Data Management

MBA 8120
Veda C. Storey

Learning Objectives

Technology
-- database management systems

Applications
-- target marketing
-- data mining

Management Issues
-- role in business practices
-- strategic applications
-- implications of outsourcing data management functions

Data Management

COLLECT
-- data sources
-- represent

STORE
-- databases
-- data warehouses

MANAGE
-- integration
-- enterprise applications
-- business process support

USE
-- data mining
-- target marketing
Data and Their Sources

- Types of data
  - public
  - private
- Sources
  - internal
  - external

Implications?

Databases for e-Commerce

"Without databases e-commerce is nothing."

-- Former dot com executive

e-Commerce architecture

[Diagram showing e-commerce architecture with layers and components]
A Vanilla Problem

- Who might have one?
- Why?

Who might have one?

- Nestle
  - Giant food conglomerate
  - Holding company (200 subsidiaries in 80 countries)

Why?

- Problem – pay same vendor 29 different prices for vanilla flavoring.
  - “Every plant would buy vanilla from the vendor, and the vendor would just get whatever it thought it could get. And the reason we couldn’t even check is because every division and every factory got to name vanilla whatever it wanted to. So you could call it 1234, and it might have a whole specification behind it, and I might call it 7778. we had no way of comparing.”
  - Jeri Dunn, CIO of Nestle USA

Global implementation

- $200 million contract for software
- $80 million for consulting and maintenance
- BEST (Business Excellence through Systems Technology)

Source: J. Senn, 2004; or R.A. Malaga, 2005; or Luftman, 2004

A Vanilla Problem (Cont’d)

Solution

- Common data structure across organization
- “1234” represents Vanilla every division / every country

Effects

- Combine purchase requests
- Manage inventory – move vanilla between locations
- $325 million saved
Databases

- **Database**
  - integrated collection of data

- **Database management software**
  - mechanism for storing and organizing data for sophisticated queries and manipulation of data

Employee and Client Data

<table>
<thead>
<tr>
<th>Employee</th>
<th>Client</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Id</td>
</tr>
<tr>
<td>Fred Smith</td>
<td>741</td>
</tr>
<tr>
<td>Sarah Thomas</td>
<td>852</td>
</tr>
<tr>
<td>Daniel McCarthy</td>
<td>963</td>
</tr>
<tr>
<td>Ellen Lewis</td>
<td>357</td>
</tr>
</tbody>
</table>

Client and Employee Data

Trend: Phenomenal Data Growth

- **Data Deluge**
  - Individual Data
    - Medical Records
    - Insurance
    - Educational
    - Personal Comm.
    - Credit History
    - Family & Kids
  - Mass Access Data
    - Movies, Videos
    - Books, Newspapers
    - Audio, Music
  - Corporate/Enterprise
    - Corporate Comm.
    - Accts/Legal/Admin
    - Inventory/Production
  - Customer/Sales

Source: Ashu Joshi
Data Explosion Problem

- Tremendous amount of data stored in databases
- Why?
  - Automated data collection techniques
  - Mature database technology

*We are drowning in data, but starving for knowledge!*
**What is a Data Warehouse?**

An integrated and consistent store of data, structured for query and retrieval to support management decision making.

- Data gathered from different operational databases
- Supports business analysis activities and decision-making tasks

---

**Data Warehouse**

- Where the information systems department puts data to be turned into information.
- One cannot just dump masses of data into a disk drive and expect it to be usable.

---

**Bigger and Bigger**

- Megabyte: 1 Million Bytes
- Gigabyte: 1 Billion Bytes
- Terabyte: 1000 Gigabytes
- Petabyte: 1000 Terabytes
- Exabyte: 1000 Petabytes
- ZettaByte: 1000 Exabytes
- Yottabyte: 1000 Zettabytes
**Data Warehouse Architecture**

Data Sources → Extract, Transform, Load → Warehouse → Analysis/Query/Reports/Data mining → Serve

- **Metadata**
- **Other Sources**
- **Operational DBs**

---

**Example: Wal-Mart**

- **Challenge:**
  - Approximately 100,000 items; 3,000 store locations
  - Merchandise decisions at individual store level
- **Solution:**
  - Data warehouse to store detailed sales data (e.g., by item, store, day)
  - Over 24 terabytes of data
  - 4,000 suppliers have access to warehouse
- **Buyers and vendors:**
  - Analyze sales trends
  - Analyze customer and seasonal buying trends
  - Make markdown decisions
  - React to merchandise volume and movement at any time
- **Feedback and strategic response:**
  - Connected to actual sales through main data center.
  - Sales in Florida after hurricane Charlie

---

**Wal-Mart**

Data warehouse: over 24 terabytes.

"Our business strategy depends on detailed data at every level - every cost, every line item is carefully analyzed enabling better merchandising decisions to be made on a daily basis. It is the foundation for maintaining Wal-Mart’s competitive edge and its continuing success in providing everyday low prices and superior customer satisfactions."

Randy Mott, Wal-Mart

---
Data Warehouses Concepts

- Information gathered from many databases
- Multidimensional
- Support OLAP, not OLTP
  - support decision making, not transaction processing

Multidimensional

OLTP versus OLAP

- Online transaction processing (OLTP)
  - gather data, process it, and update information
  - Operational databases — support OLTP

- Online analytical processing (OLAP)
  - manipulation of data to support decision making
  - Data Warehouse — special form of database that supports OLAP
Data Management

- COLLECT
- STORE
- MANAGE
- USE

Data Mining

Motivation

- Data in data warehouses not, by itself, of great intrinsic value.

- Value
  - Knowledge that can be discovered from data.
  - What do you do with it?
Data Mining

- Process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.
- Extraction of hidden, predictive information from large databases.
- Provide answers to questions a decision maker had previously not thought to ask.

Data Mining

- Search for relationships, patterns, and trends which, prior to the search were not known to exist or were not visible.

  E.g. “Find related buying patterns.”
  “There is a pattern that occurs X% of the time that when someone buys window coverings (not shades, blinds, or other specifics), and within 1 to 3 months buys linens, within the next 4 months buys furniture.”
Patterns

- Scuba gear and Australian vacations
- Skim milk and whole wheat bread
- AT&T’s stock rises at least 2% after every 3-day slump in DOW

Are “Discovered” Patterns Interesting?

- Data mining query
  - generate thousands of patterns
- Are they interesting?
- Interesting if:
  - easily understood
  - valid on new or test data with some degree of certainty
  - potentially useful
  - novel
  - validates hypothesis trying to confirm

Application: Telephone Company

How to find the customers you want to keep from among the millions?
Comb marketing data on 140 million households, each evaluated on as many as 10,000 attributes—e.g., income, lifestyle, and details about past calling habits.
Challenge: Data Mining Should Not be Used Blindly!

- Data mining find regularities from history
  - History not the same as the future
- Association does not dictate trend nor causality
  - Drink diet drinks lead to obesity!
  - David Heckerman’s counter-example
    - Barbecue source, hot dogs and hamburgers

Challenges

- Data quality
- Mining
  - Serious investment
- Products
  - Off-the-shelf, low-end mining products (careful)
  - Wide range of pricing
- Personnel
  - Do need team of statisticians to do mining right
  - Hire someone who knows what they are doing
- Gold
  - You "can" strike it!

Kimball’s User Categories

- Tourists
  - Impossible to see it all
  - Focus on the a few key cities and sites
- Farmers
  - They know their crops
  - They have seeds to plant
  - They know how to drive the tractor
- Explorers
  - Create or suggest hypotheses
  - Sometimes they find nothing, sometimes gold
- Miners
  - Help extract the nuggets
  - Statisticians

Source: B. Wixom
Target Marketing

- Effective use of customer information
- What are predictors of consumer purchase?
- Examples: Promotions, coupons, etc.

Conclusion

- **Database**
  - crucial part of business
- **Collect data**
  - internal and external sources
- **Store data**
  - Database management systems
  - Data warehouses
- **Use data**
  - Transaction processing
  - Marketing
- **Data mining**
  - powerful, strategic tool