Introduction to Database Concepts

© Department of Computer Science
Northern Illinois University
January 2018
What is a Database?

- A collection of stored operational data used by the application systems of some particular enterprise.

- As per book
  “A collection of related data”
What is a Database?

• **Enterprise**

  - generic term for any reasonably large-scale commercial, scientific, technical, or other application
    
    • manufacturing
    • financial
    • medical
    • university
    • government
What is a Database?

• Operational data
  – data maintained about the operation of an enterprise
    • products
    • accounts
    • patients
    • students
    • plans

• DOES NOT include input/output data
What is a DBMS?

• A Database Management System (DBMS)
  – a collection of programs that enables users to create and maintain a database
  – a general-purpose software system that facilitates the processes for databases of
    • defining
    • constructing
    • Manipulating
    • sharing
Defining a Database

• For the data being stored in the database, defining the database specifies:
  – the data types
  – the structures
  – the constraints
Constructing a Database

• Is the process of storing the data itself on some storage device
• The storage device is controlled by a DBMS
Manipulating a Database

- Included functions that
  - retrieve specific information in a query
  - update the database to include changes
  - generate reports from the data
Sharing a Database

• Allows multiple users and programs to access the database at the same time
• Any conflicts between applications are handled by the DBMS
Other Important Functions of a Database

- Other important functions provided by a DBMS include
  - Protection
    - System protection
    - Security protection
  - Maintenance
    - Allows updates easily
Simplified Database System Environment

- Users/Programmers
- Application Programs / Queries
- Software to Access Stored Data
- Software to Process Queries / Programs
- DBMS SOFTWARE
- Stored Database
- Stored Database Definition (Meta-Data)
• Main characteristics of a database system are
  – Self-describing nature of a database system
  – Insulation between programs and data, and data abstraction
  – Support of multiple view of the data
  – Sharing of data and multi-user transaction processing
Other Capabilities of DBMS Systems

• Support for at least one data model through which the user can view the data
  – there is at least one abstract model of data that allows the user to see the “information” in the database
  – relational, hierarchical, network, inverted list, or object-oriented
Other Capabilities of DBMS Systems

- Support for at least one data model through which the user can view the data
  - efficient file access which allows us to “Find the boss of Susie Jones”
  - allows us to “navigate” within the data
  - allows us to combine values in 2 or more databases to obtain “information”
Other Capabilities of DBMS Systems

• Support for certain high-level languages that allow the user to define the structure of data, access the data, and manipulate the data
  – Data Definition Language (DDL)
  – Data Manipulation Language (DML)
  – Data Control Language (DCL)
  – query languages access data
  – operations such as add, delete, and replace
Other Capabilities of DBMS Systems

• **Transaction management** which is the capability to provide correct, concurrent access to the database by many users at the same time
  – ability to manage simultaneously large numbers of “transactions”
    • procedures operating on the database
    • often transactions come from around the world
    • “lock-out” mechanisms
Other Capabilities of DBMS Systems

• **Access control** which is the ability to limit access to data by unauthorized users along with the capability to check the validity of the data
  – protect against loss when database crashes
  – prevent unauthorized access to portions of data
Other Capabilities of DBMS Systems

• **Resiliency** which is the ability to recover from system failures without losing data
  – must be able to recover from ANY type of failure
    • sabotage
    • acts of God
    • hardware failure
    • software failure
    • etc.
Use of Conceptual Modeling

Entity-Relationship Diagram

Conceptual Level

Conversion of Conceptual Schema to Logical Schema

Logical Database
(relational, hierarchical, network, etc.)

Logical Level
Leveled Architecture of a DBMS

View 1

View 2

View n

External Level (View Level)

External / Logical Mapping

Logical Level
(relational, hierarchical, network, etc.)

Logical Data Independence

Logical Level

Logical / Physical Mapping

Conceptual Model

Physical Level (as stored)

Physical Data Independence
Levels of a DBMS

- **External Level**
  - a view or sub-schema
    - portion of the logical database
  - may be in a higher level language
Levels of a DBMS

• Logical Level
  – abstraction of the real world as it pertains to the users of the database
  – DBMS provides a DDL to describe the logical schema in terms of a specific data model such as relational, hierarchical, network, or inverted list
Levels of a DBMS

• Physical Level
  – the collection of files and indices
  – resides permanently on secondary storage device
  – actual data itself
Instance

- an instance of a database is the current contents of the data
- extension of the database
- state of the database
- snapshot of the data
Basic Database Terminology

• Schema
  – plan of the database
  – logical plan
  – physical plan
  – intention of the database
Basic Database Terminology

Schema

Instance

Student Name
Susie Jones

Birthdate
09/03/1980

GPA
3.459
Basic Database Terminology

- **Data Independence**
  - has to do with the separation of data modeling levels
Basic Database Terminology

• Physical Data Independence
  – The separation between the physical and logical data modeling levels.

  – Consequence -
    • physical schema can be changed without modifying logical schema
Basic Database Terminology

• **Logical Data Independence**
  - The separation between the logical and conceptual (external) data modeling levels.

  - **Consequence** -
    • logical schema can be changed without modifying conceptual schema
Basic Database Terminology

• DCL, DDL and DML
  – may be completely separate (in IMS)
  – may be intermixed (DB2)
  – Host language
    • application program in which DML commands are embedded such as COBOL or PL/I
DBMS Components

User’s Query

Application / Load Program

Database Schema

Security Tables

Query Language Processor

DDL Compiler

Transaction Management Tables

Database Description Tables

File Manager

Physical Database
Overall Scenario of a using a DBMS

• Database Administrator (DBA) defines the conceptual, logical, and physical levels using DDL
• DBMS software stores instances of these in schemas
• User defines views (External Schema) in DDL
• User accesses database using DML
Advantages of a Database

- Controlled redundancy
- Reduced inconsistency in the data
- Shared access to data
- Standards enforced
- Security restrictions maintained
- Integrity maintained easier
- Providing Backup and recovery
- Permitting inferencing and actions using rules
Disadvantages of a Database

- Increased complexity in concurrency control
- Increased complexity in centralized access control
- Security needed to allow the sharing of data
- Needed redundancies cause updating complexities
Data and Information

- Data – values
- Information – computed values
- Knowledge – data/information with meaning
- Wisdom – application of knowledge