Data warehouse

Data warehouse is an integrated collection of data extracted from operational, historical and external databases that have been cleaned, transformed and cataloged for retrieval and analysis of to provide business intelligence for business decision making. Data warehouses may be divided into data marts which are the subsets of data that focus on specific aspects of a company.

- Data: Raw piece of information that is capable of being moved and store.
- Database: An organized collection of such data in which data are managed in tabular form with relationship.
- Data Warehouse: System that organizes all the data available in an organization makes it accessible & usable for the all kinds of data analysis and also allows creating lots of reports by the use of mining tools.

  "A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process."

  Data warehouse integrated: Constructed by integrating multiple, heterogeneous data sources, relational databases, flat files, on-line transaction records

  Data cleaning and data integration techniques are applied. Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources

  E.g., Hotel price: currency, tax, breakfast covered, etc.

  Data warehouse is subject oriented: Organized around major subjects, such as customer, product, sales. Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing. Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.

  Data warehouse is time variant: The time horizon for the data warehouse is significantly longer than that of operational systems. Operational database: current value data. Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)

  Data warehouse is non volatile: A physically separate store of data transformed from the operational environment. Operational update of data does not occur in the data warehouse environment.

Advantage of data warehouse to organization

1. Information processing: supports querying, basic statistical analysis and reporting using crosstabs, tables and graphs.
2. Analytical processing: multidimensional analysis of data warehouse data also supports basic OLAP operations, slice dice, drilling etc.
3. Data mining
4. It provides business users with a customer-centric view of the companies heterogeneous data by helping to integrate data from sales, service manufacturing and distribution and other customer related business system.
5. Enhanced customer service
6. More cost effective decision making
7. Business reengineering

Strategic uses of data ware housing

1. Airlines
2. Heath care
3. Personal care
4. telecommunication

Data mining

The process of Discovering meaningful patterns & trends often previously unknown, by shifting large amount of data, using pattern recognition, statistical and Mathematical techniques. A group of techniques that find relationship that have not previously been discovered. Data mining is a young discipline with wide and diverse applications, some application domains

- Biomedical and DNA data analysis
- Financial data analysis
- Retail industry
- Telecommunication industry

Application of data mining

1. Market segmentation: identifies the common characteristics of customers who buy the same products from your company.
2. Fraud detection: identifies which transaction is most likely to be fraudulent.
3. Market basket analysis: understands what products or services are commonly purchased together.
4. Find root causes of quality or manufacturing problems.
5. Cross-sell to existing customers.

Advantage of data mining

i. Automated perdition of trends and behaviors.
ii. Automated discovery of previously unknown patterns
iii. Database can be larger in both depth and breath.

Disadvantage of data mining

i. Privacy issues
ii. Security issues
iii. Misuse of information