CMPUT 391
Database Management Systems

Li-Yan Yuan
University of Alberta,
Winter 2014

Objectives

- To **understand** the fundamental concepts and advanced technology underlying database systems:
  - Implementation techniques
  - Database design theory
  - Modern applications
- To **gain** hand-on experience with web-based database application systems
  - Oracle,
  - Java Servlets/JSP, PHP, or others such as Behavior Driven Development (BDD)
- To know some new technologies in the field
  - Hadoop
  - NoSQL database systems
  - Big Data

Course Topics

- Modern database systems
  - object-relational databases
  - active databases
  - NoSQL databases
- Relational database design method
  - dependency theory
  - normal forms
- Database implementation techniques
  - Query processing and optimization
  - Transaction management
    - → ACID and BASE
    - → concurrence control
  - crash recovery
  - Security and authorization
- Data mining and data warehouse

Course Topics

- Introduction to various systems
  - Network storage systems
    - → DAS, NAS, SAN
    - → EMC, IBM
  - Database Appliance
    - → IBM Netteza
    - → Oracle Exadata
  - Hadoop
  - Distributed platform

Textbooks

- Database Systems by M. Kifer, A. Berstein, and P. Lewis second edition
- Other on line materials

CMPUT391 pages on MOODLE

- https://eclass.srv.ualberta.ca/course/view.php?id=16138

Administration

- Office Hours
  - MW 12:00 -13:00 ATH 356
- Grading
  - →Assignments 10 %
  - →Project 35 %
  - →midterm exam 20 %
  - →Final exam 35 %
- Re-examination
  - →none, except as per regulation
  - →deferred exam must be scheduled by May 1, 2014.
- Collaboration
  - →Collaborate on assignments, but do not merely copy.
- No late submission of assignment and project will be accepted
- Course web page and the newsgroup

Laboratories

- Teach Assistants
  - Talat Syed (syedtala@cs.ualberta.ca)
  - Amine Trabelsi(atrabels@cs.ualberta.ca)
  - Lengdong Wu (lengdong@cs.ualberta.ca)
- Lab times
  - H01: M 17:00-19:50, CSC 219
  - H02: W 14:00-16:50, CSC 219
  - H03: R 14:00-16:50, CSC 219

Basic Notions of Database Management Systems

Data: any information

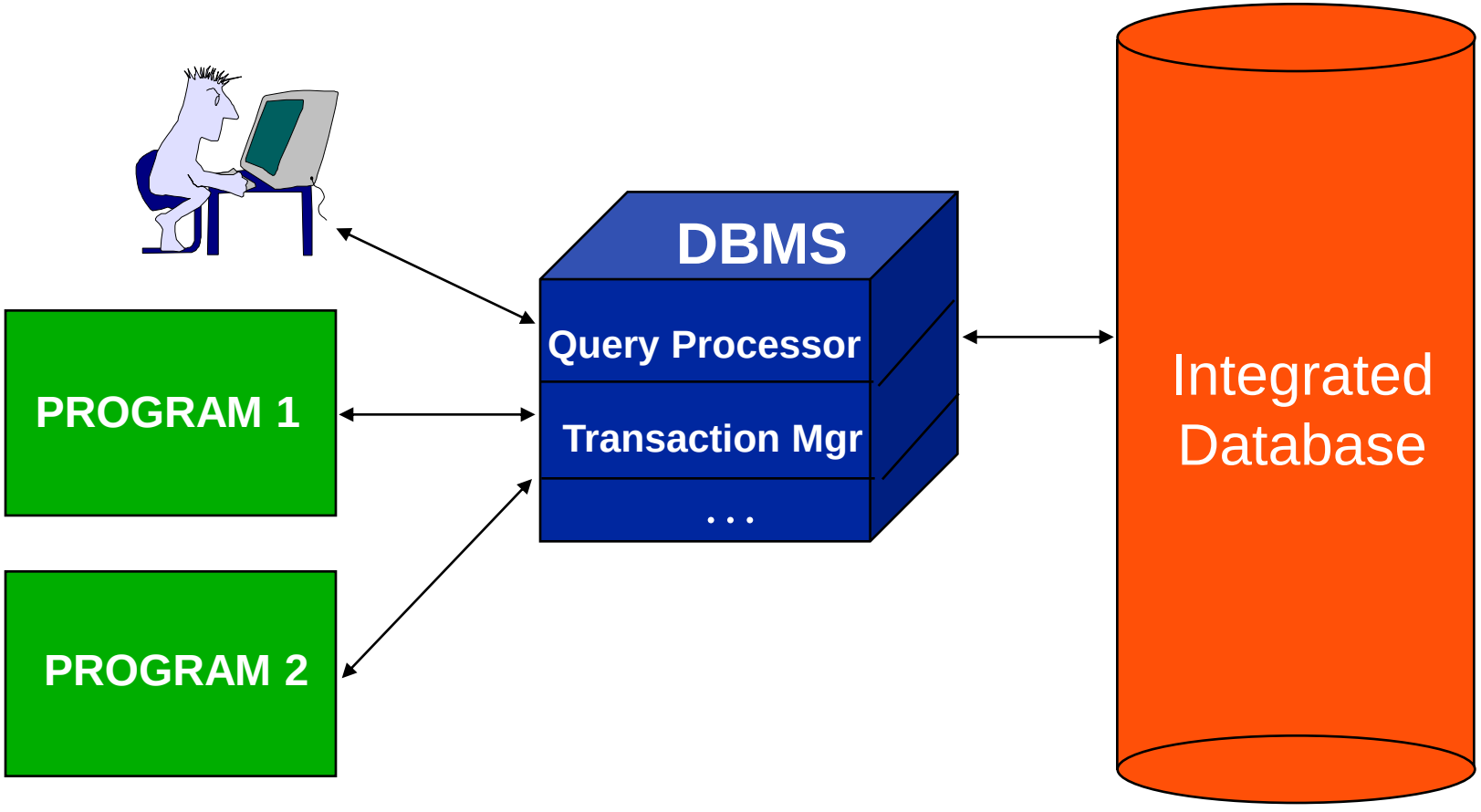
Manufacturing	Product data
University	Student data, courses
Hospital	Patient data, facilities
Bank	Account data

Database: a large collection of data
an integrated collection of data

Database management system:
a software system that provides an efficient as well as convenient environment for accessing data in a database.

Web-based database application systems (3-tier)
Client + Web-server + Database server

Database Approach



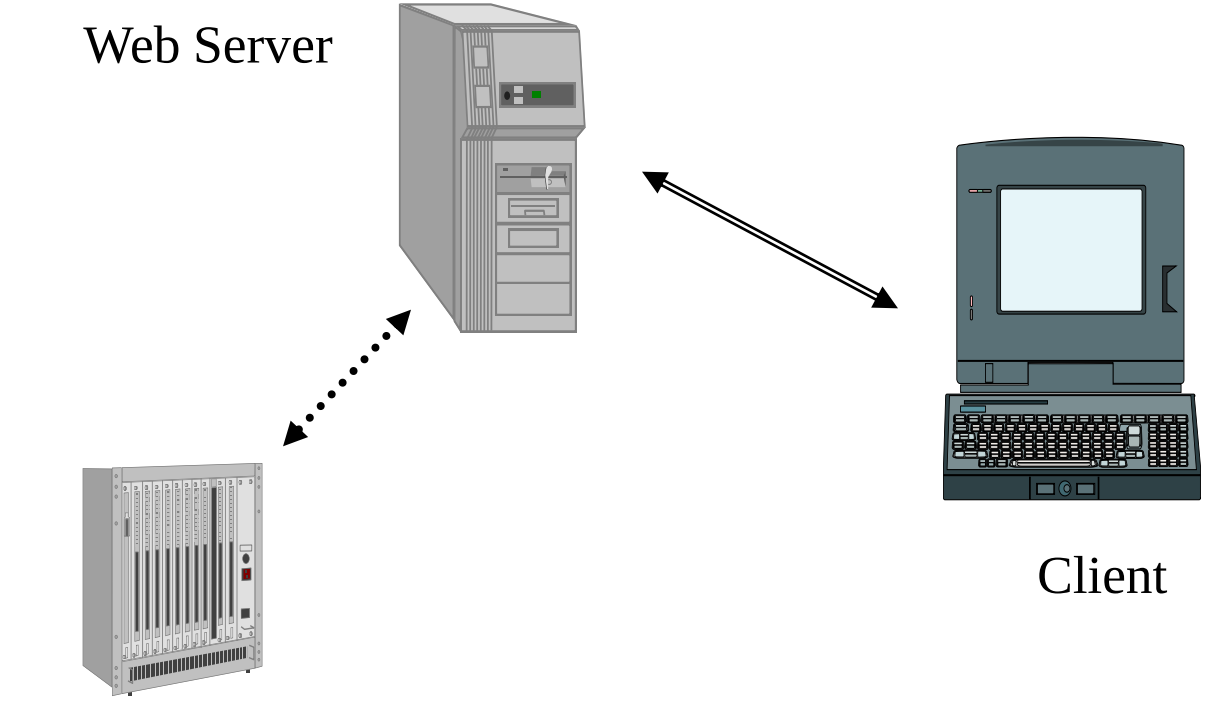
Functionalities of DBMSs

- Specifying the database structure
  - data definition language
- Manipulation of the database
  - query processing and query optimization
- Integrity enforcement
  - integrity constraints
- Concurrent control
  - multiple user environment
- Crash recovery
- Security and authorization

How to access a database system?

- SQLPlus == > Oracle
  - ad hoc
- JDBC / ODBC
  - A computer programs ==> Oracle / DB2 /MYSQL
- Client - Server Architecture
  - Terminals
  - Database Servers
- Web-based database applications

Web-based database applications



Database Applications

- On-Line Transaction Processing (OLTP)
  - Traditional mission critical applications
    - →Banking systems
    - →University registration systems
    - →On-line commerce
- On-Line Analytical Processing (OLAP)
  - Data mining
  - Decision making
- Big Data

What consists of a database system?

- Database Management System
  - Oracle / DB2 /MySQL
- Computer Servers
  - A Linux box
  - A cluster of servers
  - Mainframe computers
- Storage Systems
  - RAID
  - Storage systems
    - →EMC / IBM SAN
- Operation Systems
  - Linux, Unix, Windows, …

Any problems with such a configuration?

- Complicated and expensive
  - One needs to know all four major components
- Price Tag?
  - Nothing cheap
- Scalability
  - What if the data increase to P, not just a few Ts ?
  - What if the number of clients increases to 1 million?

Possible Solutions?

- Grid database
  - Oracle 11G: needs oracle cluster
- Database Appliance: one product for all four components
  - IBM Netezza
  - Oracle Exadata
- Distributed file systems
  - MapReduce
  - Hadoop
- NoSQL systems
  - SQL
  - Distributed database systems