Fact Tables and Dimension Tables in Data Warehousing and Data Mining

Mokshad Ketan Sankhe

In data warehousing, fact tables and dimension tables are two key components that store different types of data.

Fact Tables:

Fact tables contain quantitative data that gets measured and stored. This is usually numeric data that can be aggregated. Examples of facts include sales amounts, units sold, inventory counts, etc.

Some key characteristics of fact tables:

- Store numeric facts that can be aggregated and analysed.
- Contain foreign keys to link to dimension tables.
- Can contain large volumes of data.
- New records get added as new facts get captured.

Dimension Tables:

Dimension tables contain descriptive attributes and contextual information about the facts stored in fact tables. Dimension attributes help categorize and label facts. Examples of dimensions include product, customer, time, location etc.

Some key characteristics of dimension tables:

- Store qualitative, descriptive attributes.
- Contain low volume of data that changes infrequently.
- Gets referenced by fact tables via foreign keys.
- Can be denormalized into star schema for simplify joins.

Together, fact and dimension tables allow analytical queries that aggregate metrics by different attributes. For example, total sales by product, or revenue by geography.

In data mining, fact and dimension tables provide the structured data inputs for applying data mining algorithms. The mining models can analyze and detect patterns in fact metrics across the various dimensions. This enables gaining new business insights from historical data.

Conclusion:

Fact and dimension modelling creates the foundational data structure for data warehousing and data mining. Fact tables store the measurable business events and metrics, while dimensions add contextual descriptors to the facts. This provides a flexible analytical data framework to gain insights from across multiple perspectives.