Management Information Systems

- Data and Knowledge Management

Dr. Shankar Sundaresan

(Adapted from Introduction to IS, Rainer and Turban)

LEARNING OBJECTIVES

- Recognize the importance of data, issues involved in managing data and their lifecycle.
- Describe the sources of data and explain how data are collected.
- Explain the advantages of the database approach.
- Work with Relational Databases
- Explain the operation of data warehousing and its role in decision support.
- Define knowledge, and describe different types of knowledge.

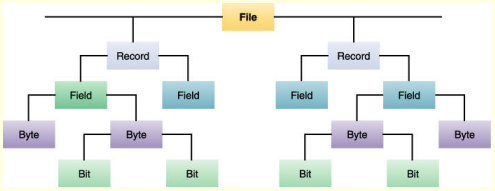
Managing Data

Difficulties in Managing Data

Data Hierarchy

- Bit
- Byte
- Field
  - is a group of logically related characters (e.g., a word, small group of words, or identification number).
- Record
  - is a group of logically related fields (e.g., student in a university database)
- File (or table)
  - is a group of logically related records.
- Database
  - is a group of logically related files.

Hierarchy of Data for a Computer-Based File



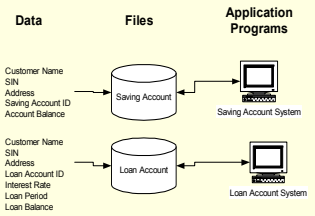
Traditional File based Approach

- Data Storage
  - Bits, bytes, fields, records, files
- File Organizations
  - Sequential Access (SAM)
    - Data records retrieved in the same order of physical storage
  - Indexed Sequential Access (ISAM)
    - Uses index of key fields to locate individual records
    - Separate index file lists the key values and data storage location
  - Random Access
    - Mathematical algorithm indicates storage location based on the key field

17

Traditional File-Based Approach

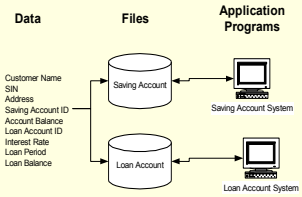
Drawbacks
- Application/Data Independence:
  Any change in the data will need a change in the programs
- Redundancy
- Inconsistency
- Integrity
- Isolation



Database Approach

A database management system (DBMS) is software that serves as an interface between a common database and various application programs.

Centralized, Shared Access
Consistent Data



A database is a collection of data accessed and utilized by many different applications.

Database Approach

- Make data a centralized resource, used by multiple applications
  - One set of basic access programs
  - Manage access to data
  - Provide metadata and query language to ease access
  - Design data structure once, reuse it
- Controlled (Vs. no) redundancy
- Easy to establish relationships among data--natural correspondence between real world and data world

20

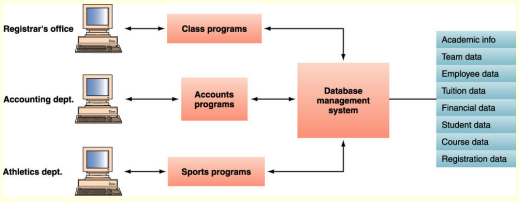
The Database Approach

- Database management system (DBMS) provides all users with access to all the data.
- DBMSs minimize the following problems:
  - Data redundancy
    - The same data are stored in many places.
  - Data isolation
    - Applications cannot access data associated with other applications.
  - Data inconsistency
    - Various copies of the data do not agree.

Database Approach (continued)

- DBMSs maximize the following issues:
  - Data security
    - Keeping the organization's data safe from theft, modification, and/or destruction.
  - Data integrity
    - Data must meet constraints (e.g., student grade point averages cannot be negative).
  - Data independence
    - Applications and data are independent of one another. Applications and data are not linked to each other, meaning that applications are able to access the same data.

Database Management Systems



Database Management Systems

- Database management system (DBMS)
  - is a set of programs that provide users with tools to add, delete, access, and analyze data stored in one location.
- Relational database model
  - is based on the concept of two-dimensional tables.
- Structured Query Language (SQL)
  - allows users to perform complicated searches by using relatively simple statements or keywords.+
- Query by Example (QBE)
  - allows users to fill out a grid or template to construct a sample or description of the data he or she wants.

Designing the Database

Data model

is a diagram that represents the entities in the database and their relationships.

- Entity
  - is a person, place, thing, or event about which information is maintained. A record generally describes an entity.
- Attribute
  - is a particular characteristic or quality of a particular entity.
- Primary key
  - **ke**y is a field that uniquely identifies a record.
- Secondary keys
  - are other field that have some identifying information but typically do not identify the record with complete uniqueness.

Relational model--introduction

- Basic structure: the "relation"
- Informally, a relation is a two-dimensional table.
- Rows correspond to individual "things" or individual relationships between things
- Columns correspond to attributes which take on **atomic** values
- Any table can be related to a different table through common data elements
- It's WYSIWYG: all data access paths are shown in the collection of relations making up the database
- Some combination of the attributes serves as a **unique identifier (primary key)** for the rows
  - No duplicate rows
- Rows unordered: no top-to-bottom order implied
- Columns unordered: no left-to-right order implied
- Example:
  - Microsoft Access, ORACLE

26

Example

EMPID	NAME	DEPT	SALARY	COURSE	COMP DATE	DEPT MGR	COURSE DURATION
100	Margaret Simpson	Mktg	42,000	SPSS	6/19/93	900	2 Days
100	Margaret Simpson	Mktg	42,000	Surveys	10/7/93	900	3 Days
140	Alan Beton	Acctg	39,000	Tax Acc	12/8/94	910	2 Days
110	Chris Lucero	IS	41,500	C++	4/22/94	920	5 Days
110	Chris Lucero	IS	41,500	SPSS	5/7/95	920	2 Days
190	Lorenzo Davis	Finance	38,000	Investing	1/12/95	930	1 Day
150	Susan Martin	Mktg	38,500	SPSS	6/19/95	900	2 Days
150	Susan Martin	Mktg	38,500	TQM	8/12/95	900	1 Day

27

Keys

- Primary Key: set of attributes in a relation which have been selected as record identifiers.
  - Must be uniquely valued
  - No part may be "null"
  - No subset of the attributes may be unique
- How to select a primary key?
- Foreign Key:
  - Attribute in one relation which is the primary key of some other relation.
  - Foreign keys can be null
  - Provide the linkage between relations

28

Example

EMPID	NAME	DEPT	SALARY	COURSE	COMP DATE	DEPT MGR	COURSE DURATION
100	Margaret Simpson	Mktg	42,000	SPSS	6/19/93	900	2 Days
100	Margaret Simpson	Mktg	42,000	Surveys	10/7/93	900	3 Days
140	Alan Beton	Acctg	39,000	Tax Acc	12/8/94	910	2 Days
110	Chris Lucero	IS	41,500	C++	4/22/94	920	5 Days
110	Chris Lucero	IS	41,500	SPSS	5/7/95	920	2 Days
190	Lorenzo Davis	Finance	38,000	Investing	1/12/95	930	1 Day
150	Susan Martin	Mktg	38,500	SPSS	6/19/95	900	2 Days
150	Susan Martin	Mktg	38,500	TQM	8/12/95	900	1 Day

29

Anomalies from Poor Design

- Insertion Anomaly
  - inability to insert a piece of information about an object without having to add spurious information about something else
  - e.g. Add a new employee
- Deletion Anomaly
  - The loss of a piece of information about one object when a piece of information about a different object is deleted
  - E.g. Delete course information on Tax Acc
- Modification Anomaly
  - A need to change the same piece of information multiple times
  - E.g. Margaret gets salary increase
- Improving database Design
  - Normalization

30

Modified Design

COURSE	COURSE DURATION
SPSS	2 Days
Surveys	3 Days
Tax Acc	2 Days
C++	5 Days
Investing	1 Day
TQM	1 Day

EMPID	NAME	DEPT	SALARY	DEPT MGR
100	Margaret Simpson	Mktg	42,000	900
140	Alan Beton	Acctg	39,000	910
110	Chris Lucero	IS	41,500	920
190	Lorenzo Davis	Finance	38,000	930
150	Susan Martin	Mktg	38,500	900

EMPID	COURSE	COMP DATE
100	SPSS	6/19/93
100	Surveys	10/7/93
140	Tax Acc	12/8/94
110	C++	4/22/94
110	SPSS	5/7/95
190	Investing	1/12/95
150	SPSS	6/19/95
150	TQM	8/12/95

31

Further Simplification

DEPT	DEPT MGR
Mktg	900
Acctg	910
IS	920
Finance	930

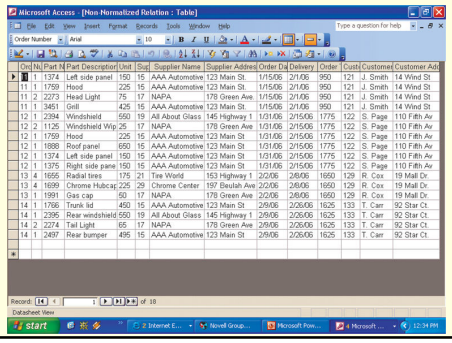
EMPID	NAME	DEPT	SALARY
100	Margaret Simpson	Mktg	42,000
140	Alan Beton	Acctg	39,000
110	Chris Lucero	IS	41,500
190	Lorenzo Davis	Finance	38,000
150	Susan Martin	Mktg	38,500

32

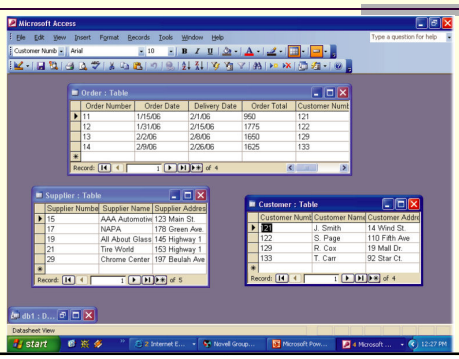
Normalization

- **Normalization** is a method for analyzing and reducing a relational database to its most streamlined form for:
  - Minimum redundancy
  - Maximum data integrity
  - Best processing performance
- Normalized data is when attributes in the table depend only on the primary key.

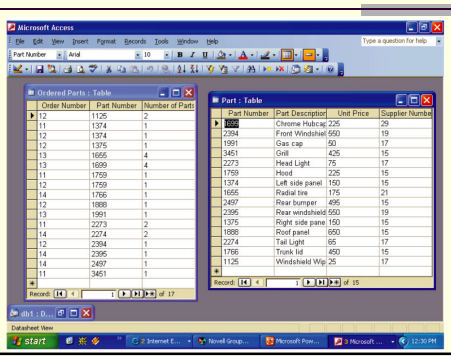
Non-Normalized Relation



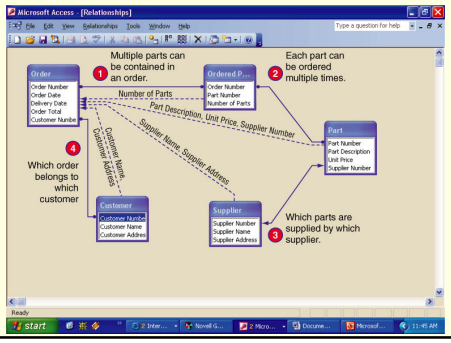
Normalizing the Database (part A)



Normalizing the Database (part B)



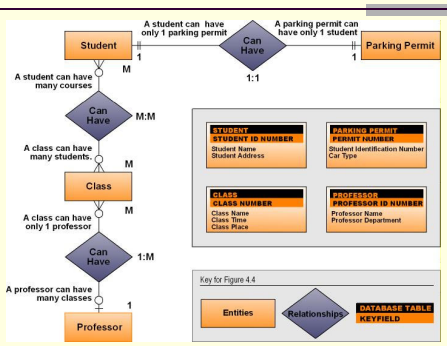
Normalization Produces Order



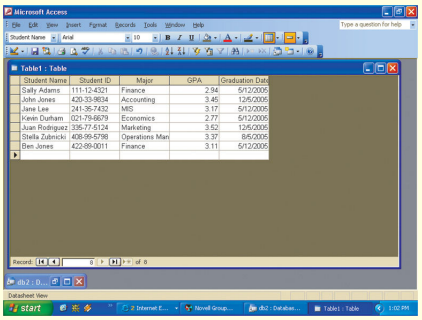
Entity-Relationship Modeling

- Database designers plan the database design in a process called **entity-relationship (ER) modeling.**
- **ER diagrams** consists of entities, attributes and relationships.
  - Entity classes
    - are groups of entities of a certain type.
  - Instance
    - the representation of a particular entity.
  - Identifiers
    - attributes that are unique to that entity instance.

Entity-Relationship Diagram Model



Student Database Example

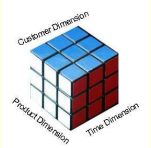


Data Warehousing

Data warehouse

- is a repository of historical data organized by subject to support decision makers in the organization.
- Data warehouses are organized by business dimension or subject.
- Data warehouses are multidimensional.

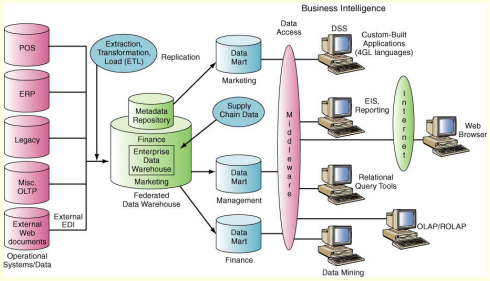
A Data Cube



Data Warehousing (continued)

- Data warehouses are historical.
  - data warehouses can be used for identifying trends, forecasting, and making comparisons over time.
- Data warehouses use online analytical processing (OLAP)
  - involves the analysis of accumulated data by end users

Data Warehouse Framework & Views



Relational Databases

(a) 2003

Product	Region	Sales
Nuts	East	50
Nuts	West	60
Nuts	Central	100
Screws	East	40
Screws	West	70
Screws	Central	80
Bolts	East	90
Bolts	West	120
Bolts	Central	140
Washers	East	20
Washers	West	10
Washers	Central	30

(b) 2004

Product	Region	Sales
Nuts	East	60
Nuts	West	70
Nuts	Central	110
Screws	East	50
Screws	West	80
Screws	Central	90
Bolts	East	100
Bolts	West	130
Bolts	Central	150
Washers	East	30
Washers	West	20
Washers	Central	40

(c) 2005

Product	Region	Sales
Nuts	East	70
Nuts	West	80
Nuts	Central	120
Screws	East	60
Screws	West	90
Screws	Central	100
Bolts	East	110
Bolts	West	140
Bolts	Central	160
Washers	East	40
Washers	West	30
Washers	Central	50

Benefits of Data Warehousing

- End users can access data quickly and easily via Web browsers because they are located in one place.
- End users can conduct extensive analysis with data in ways that may not have been possible before.
- End users have a consolidated view of organizational data.

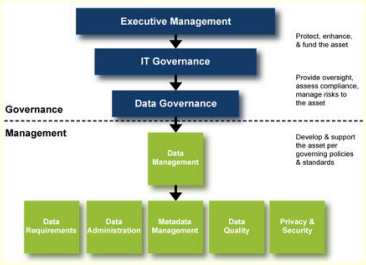
Data Marts

A **data mart** is a small data warehouse, designed for the end-user needs in a strategic business unit (SBU) or a department.

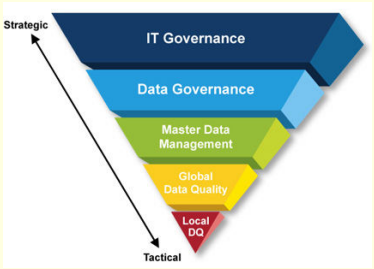
4.5 Data Governance

- Data governance
  - is an approach to managing data and information across an entire organization.
- Master data management
  - is a method that organizations use in data governance.
- Master data
  - are a set of core data that span all enterprise information systems.

Data Governance (continued)



Data Governance (continued)



Knowledge Management

- Knowledge management (KM)
  - is a process that helps organizations manipulate important knowledge that is part of the organization's memory, usually in an unstructured format.
- Knowledge
  - that is contextual, relevant, and actionable.
- Intellectual capital (or intellectual assets)

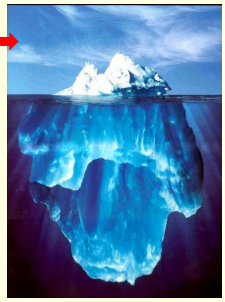
Knowledge Management (continued)

Explicit Knowledge
(above the waterline)

objective, rational, technical knowledge that has been documented.
Examples: policies, procedural guides, reports, products, strategies, goals, core competencies

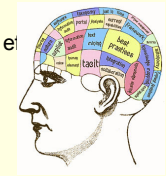
Tacit Knowledge
(below the waterline)

cumulative store of subjective or experiential learning.
Examples: experiences, insights, expertise, know-how, trade secrets, understanding,
skill sets, and learning



Knowledge Management (continued)

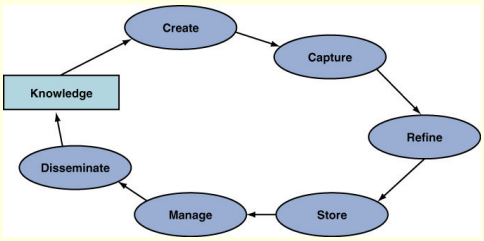
- Knowledge management systems (KMSs)
  - systems refer to the use of information technologies to systematize, enhance, and expedite intrafirm and interfirm knowledge management.
- Best practices
  - are the most effective and e
    doing things.



Knowledge Management System Cycle

- Create knowledge
- Capture knowledge
- Refine knowledge
- Store knowledge
- Manage knowledge
- Disseminate knowledge

Knowledge Management System Cycle



Stocks Database