Short Topics

Logical Databases

Logical databases provide read only access to a group of related tables. Logical databases are SQL statements which read data from other databases just like a stored query in MS-Access. It provides a view of database tables.

The usage of Logical databases:

- Checking functions for appropriateness of user input.
- Summarization of data.
- Centralized authorization for DB access.
- Better read performance for tables hierarchically linked.

Data Warehousing

Def-1 A data warehouse is a subject oriented, Integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making.
Def-2- It is a logical collection of information gathered from many different operational databases and used to create business intelligence that supports business analysis activities and decision making tasks.
Def-3- It is an integrated and consistent store of subject-oriented data that is obtained from a variety of resources and formatted into a meaningful context to support decision making in an organization.

The need for Data Warehousing:

- Integrated Organization-wide view of contextual information.
- Separation of Operational data and Informational data.
- Data stores hold current data, not historical. So historical summarization is not possible in operational data stores.
- Decision making need integrated, analyzed, most recent information brought from operational data.

The Data Warehousing helps in:

- Study of historical trends.
- DSS Queries need not depend on Transaction Processing systems.

Obtaining Data for the warehouse:

- Data sources are checked in regular intervals for new data and extracted.
- Generally data for the warehouse is obtained in non-peak hours as it is quite acceptable to have slightly out dated data in the warehouse.
- Many a times, during extraction, data from different tables are mixed to suit the needs of the warehouse.
- Warehouse schema may be a materialized view. (summarization, generally aggregate functions are applied and data like subtotals, summations, averages are extracted.)

Data Mining

Data mining is a semi-automatic process of analyzing large databases to find interesting and useful pattens applicable for business decision making.

Generally data mining is used to find inter-relations between different aspects of business like buying patterns, seasonal sales, multi-product mix etc. Generally these patterns are used to predict the future of the business based on patterns obtained from past data. We study the patterns available through computerized projections and try to get a theory out of it.

Some general usage of Data Mining:

- Classification
  Given a new item whose class is unknown, prediction will be made about its class.
- Regression
  Given a set of mappings of an unknown function, prediction will be made on the result of the function on a new set of values.
- Association
  Preparing associations by analysis of past data (Put forth a theory from past patterns like relation between Gutkha and Oral-cancer).
- Clustering
  Preparing association of masses to single causes. (Put forth a theory linking a single event with some large scale outcome of similar pattern like Increased Toxin in blood for people living near Ammonia Factory)

Why to use Data-Mining?

Human efforts can't analyze the large volumes of business and social data which grows by leaps and bounds day by day.

Computerized analysis is possible due to High processing power, Large storage, Ready made software packages and technical expertise availability.

Data Mining Process:

- Selection of data
- Cleaning data for not-to-be-used extraneous information.
- Remove noise (Data that can affect the purity of other data)
- Transformation of data with calculations.
- Getting commonalities between different data.
- Generation of new information.

OLAP and OLTP

On-line Analytical Processing	On-line Transaction Processing
Act on Historical data	Act on Current Data
M-D Schema (Store – show etc.)	I-D Schema
Finding out a summary	Processing of real data
Acts on De-normalized data	Acts on Normalized data
Queries are very long and complex	Many small transactions
Used by the decision maker (Manager, Analyst, Top officials)	Used by Low-end Users.