to digital landscapes, the utilization of data-driven strategies has become imperative. This case study offers an in-depth exploration of the amalgamation of data warehousing and mining to extract valuable insights from the vast reservoir of data encompassing user activities, resource utilization, and operational records within a library system.

The case study delves into the essence of data mining, underscoring its pivotal role in unearthing concealed patterns, trends, and interconnections within the realm of library data.

Star Schema:

A star schema is a type of data modeling technique used in data warehousing to represent data in a structured and intuitive way. In a star schema, data is organized into a central fact table that contains the measures of interest, surrounded by dimension tables that describe the attributes of the measures.

The fact table in a star schema contains the measures or metrics that are of interest to the user or organization. For example, in a sales data warehouse, the fact table might contain sales revenue, units sold, and profit margins. Each record in the fact table represents a specific event or transaction, such as a sale or order.

The dimension tables in a star schema contain the descriptive attributes of the measures in the fact table. These attributes are used to slice and dice the data in the fact table, allowing users to analyze the data from different perspectives. For example, in a sales data warehouse, the dimension tables might include product, customer, time, and location.

In a star schema, each dimension table is joined to the fact table through a foreign key relationship. This allows users to query the data in the fact table using attributes from the dimension tables.

Some features of star schema:

* Central Fact Table
* Dimension Tables
* Denormalized structure
* Simple queries
* Aggregated data
* Fast performance
* Easy to understand

Advantages of Star Schema :

* Simpler Queries –

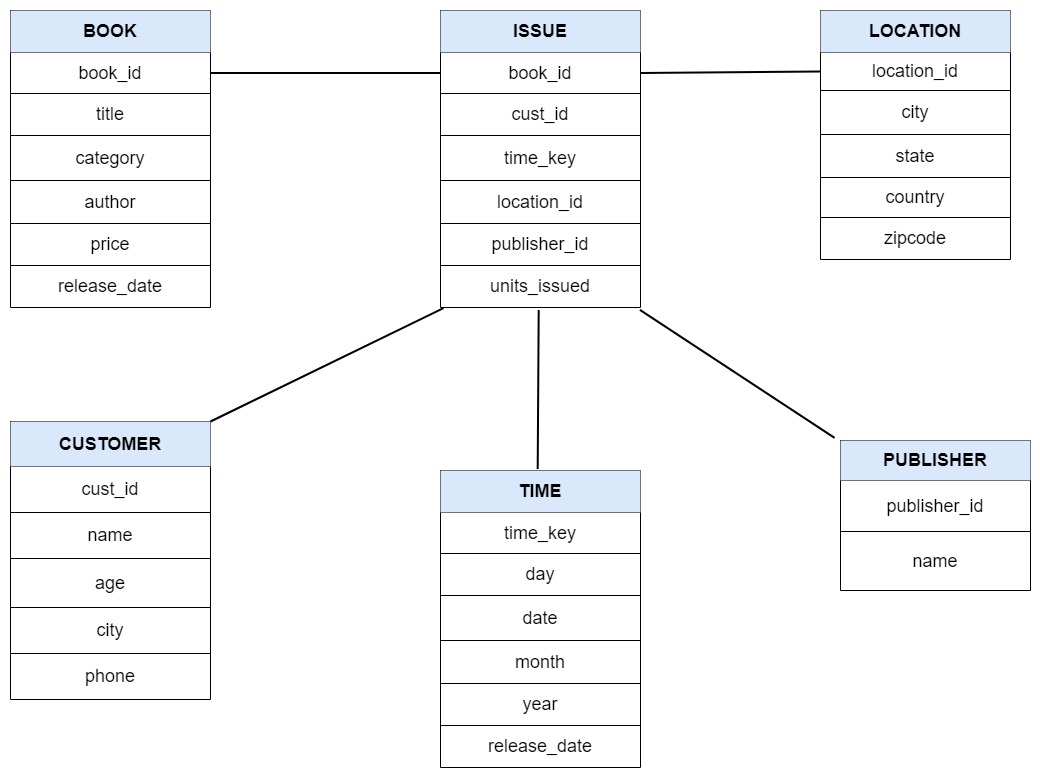
Join logic of star schema is quite cinch in comparison to other join logic which are needed to fetch data from a transactional schema that is highly normalized.

* 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.

* 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.



**Fig 2.1 Star Schema of Library Management**

Snowflake Schema:

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.

A snowflake schema is a type of data modeling technique used in data warehousing to represent data in a structured way that is optimized for querying large amounts of data efficiently. In a snowflake schema, the dimension tables are normalized into multiple related tables, creating a hierarchical or “snowflake” structure.

In a snowflake schema, the fact table is still located at the center of the schema, surrounded by the dimension tables. However, each dimension table is further broken down into multiple related tables, creating a hierarchical structure that resembles a snowflake.

For Example, in a sales data warehouse, the product dimension table might be normalized into multiple related tables, such as product category, product subcategory, and product details. Each of these tables would be related to the product dimension table through a foreign key relationship.

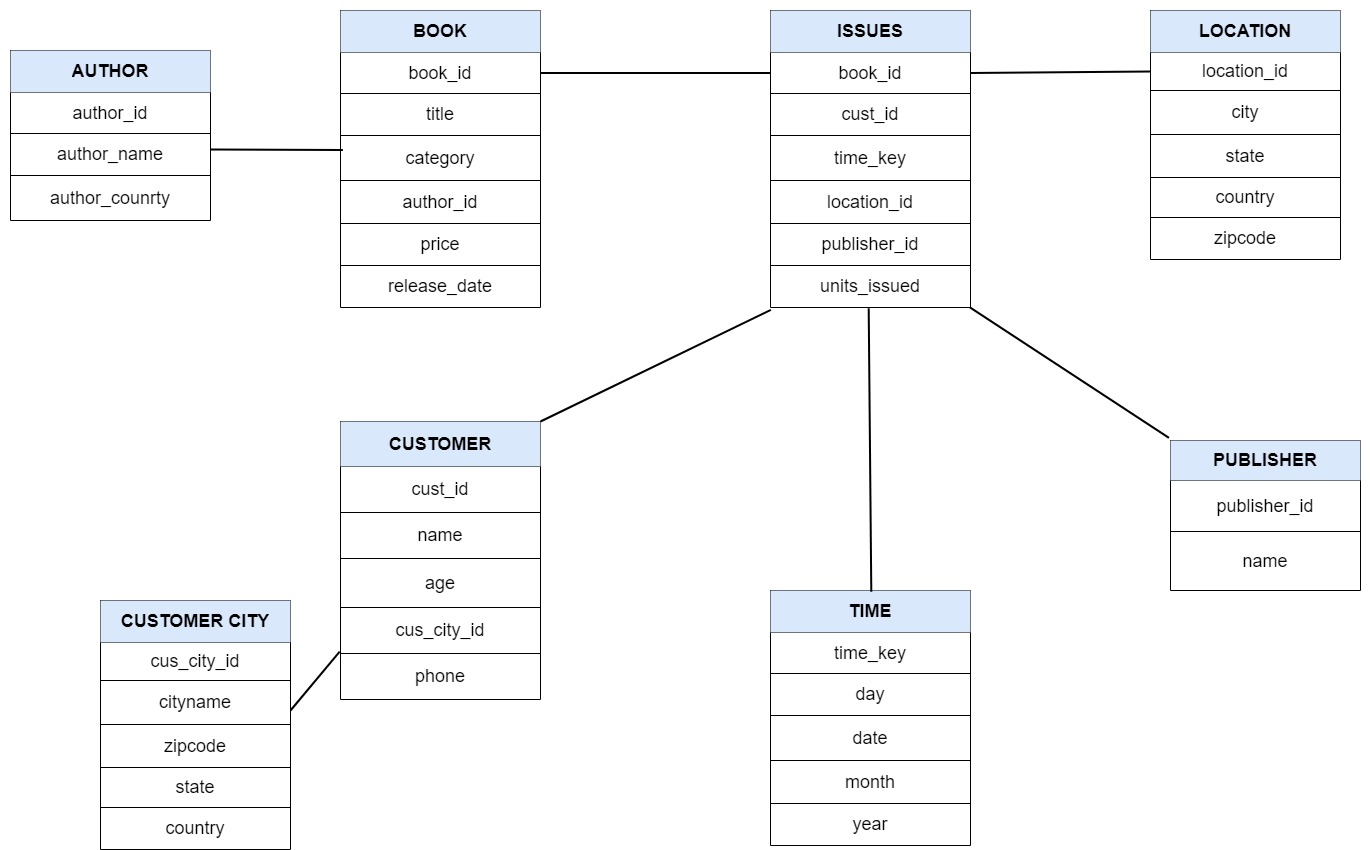
The snowflake design is the result of further expansion and normalization of the dimension table. In other words, a dimension table is said to be snowflaked if the low-cardinality attribute of the dimensions has been divided into separate normalized tables.

Features of Snowflake Schema:

* Normalization
* Hierarchical
* Multiple levels
* Joins
* Scalability

Advantages of Snowflake Schema

* It provides structured data which reduces the problem of data integrity.
* It uses small disk space because data are highly structured.



**Fig 2.2 Snowflake schema of Library Management**

**Conclusion:**

In conclusion, star and snowflake schema both provide an aggregated view of the data sets and both of them have their own differences, advantages and disadvantages. While star schema provides a more aggregated straight forward perspective of the data, while snowflake provides a normalised structure which helps in minimising redundancy. he 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.

However, the snowflake schema can also be more complex to query than a star schema because it requires more table joins.

The decision to use a snowflake schema versus a star schema in a data warehousing project will depend on the specific requirements of the project and the trade-offs between query performance, schema complexity, and data integrity.