Computer Science 202

Data-Warehousing

Objectives

- The difference between operational data and decision support
- What a data warehouse is and how its data are prepared
- What star schemas are and how they are constructed
- What steps are required to implement a data warehouse successfully
- What data mining is and what role it plays in decision support

---

Why Data Analysis?

- External and internal forces require tactical and strategic decisions
- Search for competitive advantage
- Business environments are dynamic
- Decision-making cycle time is reduced
- Different managers require different decision support systems (DSS)

| COMPANY	PROBLEM	BENEFIT
Moen, Inc. Manufacturer of bathroom and kitchen fixtures and supplies Source: Cognos Corp. *www.cognos.com*	• Information generation very limited and time-consuming • Only five people knew how to extract data using a 3GL • Response time unacceptable for Managers' decision-making purposes	• Provided quick answers to ad hoc questions for decision making • Provided access to data for decision-making purposes • Got in-depth view of product performance and customer margins
Pacific Gas Transmission Co. Natural gas provider in Pacific Northwest Source: Oracle Corp. *www.oracle.com*	• Rapid changes in markets due to deregulation • Diminishing profits in traditional services	• Managers able to analyze data quickly • Positioned company to quickly identify market trends • Created new services and pricing structures
Sega Corporation Interactive entertainment systems and video games Source: Oracle Corp. *www.oracle.com*	• Needed way to rapidly analyze great amount of data • Needed to track advertising, coupons, and rebates associated with effects of pricing changes • Used to do it with Excel spreadsheets, leading to human errors	• Eliminated data-entry errors • Identified successful marketing strategies to dominate interactive entertainment niches • Used product analysis to identify better markets/product offerings

---

Decision Support Systems

- Decision Support
  - Is a methodology
  - Extracts information by processing data
  - Uses information as basis for decision making
- Decision support system (DSS)
  - Arrangement of computerised tools
  - Used to assist managerial decision
  - Extensive data "massaging" to produce information
  - Used at all levels in organization
  - 'Drill Down' functionality
  - Interactive
  - Provides *ad hoc* query tools

DSS components

- Data Store
  - The large component that holds the organisational data
- Data Extraction and Filtering
  - Retrieval of information from the store and filtering it based on user supplied criteria
- Query Tool
  - Allows the users to generate Ad-hoc queries
- Presentation Tool
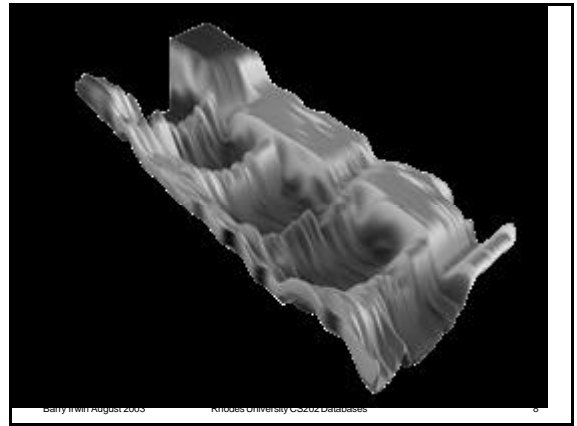 - Used to organise and present the data in a suitable format

DSS Components
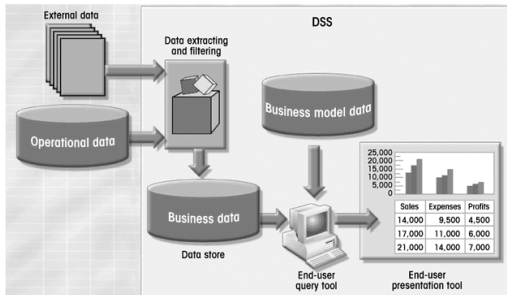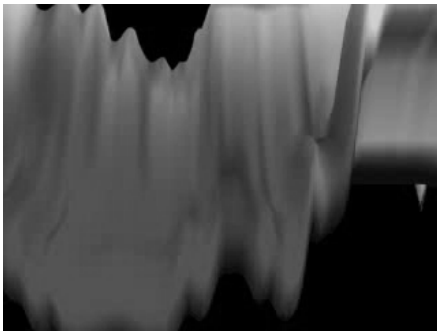
Operational Data vs DS Data

- Operational data
  - Relational, normalized database
  - Optimized to support transactions
  - Real time updates
- DSS
  - Snapshot of operational data
  - Summarized
  - Large amounts of data
- Data analyst viewpoint
  - Timespan
  - Granularity
  - Dimensionality

The DSS Database

- Database schema
  - Support complex (non-normalized) data
  - Extract multidimensional time slices
- Data extraction and filtering
- End-user analytical interface
- Database size
  - Very large databases (VLDBs)
  - Contains redundant and duplicated data
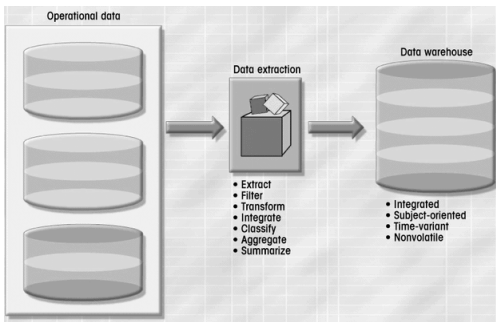
The Data Warehouse

- Integrated
  - Centralized
  - Holds data retrieved from entire organization
- Subject-Oriented
  - Optimized to give answers to diverse questions
  - Used by all functional areas
- Time Variant
  - Flow of data through time
  - Projected data
- Non-Volatile
  - Data never removed
  - Always growing

Data Warehouse

Operational data

Data extraction

Data warehouse

- Extract
- Filter
- Transform
- Integrate
- Classify
- Aggregate
- Summarize

- Integrated
- Subject-oriented
- Time-variant
- Nonvolatile

Data Marts

- Single-subject data warehouse subset
- Decision support to small group
- Can be test for exploring potential benefits of Data warehouses
- Address local or departmental problems
- Similar to a warehouse but smaller in scale

DSS Architectures

- Traditional mainframe-based OLTP
- Managerial information system (MIS) with 3GL
- First-generation departmental DSS
- First-generation enterprise data warehouse using RDMS
- Second-generation data warehouse using MDBMS
- See Table 13.7 page 627 for details

Twelve Data warehouse Rules

- Separated from operational environment
- Data are integrated
- Contains historical data over long time horizon
- Snapshot data captured at given time
- Subject-oriented data
- Mainly read-only data with periodic batch updates from operational source, no online updates
- Development life cycle differs from classical one, data driven not process driven


Twelve Data warehouse Rules

- Contains different levels of data detail
  - Current and old detail
  - Lightly and highly summarized
- Characterized by read-only transactions to large data sets
- Environment has system to trace data resources, transformation, and storage
- Metadata critical components
  - Identify and define data elements
  - Provide the source, transformation, integration, storage, usage, relationships, and history of data elements
- Contains charge-back mechanism for usage
  - Enforces optimal use of data

Data Warehouse Implementation Road Map
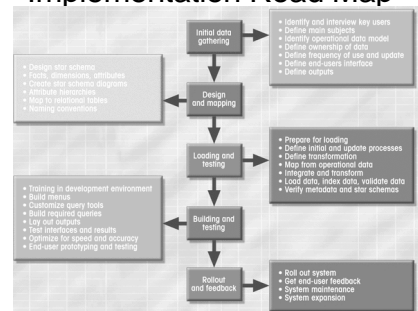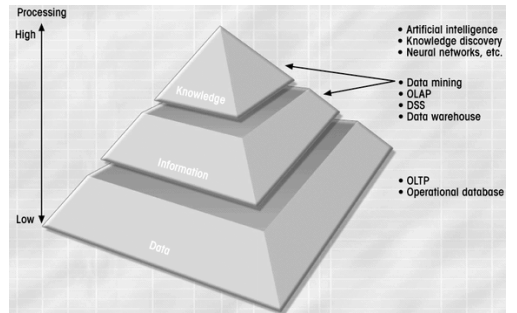
Data Mining

- Seeks to discover unknown data characteristics
- Automatically searches data for anomalies and relationships
- Data mining tools
  - Analyze data
  - Uncover problems or opportunities
  - Form computer models based on findings
  - Predict business behavior with models
  - Require minimal end-user intervention

Extraction of Knowledge from Data

Data Mining Process

Figure 13.23
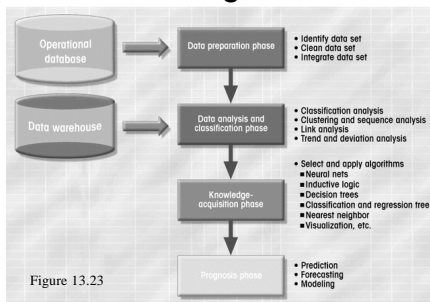
Star Schemas

- Used for mapping of Multi-Dimensional Data into a relational Database