Aim: To study and design Star schema and Snowflake schema

Theory:

- Star and snowflake schema designs are mechanisms to separate facts and dimensions into separate tables
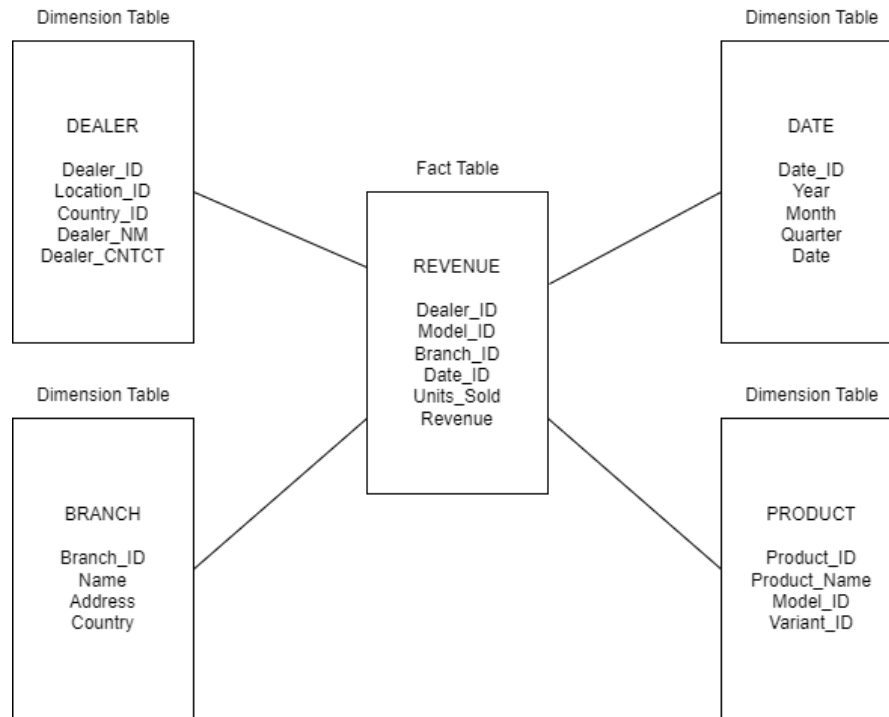
STAR SCHEMA

- Star schema is the fundamental schema among the data mart schema and it is simplest.
- 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.

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.

Dimension Table

DEALER

Dealer_ID
Location_ID
Country_ID
Dealer_NM
Dealer_CNTCT

Fact Table

REVENUE

Dealer_ID
Model_ID
Branch_ID
Date_ID
Units_Sold
Revenue

Dimension Table

DATE

Date_ID
Year
Month
Quarter
Date

Dimension Table

BRANCH

Branch_ID
Name
Address
Country

Dimension Table

PRODUCT

Product_ID
Product_Name
Model_ID
Variant_ID

SNOWFLAKE SCHEMA

- The snowflake schema is a variant of the star schema.
- Here, the centralized fact table is connected to multiple dimensions.
- 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.
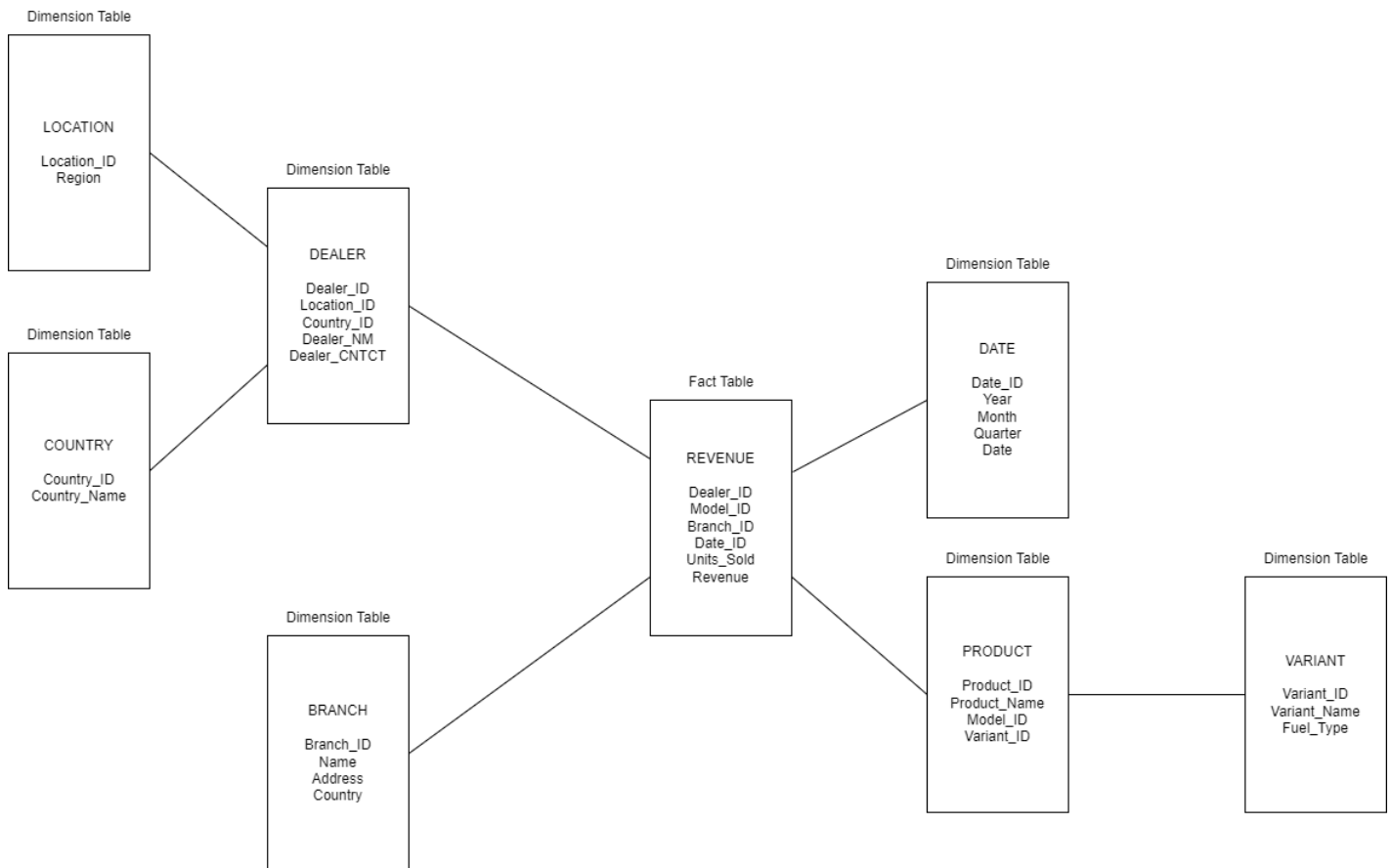
Advantages of Snowflake schema:

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

Disadvantages of Snowflake schema:

1. Snowflaking reduces space consumed by dimension tables but compared with the entire data warehouse the saving is usually insignificant.
2. Avoid snowflaking or normalization of a dimension table, unless required and appropriate.
3. Do not snowflake hierarchies of one dimension table into separate tables. Hierarchies should belong to the dimension table only and should never be snowflakes.

4. Multiple hierarchies that can belong to the same dimension have been designed at the lowest possible detail.

Dimension Table

LOCATION
Location_ID
Region

Dimension Table

DEALER
Dealer_ID
Location_ID
Country_ID
Dealer_NM
Dealer_CNTCT

Dimension Table

COUNTRY
Country_ID
Country_Name

Fact Table

REVENUE
Dealer_ID
Model_ID
Branch_ID
Date_ID
Units_Sold
Revenue

Dimension Table

DATE
Date_ID
Year
Month
Quarter
Date

Dimension Table

PRODUCT
Product_ID
Product_Name
Model_ID
Variant_ID

Dimension Table

VARIANT
Variant_ID
Variant_Name
Fuel_Type

Dimension Table

BRANCH
Branch_ID
Name
Address
Country

Difference between Star and Snowflake Schemas:

S.NO	Star Schema	Snowflake Schema
1.	In star schema, The fact tables and the dimension tables are contained.	While in snowflake schema, The fact tables, dimension tables as well as sub dimension tables are contained.
2.	Star schema is a top-down model.	While it is a bottom-up model.
3.	Star schema uses more space.	While it uses less space.
4.	It takes less time for the execution of queries.	While it takes more time than star schema for the execution of queries.
5.	In star schema, Normalization is not used.	While in this, Both normalization and denormalization are used.
6.	It's design is very simple.	While it's design is complex.
7.	The query complexity of star schema is low.	While the query complexity of snowflake schema is higher than star schema.

8.	It's understanding is very simple.	While it's understanding is difficult.
9.	It has less number of foreign keys.	While it has more number of foreign keys.
10.	It has high data redundancy.	While it has low data redundancy.

Conclusion:

Successfully studied and designed Star schema and Snowflake schema.