Data Warehouse

It is a collection of time-variant, non-volatile, subject-oriented data of an organization.

Data Warehouse (DW) is stored as a multidimensional cube.

![Diagram of a multidimensional cube]

\[
\begin{array}{c}
10000 \quad \text{Product ID} \\
1000 \quad \text{Customer ID} \\
1 \quad \text{Week} \end{array}
\]

\[
52 \times 1000 \times 10000
\]

Cell \((3, 150, 35) \Rightarrow \text{int value (sales quantity)}\)

Since many cells would be empty, sparse matrix is used for data representation.

Purpose of DW analysis is to predict the behavioral patterns (of customers) from the past data set.

Operations on a DW

- **Slicing** reduces the dimensions
  - 4D \(\to\) 3D, 3D \(\to\) 2D, 2D \(\to\) 1D
- **Dicing**
- **Drill down** start with an N-D to a specific cell
- **Roll up** start with a cell to N-D cube
Datawarehouse Model

a) Star Schema

- Customer
- Sale
- Key Event
- Brand
- Time
- Product
- Discount

1-hop

b) Snowflake Schema (Graph) multi-hop

Processing

- OLAP
  - Online Analytical Processing
  - Complex operations
  - Requires substantial processing

- OLTP
  - Online Transaction Processing
  - Smaller transactions
  - Requires quick response (short response time)