INFO20003
DATABASE SYSTEMS

Complete Subject Notes

Semester 2 2016

Contents

Table of Contents

Data, Information & Knowledge

Data vs Information

Data

- Just fact stored and recorded
    - Numbers
    - Dates
    - Text
    - Images
    - Sound
    - Video

Information

- Data in context
    - Can be summarised
- Data that has been processed to increase its value, and user's knowledge about that data
- Makes data more useful
- Examples:
    - Put raw data in tables
    - Put raw data in charts

Metadata

- Data about data
    - i.e. Definitions of each column in a table
- Can include:
    - Structure
    - Rules
    - Constraints
- Why is metadata needed?
    - Consistency
    - Clarity
    - Meaning
- Metadata standards
    - Use XML, HTML, etc.
- **Data dictionaries** are generated as part of analysis of system requirements
    - Contains metadata

What is a Database?

- Collection of logically related data and its description
    - Usually in tables, spreadsheets

Database Management System (DBMS)

- A software system that enables users to define and control the database
    - Need to define, create, maintain and control database access
- Provides specific capabilities:

- **Query Language**
  - Data definition language (DDL): define and set up database
  - Data manipulation language (DML): maintain, alter and use database
  - Data control language (DCL): control access to the database
- **Other capabilities**
  - Integrity of data
  - Concurrency (shared access)
  - Recovery
  - Data descriptions

File Processing Systems

- Old method:
  - Multiple computers for each department, each contain flat file data of products and/or customer information
  - Problems:
    - Program-data dependence
      - If file structure changes, so does the program
      - Program knows too much low-level data structure
    - Duplication of data
      - Each department independently has to keep track of customers and products
      - Wasteful, inefficient, loss of integrity
      - Loss of metadata integrity
    - Limited data sharing
      - Data is tied to application in each department
    - Lengthy development times
      - Application has to do low-level data management, figuring out the file format each time
    - Excessive program maintenance
      - Up to 80% of development time is used for maintenance
- New method: Database systems
  - Advantages:
    - Data independence
      - Separation of data and program
      - Central data repository
    - Minimal data redundancy
      - Redundancy can be control using normalisation
    - Improve data consistency
      - Single store of data means to disagreement or update problems
    - Improve data sharing
      - Data is shared, not necessary for a single application
      - External users can be allowed access
    - Reduce data maintenance
      - Data structure can change without application changing
    - Increased productivity of application development
      - Data already collected and structures already known
      - DMBS have many tools to help access and maintain the data
    - Enforcement of standards
      - Centralised data management
      - Documented policy for data management
      - Data definition and dictionary for metadata

- Improved data quality
    - Constraints built into the database
    - Scrubbing data if necessary
- Data access “without programming”
    - SQL