Aim: To Build a DataWarehouse – Star Schema and Snowflake Schema

Objective: A data warehouse is a large store of data collected from multiple sources within a business. The objective of data ware house system is to provide consolidated, flexible, meaningful data storage to the end user for reporting and analysis.

Theory:

In general, the warehouse design process consists of the following steps:

1. Choose a business process to model (e.g., orders, invoices, shipments, inventory, account administration, sales, or the general ledger). If the business process is organizational and involves multiple complex object collections, a data warehouse model should be followed. However, if the process is departmental and focuses on the analysis of one kind of business process, a data mart model should be chosen.
2. Choose the business process grain, which is the fundamental, atomic level of data to be represented in the fact table for this process (e.g., individual transactions, individual daily snapshots, and so on).
3. Choose the dimensions that will apply to each fact table record. Typical dimensions are time, item, customer, supplier, warehouse, transaction type, and status.
4. Choose the measures that will populate each fact table record. Typical measures are numeric additive quantities like dollars sold and units sold.

Star Schema:

The most common modeling paradigm is the star schema, in which the data warehouse contains:

- a large central table (fact table) containing the bulk of the data, with no redundancy, and
- a set of smaller attendant tables (dimension tables), one for each dimension.

Snowflake Schema:

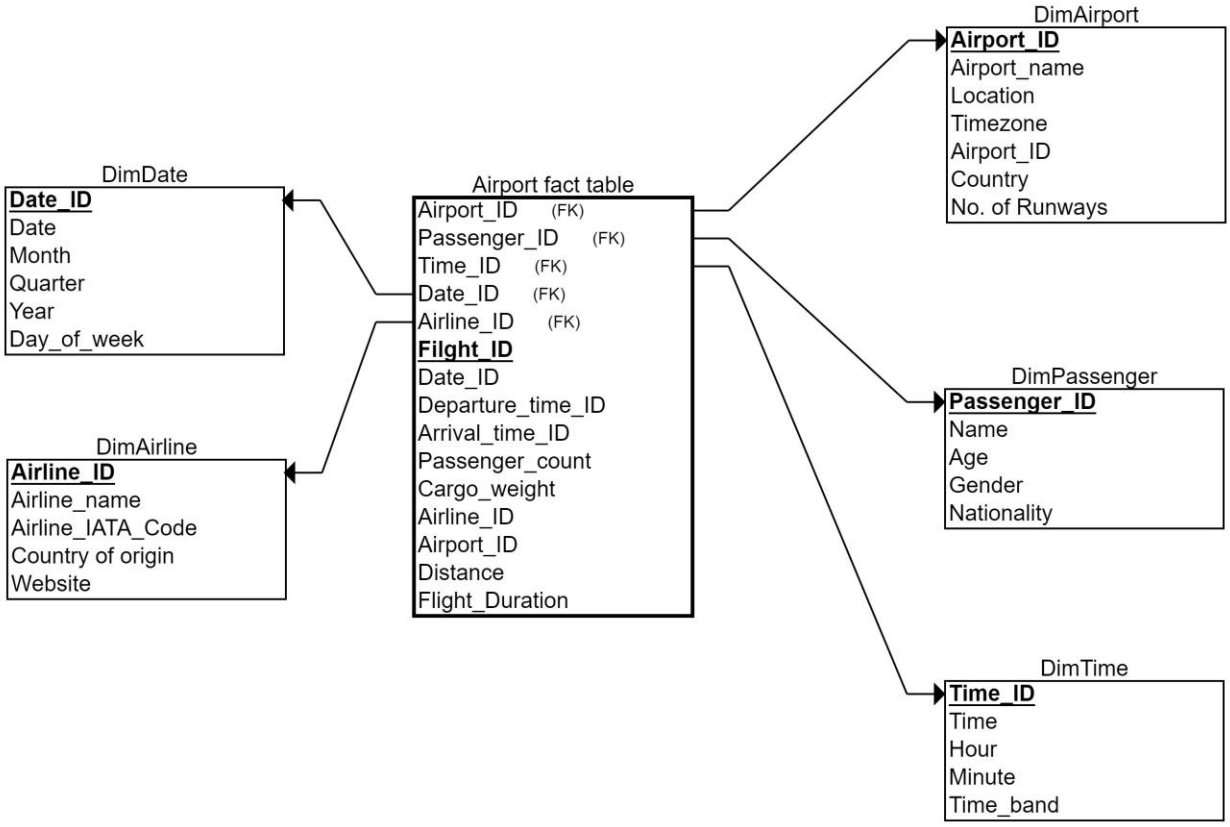
- The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables.
- The resulting schema graph forms a shape similar to a snowflake.
- The major difference between the snowflake and star schema models is that the dimension tables of the snowflake model may be kept in normalized form to reduce redundancies.
- Such a table is easy to maintain and saves storage space.
- However, this space savings is negligible in comparison to the typical magnitude of the fact table.

Problem Statement:

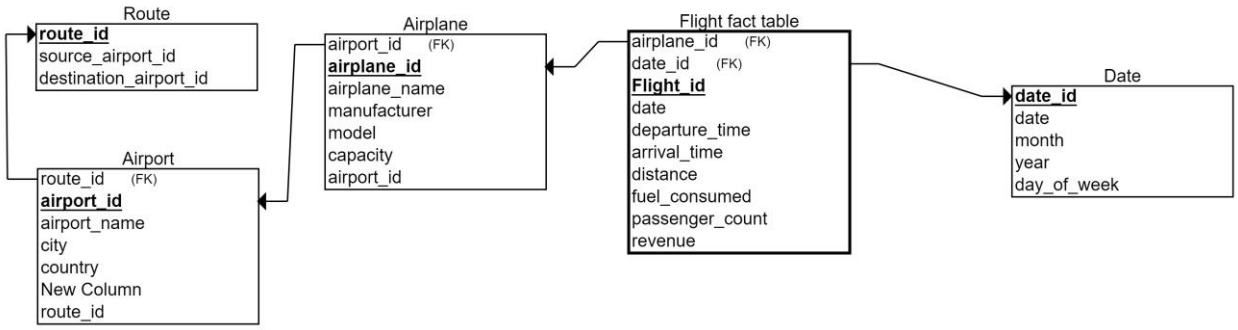
The problem is to design and implement a data warehousing solution for a bookstore that optimally organizes and manages its vast data, including sales, inventory, customer information, and more, to facilitate efficient reporting and analytics. This involves creating both a star schema and a snowflake schema to support various business intelligence and decision-making processes, while ensuring data accuracy, integrity, and performance

Construction of Star schema and Snowflake schema:

Star Schema:



Snowflake Schema:



Conclusion:

In conclusion, the creation of star and snowflake schemas for the bookstore's data warehousing project will provide an efficient and effective solution for managing and analyzing data, enabling informed decision-making and enhancing overall business performance.