Data Concepts
Database Management
Types of Databases

Harris
Fall 2006
Some Questions

- Why do we store so much data TODAY?
- What had we used to do?
- Value of DBMS?
DBMS vs. Spreadsheet Software

*Create customized analytical models
  – what if?, goal-seek, trends
*Create charts & graphs to show results
*Operate on modest data sets
  – Import “clean” data from files, DBMS
DBMS vs. Spreadsheet Software

* Create storage structures for data
* Provide end-user interfaces to data
  - query, reporting, & input tools
* Provide program-interface to data
* Manage enormous amounts of data
Case 1: Harrah’s Entertainment & Others

- For casinos, one of their most important assets is the data about the high-roller customers
- What steps can be taken to prevent employees from stealing this data?
  - Managerial
  - Legal
  - Technical
Case Study Questions

1. Why have developments in IT helped to increase the value of the data resources of many companies?

2. How have these capabilities increased the security challenges associated with protecting a company’s data resources?

3. How can companies use IT to meet the challenges of data resource security?
Real World Internet Activity

1. Companies are increasingly adopting a position that data is an asset that must be managed with the same level of attention as that of cash and other capital. Using the Internet,
   - See if you can find examples of how companies treat their data.
   - Does there seem to be any relationship between companies that look at their data as an asset and companies that are highly successful in their respective industries?
Real World Group Activity

- The case illustrates how valuable data resources are to the casino industry. In small groups,
  - Discuss other industries where their data are clearly their lifeblood.
  - For example, it has been estimated that any firm in the financial industry would have a life expectancy of less than 100 hours if they were placed in a position where they could not access their organizational data. Do you agree with this estimate?
Fundamental Data Concepts

- **Character**: single alphabetic, numeric or other symbol
- **Field or data item**: a grouping of related characters
  - Represents an attribute (a characteristic or quality) of some entity (object, person, place or event)
  - Example: salary
- **Record**: grouping of all the fields used to describe the attributes of an entity
  - Example: payroll record with name, SSN & rate of pay
- **File or table**: a group of related records
- **Database**: an integrated collection of logically related data elements
Example: An Electric Utility Database

Entities:
- Customers
- Meters
- Bills
- Payments
- Meter Readings

Relationships:
- Bills Sent to Customers
- Customers Make Payments
- Customers Use Meters

Billing

Payment Processing

Meter Reading

Service Start/Stop
Logical Database Structures

- Hierarchical
- Network
- Relational
- Multidimensional
- Object-oriented
Hierarchical Structure

Dept

- Project A
  - Employee 1
  - Employee 2

- Project B

Early DBMS structure
Records arranged in tree-like structure
Relationships are one-to-many
Network Structure

- Dept A
- Dept B
- Employee 1
- Employee 2
- Employee 3
- Project A
- Project B

Used in some mainframe DBMS packages
Many-to-many relationships
Relational Structure

- Most widely used structure
- Data elements are viewed as being stored in tables
- Row represents record
- Column represents field
- Can relate data in one file with data in another file if both files share a common data element
<table>
<thead>
<tr>
<th>Deptno</th>
<th>Dname</th>
<th>Dloc</th>
<th>Dmgr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dept B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dept C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Empno</th>
<th>Ename</th>
<th>Etitle</th>
<th>Esalary</th>
<th>Deptno</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emp 1</td>
<td></td>
<td></td>
<td></td>
<td>Dept A</td>
</tr>
<tr>
<td>Emp 2</td>
<td></td>
<td></td>
<td></td>
<td>Dept A</td>
</tr>
<tr>
<td>Emp 3</td>
<td></td>
<td></td>
<td></td>
<td>Dept B</td>
</tr>
<tr>
<td>Emp 4</td>
<td></td>
<td></td>
<td></td>
<td>Dept B</td>
</tr>
<tr>
<td>Emp 5</td>
<td></td>
<td></td>
<td></td>
<td>Dept C</td>
</tr>
<tr>
<td>Emp 6</td>
<td></td>
<td></td>
<td></td>
<td>Dept B</td>
</tr>
</tbody>
</table>
Relational Operations

- **Select:**
  - Create a subset of records that meet a stated criterion
  - Example, select employees who make more than $30,000

- **Join**
  - Combine two or more tables temporarily
  - Looks like one big table

- **Project**
  - Create a subset of columns in a table
Multidimensional Structure

- Variation of relational model
- Uses multidimensional structures to organize data
- Data elements are viewed as being in cubes
- Popular for analytical databases that support Online Analytical Processing (OLAP)
### The Multidimensional Structure

#### East
- **Sales**
  - Camera
  - TV
  - VCR
  - Audio

- **Margin**
  - Camera
  - TV
  - VCR
  - Audio

#### West
- **Sales**
  - Camera
  - TV
  - VCR
  - Audio

- **Margin**
  - Camera
  - TV
  - VCR
  - Audio

#### April
- **Sales**
  - East
  - West
  - South
  - Total

- **Margin**
  - East
  - West
  - South
  - Total

#### February
- **Sales**
  - January
  - February
  - March
  - Qtr 1

- **Margin**
  - January
  - February
  - March
  - Qtr 1

#### March
- **Sales**
  - East
  - Budget
  - Margin

- **Margin**
  - East
  - Budget
  - Margin

#### Qtr 1
- **Sales**
  - Actual
  - TV
  - VCR

- **Margin**
  - Actual
  - TV
  - VCR

- **Sales**
  - Total
  - Budget
  - Forecast
  - Variance

- **Margin**
  - Total
  - Budget
  - Forecast
  - Variance
Object-oriented Structure

- **Object** consists of
  - Data values describing the attributes of an entity
  - Operations that can be performed on the data
- **Encapsulation:**
  - Combine data and operations
- **Inheritance:**
  - New objects can be created by replicated some or all of the characteristics of parent objects
The Object-Oriented Structure

Bank Account Object

Attributes
- Customer
- Balance
- Interest

Operations
- Deposit (Amount)
- Withdraw (Amount)
- Get Owner

Checking Account Object

Attributes
- Credit Line
- Monthly Statements

Operations
- Calculate Interest
- Print Monthly Statement

Savings Account Object

Attributes
- Number of Withdrawals
- Quarterly Statement

Operations
- Calculate Interest Paid
- Print Quarterly Statement

Inheritance

Inheritance
Object-oriented Structure

- Used in Object-oriented database management systems (OODBMS)
- Supports complex data types
  - Examples, graphic images, video clips, web pages
Object-Oriented Graphical Display
Evaluation of Database Structures

- Hierarchical
  - Worked for structured routine transaction processing
  - Can’t handle many-to-many relationships
  - Unable to handle ad hoc requests

- Network
  - More flexible than hierarchical
  - Handles many-many relationships
  - Unable to handle ad hoc requests

- Relational
  - Easily respond to ad hoc requests
  - Easier to work with and maintain
  - Not as efficient or quick as hierarchical or network
Database Development

- **Database Administrator (DBA)**
  - In charge of enterprise database development

- **Data Definition Language (DDL)**
  - Develop and specify the data contents, relationships and structure
  - These specifications are stored in data dictionary

- **Data dictionary**
  - Data base catalog containing metadata
  - Metadata – data about data
1. Data Planning
   Develops a model of business processes

2. Requirements Specification
   Defines information needs of end users in a business process

3. Conceptual Design
   Expresses all information requirements in the form of a high-level model

4. Logical Design
   Translates the conceptual models into the data model of a DBMS

5. Physical Design
   Determines the data storage structures and access methods

- Physical Data Models
  Storage representations and access methods

- Logical Data Models
  E.g., relational, network, hierarchical, multidimensional, or object-oriented models

- Conceptual Data Models
  Often expressed as entity relationship models

Description of users' needs may be represented in natural language or using the tools of a particular design methodology.
Data Planning Process

- **Enterprise Model**
  - Defines basic business process of the enterprise
  - Defined by DBAs and designers with end users

- **Data Modeling**
  - Relationships between data elements
  - *Entity Relationship Diagram (ERD)* common tool for modeling
Database Design Process

- Logical design
  - Schema – overall logical view of relationships
  - Subschema – logical view for specific end users
  - Data models for DBMS

- Physical design
  - How data are to be stored and accessed on storage devices
Logical and Physical Database Views

Logical User Views
Data elements and relationships (the subschemas) needed for checking, savings, or installment loan processing

Data elements and relationships (the schema) needed for the support of all bank services

Software Interface
The DBMS provides access to the bank’s databases

Physical Data Views
Organization and location of data on the storage media
Case 2: Emerson and Sanofi
Data stewards seek data conformity

- Data stewards: dedicated to establishing and maintaining the quality of data
- Data quality team requires business, technology and diplomatic skills
- Focus on data content
Case Study Questions

1. Why is the role of a data steward considered to be innovative? Explain.

2. What are the business benefits associated with the data steward program at Emerson?

3. How does effective data resource management contribute to the strategic goals of an organization? Provide examples from Emerson and others.
1. The role of data steward is relatively new, and its creation is motivated by the desire to protect the valuable data assets of the firm. There are many job descriptions in the modern organization associated with the strategic management of data resources. Using the Internet,

- See if you can find evidence of other job roles that are focused on the management of an organization’s data.
- How might a person train for these new jobs?
Real World Group Activity

- As more and more data are collected, stored, processed, and disseminated by organizations, new and innovative ways to manage them must be developed. In small groups,
  - Discuss how the data resource management methods of today will need to evolve as more types of data emerge.
  - Will we ever get to the point where we can manage our data in a completely automated manner?
Section 2
Managing Data Resources

Harris
Fall 2006
Types of Databases

- Operational Databases
- Distributed Databases
- External Databases
- Hypermedia Databases
Distributed Databases

- Copies or parts of databases on servers at a variety of locations
- Challenge: any data change in one location must be made in all other locations
- Replication:
  - Look at each distributed database and find changes
  - Apply changes to each distributed database
  - Very complex
- Duplication
  - One database is master
  - Duplicate that database after hours in all locations
  - Easier
Operational Databases

Store detailed data to support business processes
Examples, customer database, inventory database
**Web-Based Information System**

Hypermedia Database: Website database
Consists of hyperlinked pages of multimedia
(text, graphics, video clips, audio segments)
Data Warehouse

- Stores data that has been extracted from the operational, external and other databases
- Data has been cleaned, transformed and cataloged
- Used by managers and professionals for
  - Data mining,
  - Online analytical processing,
  - Business analysis,
  - Market research,
  - Decision support
- Data mart is subset of warehouse for specific use of department
Data Warehouses: \textit{Components}

Operational, External, and Other Databases

Data Acquisition (Capture, clean, transform, transport, load/apply)

Warehouse Design

Data Management

Analytical Data Store
Enterprise Warehouse
Data Marts

Metadata Management

Metadata Directory
Metadata Repository

Data Analysis (Query, report, analyze, mine, deliver)

Web Information System
Meta data is data about data.
Data Mining

Data in data warehouse are analyzed to reveal hidden patterns and trends

Examples:

- Perform market-basket analysis to identify new business processes
- Find root causes to quality problems
- Cross sell to existing customers
- Profile customers with more accuracy
Data Mining: Examples

- Identify new product bundles
- Find root causes to quality or manufacturing problems
- Prevent customer attrition
- Cross-sell to existing customers
- Profile customers with more accuracy
Using Data Mining to Develop Business Knowledge

Database

Selection → Target Data

Data Transformation → Target Warehouse

Data Mining → Patterns

Implementation/Evaluation → Business Knowledge
Traditional File Processing
Example – Bank File Processing Systems
Traditional File Processing

- Data stored in independent files
- Problems:
  - Data redundancy
  - Lack of data integration
  - Data dependence – files, storage devices, and software are dependent on each other
  - Lack of data integrity or standardization
Database Management Approach
Database Management Approach

- Consolidate data into databases that can be accessed by different programs
- Use a database management system (DBMS)
- DBMS serves as interface between users and databases
DBMS Major Functions

• Create: Database and Application Development
• Maintain: Database Maintenance
• Use: Database Interrogation
Database Interrogation

- End users use a DBMS by asking for information via a query or a report generator
- **Query language** – immediate responses to ad hoc data requests
  - **SQL** *(Structured Query Language)* an international standard query language
  - **Graphical Queries** – Point-and-click methods
  - **Natural Queries** – similar to conversational English
- **Report generator** – quickly specify a report format for information you want printed in a report
A Sample Natural Language-to-SQL Translation for Microsoft Access

Natural Language

WHAT CUSTOMERS HAD NO ORDERS LAST MONTH?

SQL

SELECT [Customers].[Company Name],[Customers].[Contact Name]
FROM [Customers]
WHERE not Exists (SELECT [Ship Name] FROM [Orders]
    WHERE Month {{Order Date}}=1 and Year {{Order Date}}=2004 and
    [Customers].[Customer ID]=[Orders].{{Customer ID}})
Graphical Query

Source: Courtesy of Microsoft Corp.
Case 3: Acxiom Corporation
Data Demands Respect

- Acxiom does three things:
  - Managing large volumes of data
  - Cleaning, transforming & enhancing that data
  - Distilling business intelligence from that data to drive smart decisions
- Provides information products
- Manages clients’ data
Case Study Questions

1. Acxiom is in a unique type of business. How would you describe the business of Acxiom? Are they a service- or a product-oriented business?

2. From the case, it is easy to see that Acxiom has focused on a wide variety of data from different sources. How does Acxiom decide which data to collect and for whom?

3. Acxiom’s business raises many issues related to privacy. Is the data collected by Acxiom really private?
Real World Internet Activity

1. In the case, it was stated that Acxiom started as the result of a spin-off from a bus company. Using the Internet,
   - See if you can find the history of Acxiom.
   - How does a bus company evolve into a data collection and dissemination company?
Real World Group Activity

The privacy problems faced by Acxiom were associated with the accidental dissemination of data deemed sensitive by a third party. In small groups,

- Discuss the privacy issues associated with Acxiom’s business.
- Do you think they are doing anything wrong?
The End

Harris
Fall 2006