BASIC DATABASES

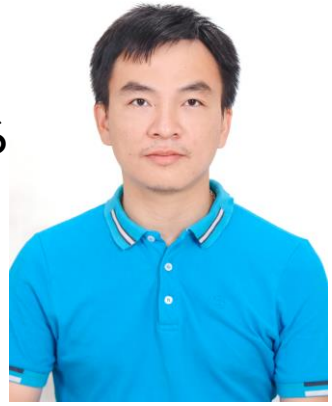
Introduction to Databases

NGUYEN Hoang Ha

Email: nguyen-hoang.ha@usth.edu.vn

About lecturer

- Full name: Hoang Ha NGUYEN
- Email: nguyen-hoang.ha@usth.edu.vn
- Diploma: PhD in Computer Science at Aix-Marseille University 2016
- Position: Lecturer at University of Science and Technology of Hanoi
- Research interests:
  - Computer graphics:
    - 3D reconstruction: building 3D models of objects from point cloud
    - Object modelling, synthesizing realistic images of objects.
    - Mesh optimization
    - Augmented Reality
  - Computer Vision
    - Object recognition and classification: landmarks on insect wings, hand gestures
    - 3D reconstruction from multiple views



Course information

- Credit: 4
- Moodle page:
  - Materials
  - Assignment submissions
- Prescribed book: Jeffrey D. Ullman, Jennifer Widom: A First Course in Database Systems, Pearson, 3rd Edition (2007)
- Softwares: My SQL
- Assessment
  - Attendance: 10%
  - Middle term: 40%
    - Rewards (+2, +1), Penalties (-2, - 1)
  - Final project: 50%

Objectives

- Understand concepts of
  - Information,
  - Data,
  - Database,
  - DBMS
  - DBS

References

- Jeffrey D.Ullman and Jennifer Widom, *A First Course In Database System Chapter 1;* Prentice-Hall Internaltional
- *Fundamentals of Database System Chapter 1,2;* Third Edition, Addison-Wesley

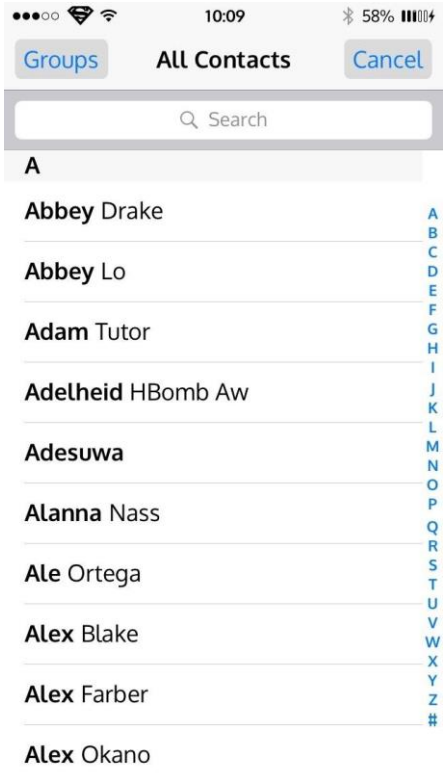
Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantage and disadvantage

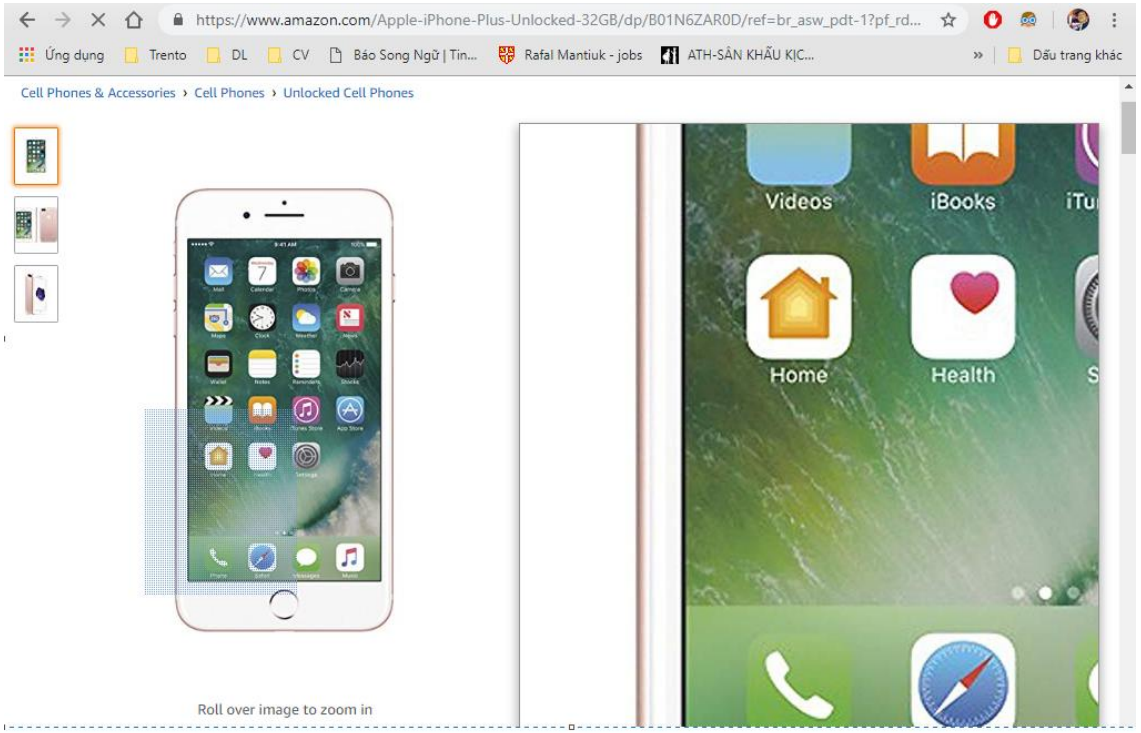
Why study Dabases

- Shift from *computation* to *information*
  - at the "low end": scramble to webspace
  - at the "high end": scientific applications
- DBMS relates to most of CS
  - Information system, OS, languages, "AI", multimedia
- Datasets increasing in diversity and volume.
  - Airline Reservation, Banking, Medicine, Corporate
  - Digital libraries, interactive video, Human Genome project, EOS project
  - ...

Database application examples



Contact List



Items description on Amazon

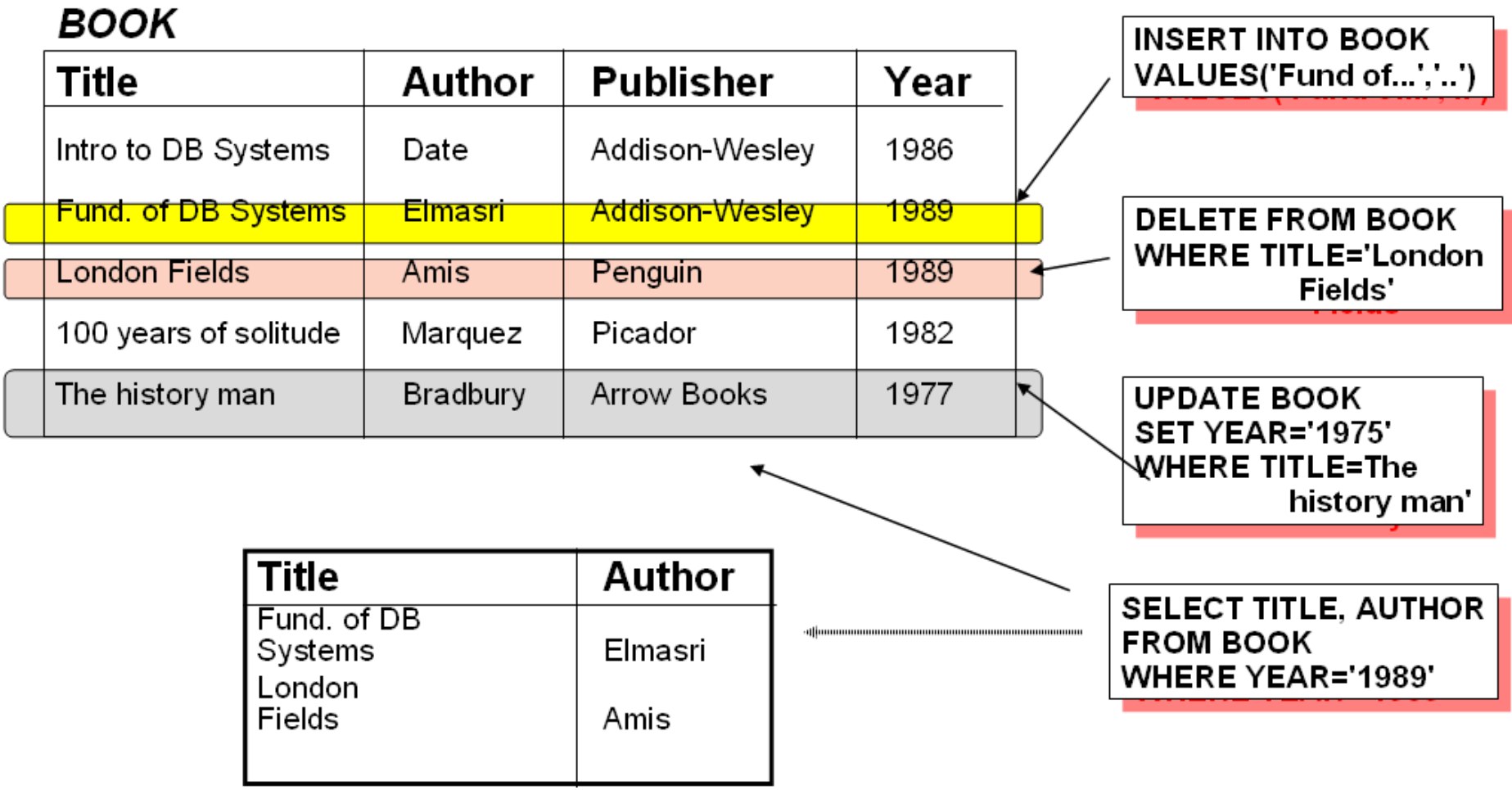
Databases are everywhere

- Applications:
  - Banking: transactions
  - Airlines: reservations, schedules
  - Universities: registration, grades
  - Sales: customers, products, purchases
  - Online retailers: order tracking, customized recommendations
  - Manufacturing: production, inventory, orders, supply chain
  - Human resources: employee records, salaries, tax deductions

Basic Definitions

- **Data**:
  - Known facts that can be recorded and have an implicit meaning.
  - Anything in a form suitable for use with a computer. Distinguished from program (Wikipedia)
- **Database**:
  - Nothing more than a collection of data existing over a long period of time
  - Purposes
    - To store data
    - To provide an organizational structure for data
    - To provide a mechanism for creating, modifying, deleting, and querying data
- **Database Management System (DBMS)**
  - A software package/ system to facilitate the creation and maintenance of a computerized database.
- **Database System**
  - The DBMS software together with the data itself. Sometimes, the applications are also included.

A Sample Database



Main Characteristics of the Database Approach

- Self-describing nature of a database system: A DBMS **catalog** stores the *description* of the database. The description is called **meta-data**).
- Isolation between programs and data: **program-data independence**. Allows changing data storage structures and operations without having to change the DBMS access programs.

Main Characteristics of the Database Approach

- Data Abstraction: A data model is used to hide storage details and present the users with a *conceptual view* of the database.
- Support of multiple views of the data: Each user may see a different view of the database, which describes *only* the data of interest to that user.

Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantage and disadvantage

History of DB Technology

The first DBMS evolved from file systems

Late 60s:

- 1969: Charles Bachman network data model
- IBM IMS hierarchical data model

- 70s:
  - Edgar Codd relational model
  - SQL was developed by IBM
  - 1979: Oracle Version 2, the first commercial RDBMS product using SQL

History of DB Technology (cont)

- 80s: SQL IBM R was introduced in 1981 (based on Codd's research)
- Late 80s-90s:
  - DB2, Oracle, Informaix, Sybase
  - OODBMSs were introduced
- 90s:
  - SQL was standardized by ANSI in 1992
- From 2000:
  - XML
  - db40
  - NoSQL: MongoDB (2007)

Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantage and disadvantage

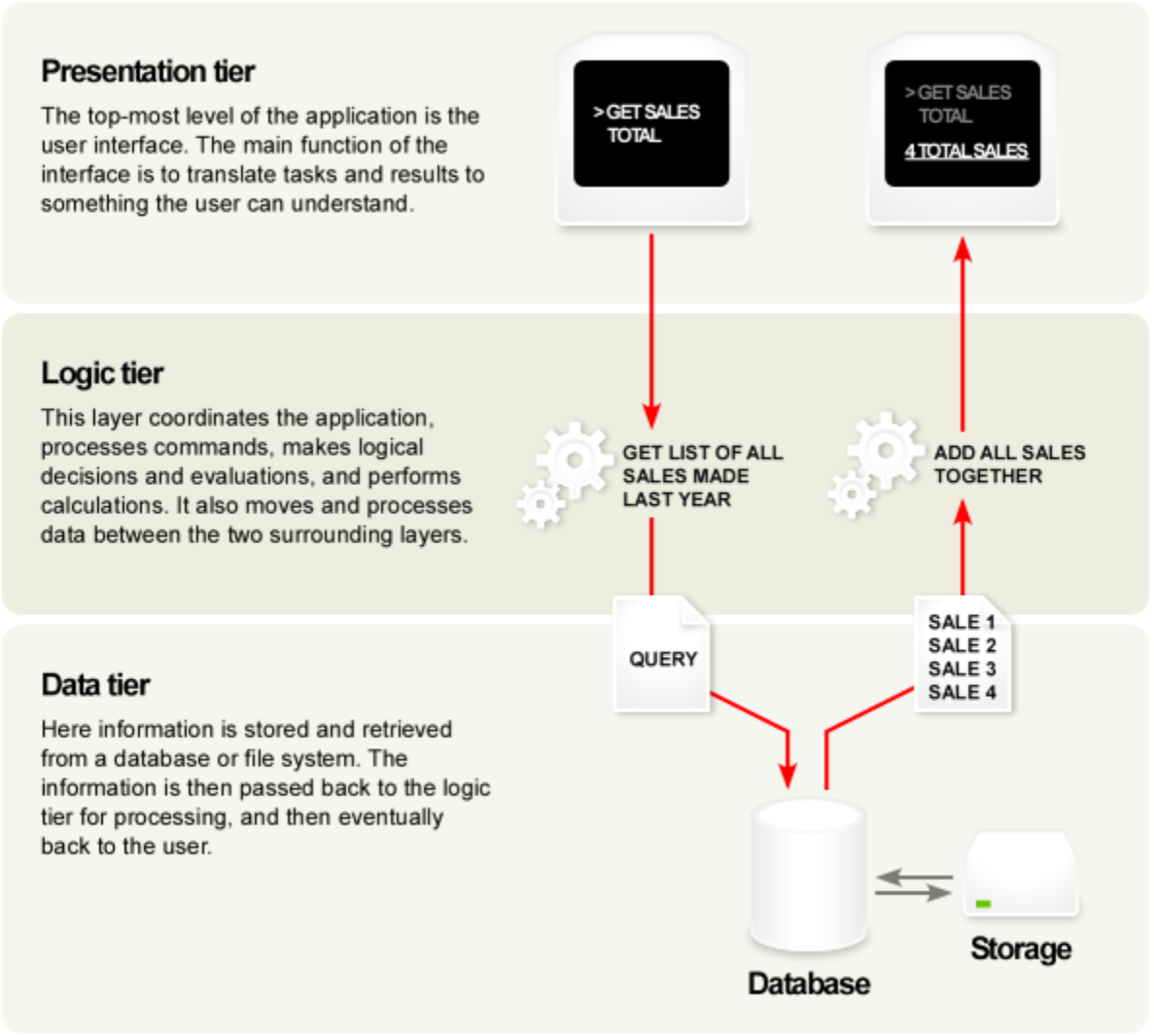
DB Technology's Trends

- Smaller and Smaller Systems
  - Originally: DBMS's were large, expensive software running on large computer
  - Today: can run on PC, Mobile…
- Bigger and Bigger Systems
  - Size of data has been increasing continuously
  - Parallel computing

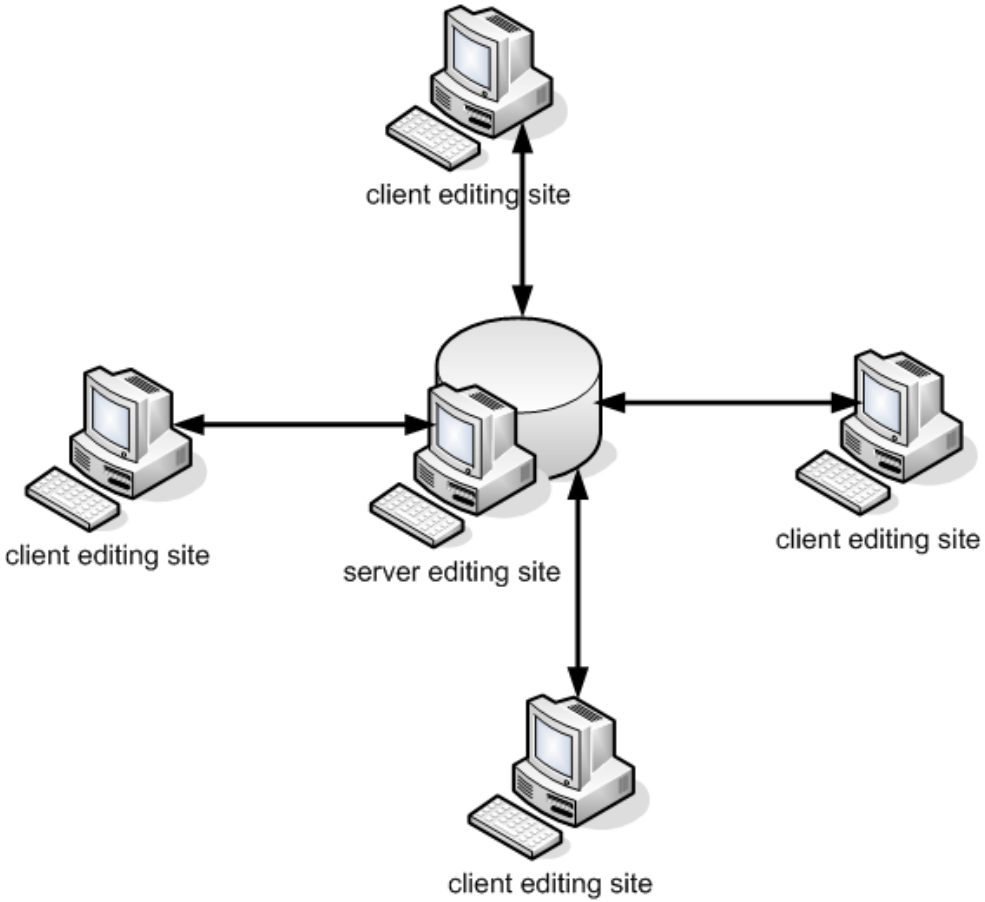
DB Technology's Trends (cont)

- Client-Server and Multi-Tier Architecture
  - DBMS is a server, application is client
  - Two – tier, three - tier (Website) Architecture
- Multimedia Data
  - Common form of multimedia data: Video, audio, radar signals, satellite images..
  - Big size
- Information Integration
  - Data Warehouse
  - Data Mining

3-tier Model


Presentation tier
The top-most level of the application is the user interface. The main function of the interface is to translate tasks and results to something the user can understand.
>GET SALES TOTAL
>GET SALES TOTAL
4 TOTAL SALES
Logic tier
This layer coordinates the application, processes commands, makes logical decisions and evaluations, and performs calculations. It also moves and processes data between the two surrounding layers.
GET LIST OF ALL SALES MADE LAST YEAR
ADD ALL SALES TOGETHER
Data tier
Here information is stored and retrieved from a database or file system. The information is then passed back to the logic tier for processing, and then eventually back to the user.
QUERY
SALE 1
SALE 2
SALE 3
SALE 4
Storage
Database

Client-Server Model



DB Technology's Trends (cont)

- **Data on the Web and E-commerce Applications**
  - **XML** (eXtensible Markup Language).

```
<?xml version = "1.0" encoding = "utf-8" ?>
<BARS>
    <BAR><NAME>Joe's Bar</NAME>
        <BEER><NAME>Bud</NAME>
            <PRICE>2.50</PRICE></BEER>
        <BEER><NAME>Miller</NAME>
            <PRICE>3.00</PRICE></BEER>
    </BAR>
    <BAR> …
</BARS>
```

DB Technology's Trends (cont)

- **New demand, new functionality**
  - Scientific Applications
  - Image Storage and Management
  - Audio and Video data management
  - Data Mining
  - Time Series and Historical Data Management
  - **→ Need more research and development of DB systems**

Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantage and disadvantage

What is DBMS

- A *Database Management System (DBMS)* is a software package designed to maintain and utilize databases
  - A very large, integrated collection of data.
  - Models real-world *enterprise*
    - Entities (e.g., students, courses)
    - Relationships (e.g., how students relate to courses)

- Software that enables users to define, create and maintain the database and provides controlled access to the database

Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
    - Database languages
    - Relational databases
    - Advantage and disadvantage

Database Users

- Actors on the scene
  - **Database administrators (DBA):** responsible for authorizing access to the database, for co-ordinating and monitoring its use, acquiring software, and hardware resources, controlling its use and monitoring efficiency of operations.
  - **Database Designers:** responsible to define the content, the structure, the constraints, and functions or transactions against the database. They must communicate with the end-users and understand their needs.
  - **End-users:** they use the data for queries, reports and some of them actually update the database content.

Database Users (cont)

- Workers behind the scene:
  - **DBMS system designers and implementers:** design and implement the DBMS modules and interfaces as a software package
  - **Tool developers:** design and implement tool - the software packages that facilitate database system design and use, and help improve performance
  - **Operators and maintenance personnel:** system administration personnel who are responsible for the actual running and maintenance of the hardware and software environment for the database system

Typical DBMS Functionalities

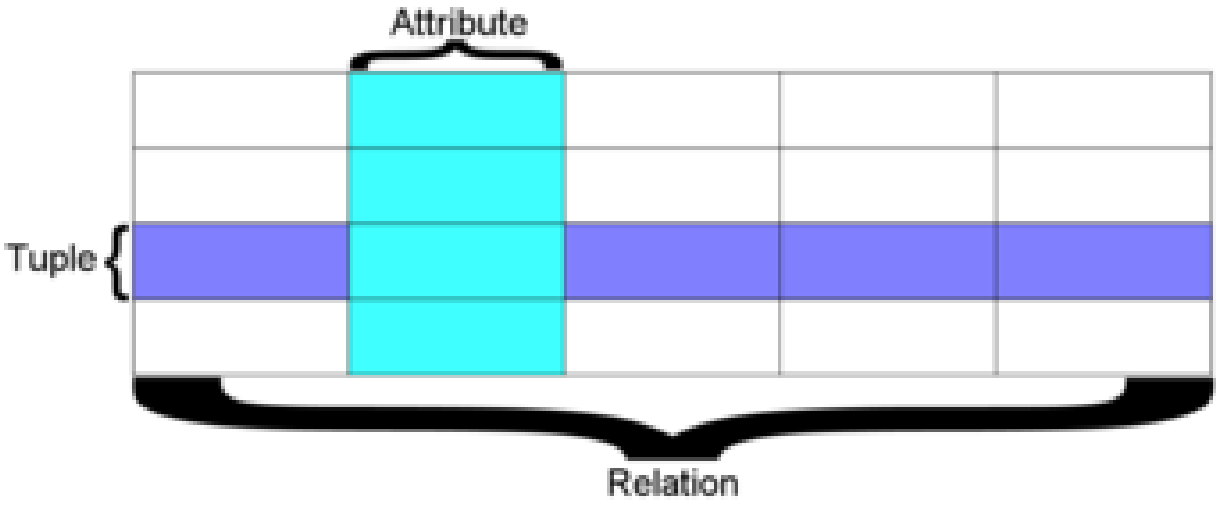
- **Define a database** : in terms of data types, structures and constraints
- **Construct or Load** the Database on a secondary storage medium
- **Manipulating the database** : querying, generating reports, insertions, deletions and modifications to its content
- **Concurrent Processing and Sharing** by a set of users and programs – yet, keeping all data valid and consistent

Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantage and disadvantage

Relational Database

- Base on Codd's theory
- Is Database that conforms to the relational model
- The most common DB model today



Database Languages

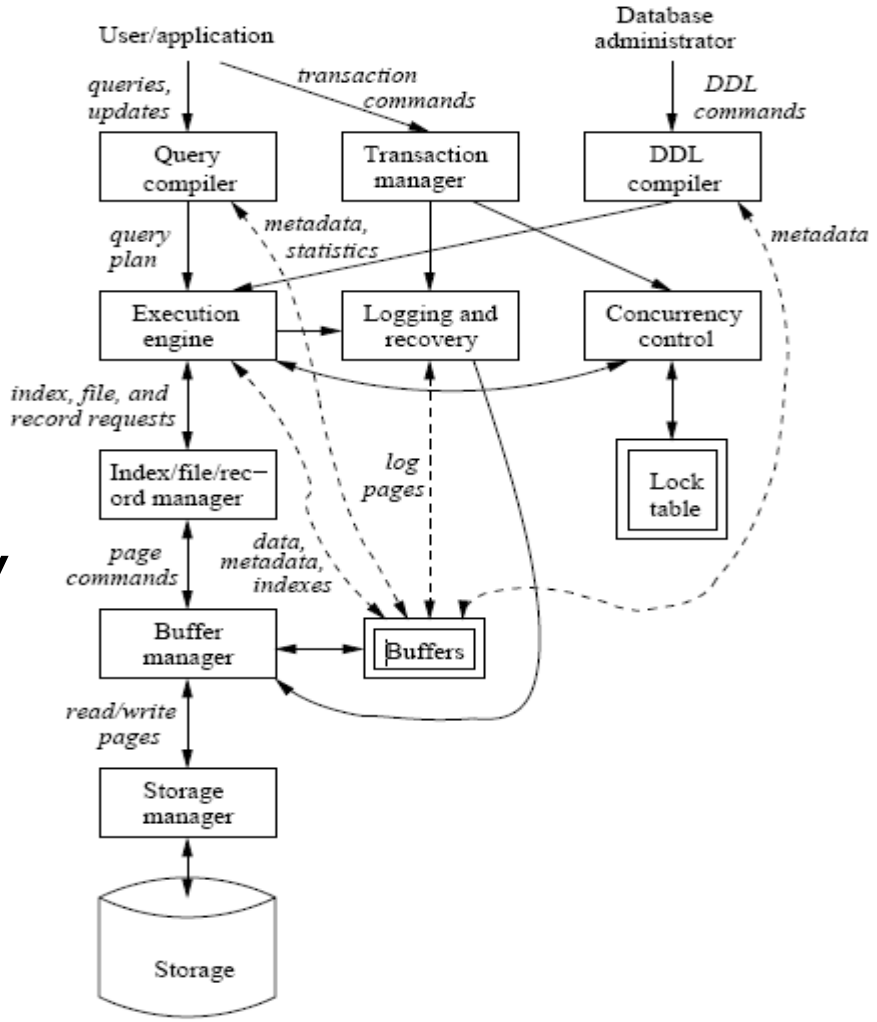
- DDL (Data-Definition Language)
  - Is Computer language for defining data structure
  - Initial: a subset of SQL: CREATE, DROP, ALTER
  - Generic sense: any formal language for describing data or information structures, like XML schemas.

Database Languages (cont)

- DML (Data-manipulation language)
  - Computer language used by computer programs or database users to retrieve, insert, delete and update data
  - Most Popular is SQL: SELECT, INSERT, UPDATE, DELETE
  - Other: IMS/DL1, CODASYL databases (such as IDMS)

DBMS Structure

- Single box: system component
- Double box: memory data structure
- Solid line: control & data flow
- Dashed line: data flow only



Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantage and disadvantage

RDBMS Products Dominate the DBMS Industry

- IBM has DB2
- Microsoft has SQL Server
- Oracle has 9i
- Sybase has SQL Anywhere
- Teradata has V2R5.0
  - Teradata is also one of the industry leaders in data warehouse/store software and data mining
  - Data mining derives knowledge from information in data files

Content

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantage and disadvantage

Advantages of Using the Database Approach

- Providing backup and recovery services.
- Providing multiple interfaces to different classes of users.
- Representing complex relationships among data.
- Enforcing integrity constraints on the database.
- Drawing Inferences and Actions using rules

When not to use a DBMS

- **Main inhibitors (costs) of using a DBMS:**
  - High initial investment and possible need for additional hardware.
  - Overhead for providing generality, security, concurrency control, recovery, and integrity functions.
- **When a DBMS may be unnecessary:**
  - If the database and applications are simple, well defined, and not expected to change.
  - If there are stringent real-time requirements that may not be met because of DBMS overhead.
  - If access to data by multiple users is not required.

When not to use a DBMS (cont)

- **When no DBMS may suffice:**
  - If the database system is not able to handle the complexity of data because of modeling limitations
  - If the database users need special operations not supported by the DBMS.

Summary

- Introduction, basic definitions
- History of DB
- Trends in DB Technology
- DMBS
  - Database Users
  - Database languages
  - Relational databases
  - Advantages and disadvantages