DATA WAREHOUSING LECTURE 8

ENGR. MADEHA MUSHTAQ

DEPARTMENT OF COMPUTER SCIENCE

IQRA NATIONAL UNIVERSITY

THE SNOWFLAKE SCHEMA

- "Snowflaking" is a method of normalizing the dimension tables in a STAR schema.

- When we completely normalize all the dimension tables, the resultant structure resembles a snowflake with the fact table in the middle.

- Let us begin with the below Figure, which shows a simple STAR schema for sales in a manufacturing company.
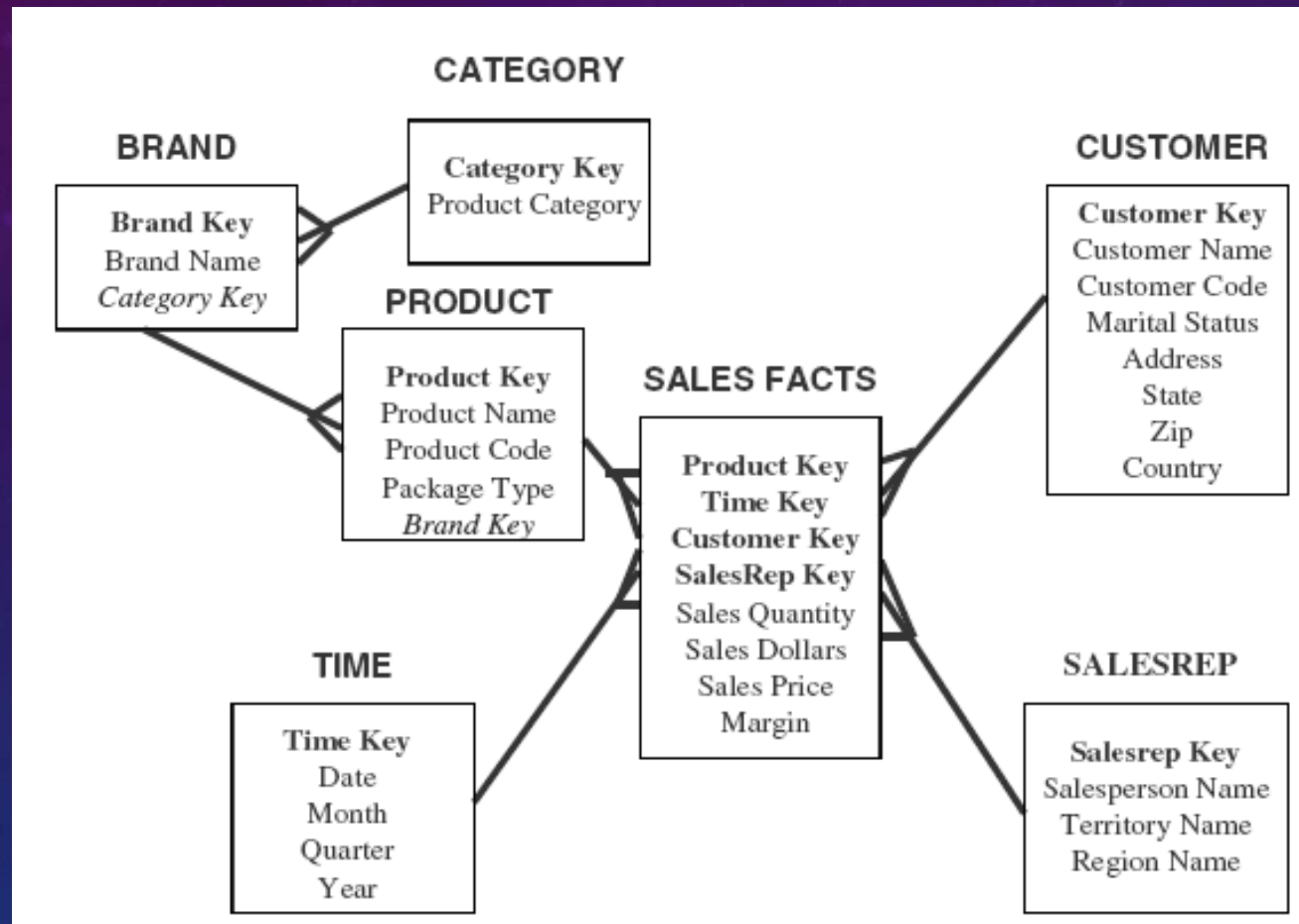
THE SNOWFLAKE SCHEMA

- This is a classic STAR schema, denormalized for optimal query access involving all or most of the dimensions.

- The model is not in the third normal form.

THE SNOWFLAKE SCHEMA
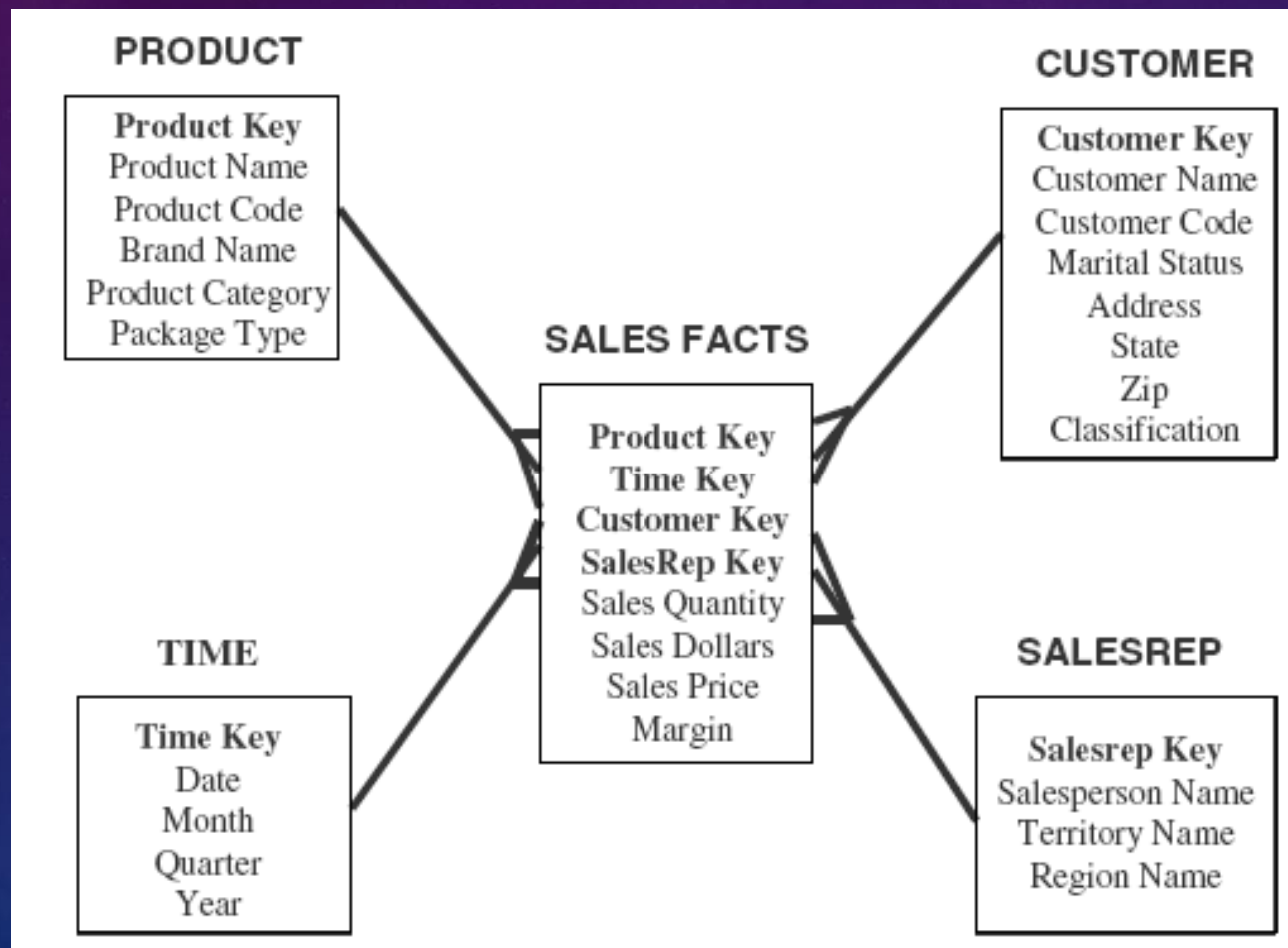
THE SNOWFLAKE SCHEMA

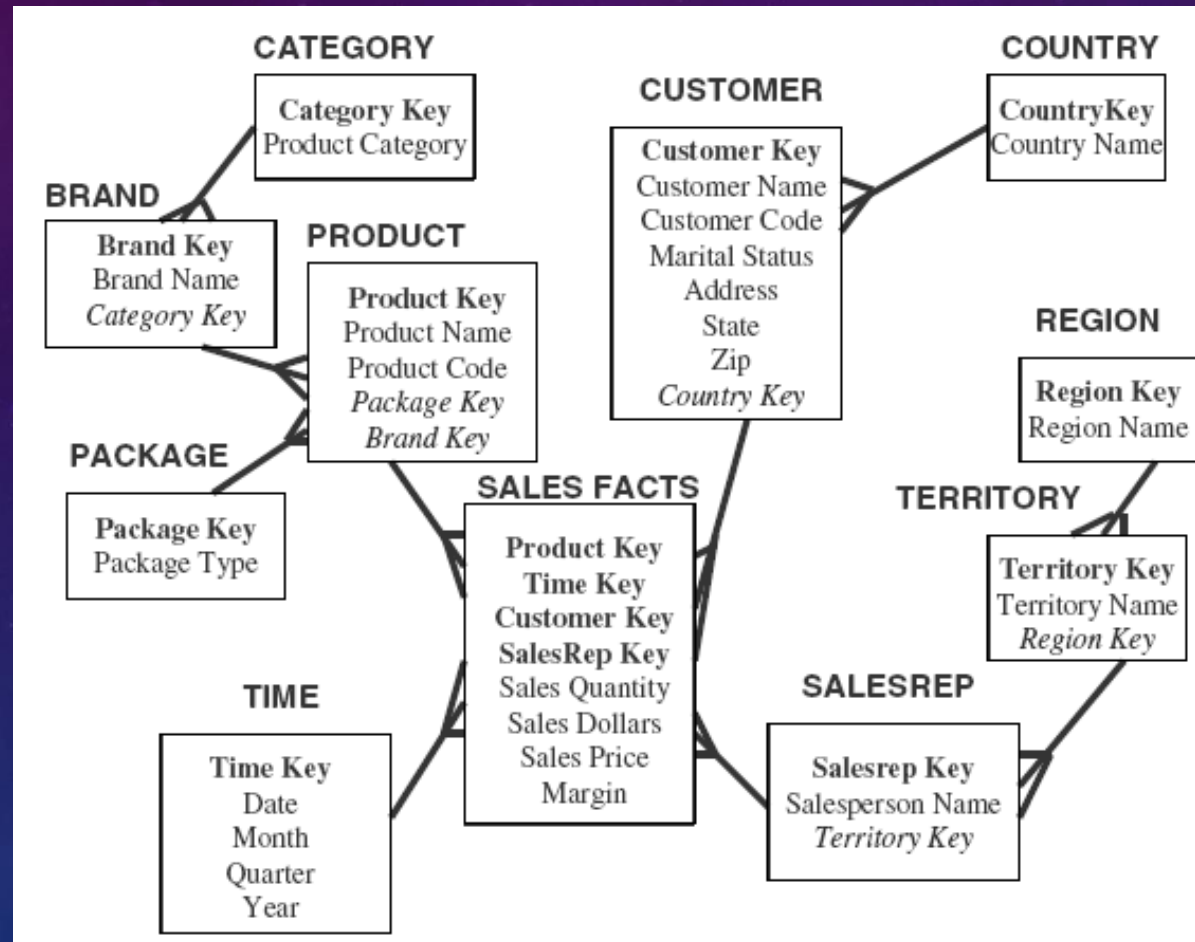Product dimension: partially normalized

THE SNOWFLAKE SCHEMA

- In Figure above, we have not completely normalized the product dimension.

- We can also move other attributes out of the product dimension table and form normalized structures.

- "Snowflaking" or normalization of the dimension tables can be achieved in a few different ways.

THE SNOWFLAKE SCHEMA

- The following options indicate the different ways we may want to consider for normalization of the dimension tables:
  - Partially normalize only a few dimension tables, leaving the others intact
  - Partially or fully normalize only a few dimension tables, leaving the rest intact
  - Partially normalize every dimension table
  - Fully normalize every dimension table.

THE SNOWFLAKE SCHEMA

Every dimension table partially or fully normalized

THE SNOWFLAKE SCHEMA

- The original STAR schema for sales contains only five tables.

- Whereas the normalized version now extends to eleven tables.

- We will notice that in the snowflake schema, the attributes with low cardinality in each original dimension table are removed to form separate tables.

- These new tables are linked back to the original dimension table through artificial keys.

ADVANTAGES

- Small savings in storage space
- Normalized structures are easier to update and maintain.

DISADVANTAGES

- Schema less intuitive and end-users are put off by the complexity.

- Ability to browse through the contents difficult .

- Degraded query performance because of additional joins.

- Snow flaking is not generally recommended in a data warehouse environment.

- Query performance takes the highest significance in a data warehouse and snow flaking hampers the performance.

END OF SLIDES