Introduction to Database Systems
CSE 444

Lecture #1

March 26, 2007

About Me

Dan Suciu:

- Joined the department in 2000
- Before that: Bell Labs, AT&T Labs

Research:

- Past: XML and semi-structured data:
  - Query language: XML-QL (later XQuery)
  - Compressor: XMill
  - Theory: XPath containment, XML typechecking
- Present: Probabilistic databases: MystiQ

Staff

- Instructor: Dan Suciu
  - Allen, Room 662, suciu@cs.washington.edu
    Office hours: Fridays 11:30-12:30 (appointment strongly recommended)
- TAs:
  - Cam Thach Nguyen, ncthach@cs.washington.edu
    Office hours: TBA

Communications

- Web page: http://www.cs.washington.edu/444/
  - Lectures will be available here
  - Homework will be posted here (HW1 is posted)
  - The project description will be here
- Mailing list:
  - Announcements, group discussions
  - Please subscribe

Textbook(s)

Main textbook, available at the bookstore:

- *Database Systems: The Complete Book*, Hector Garcia-Molina, Jeffrey Ullman, Jennifer Widom

Most chapters are good. Some are not (functional dependecies).
COME TO CLASS ! ASK QUESTIONS ! READ SLIDES !

Other Texts

Available at the Engineering Library (not on reserve):

- *Database Management Systems*, Ramakrishnan
- *XQuery from the Experts*, Katz, Ed.
- *Fundamentals of Database Systems*, Elmasri, Navathe
- *Foundations of Databases*, Abiteboul, Hull, Vianu
- *Data on the Web*, Abiteboul, Buneman, Suciu

Outline of Today's Lecture

1. Overview of DBMS

2. DBMS through an example

3. Course outline

4. Assignment 1, Homework 1

Database

What is a database ?

Give examples of databases

Database

What is a database ?

- A collection of files storing related data

Give examples of databases

- Accounts database; payroll database; UW’s students database; Amazon’s products database; airline reservation database

Database Management System

What is a DBMS ?

Give examples of DBMS

Database Management System

What is a DBMS ?

- *A big C program written by someone else that allows us to manage efficiently a large database and allows it to persist over long periods of time*

Give examples of DBMS

- DB2 (IBM), SQL Server (MS), Oracle, Sybase
- MySQL, Postgres, …

Market Shares

From 2004 www.computerworld.com

- IMB: 35% market with $2.5BN in sales
- Oracle: 33% market with $2.3BN in sales
- Microsoft: 19% market with $1.3BN in sales

An Example

The Internet Movie Database
http://www.imdb.com

- Entities:
  Actors (800k), Movies (400k), Directors, …
- Relationships:
  who played where, who directed what, …

Tables

Directors:

id	**fName**	**lName**
15901	Francis Ford	Coppola
. . .		

Movie_Directors:

id	**mid**
15901	130128
. . .	

Movies:

mid	**Title**	**Year**
130128	The Godfather	1972
. . .		

What the Database Systems Does

1. Create/store large datasets
2. Search/query/update
3. Change the structure
4. Concurrent access to many user
5. Recover from crashes
6. Security

Possible Organizations

- Files
- Spreadsheets
- DBMS

1. Create/store Large Datasets

- Files

Yes, but…

- Spreadsheets

Not really…

- DBMS

Yes

2. Search/Query/Update

- Files

Simple queries (grep);
Updates are difficult

- Spreadsheets

Simple queries;
Simple updates

- DBMS

All

Updates: generally OK

3. Change the Structure

Add *Address* to each Actor

Very hard

- Files

Yes

- Spreadsheets

Yes

- DBMS

4. Concurrent Access

Multiple users access/update the data concurrently

Lost updates; inconsistent reads,…

- What can go wrong ?
- How do we protect against that in OS ?
- This is insufficient in databases; why ?

locks

A logical action consists of multiple updates

5. Recover from crashes

- Transfer $100 from account #4662 to #7199:

```
X = Read(Account, #4662);
X.amount = X.amount - 100;
Write(Account, #4662, X);

Y = Read(Account, #7199);
Y.amount = Y.amount + 100;
Write(Account, #7199, Y);
```

CRASH !

What is the problem ?

6. Security

- Files

File-level access control

- Spreadsheets

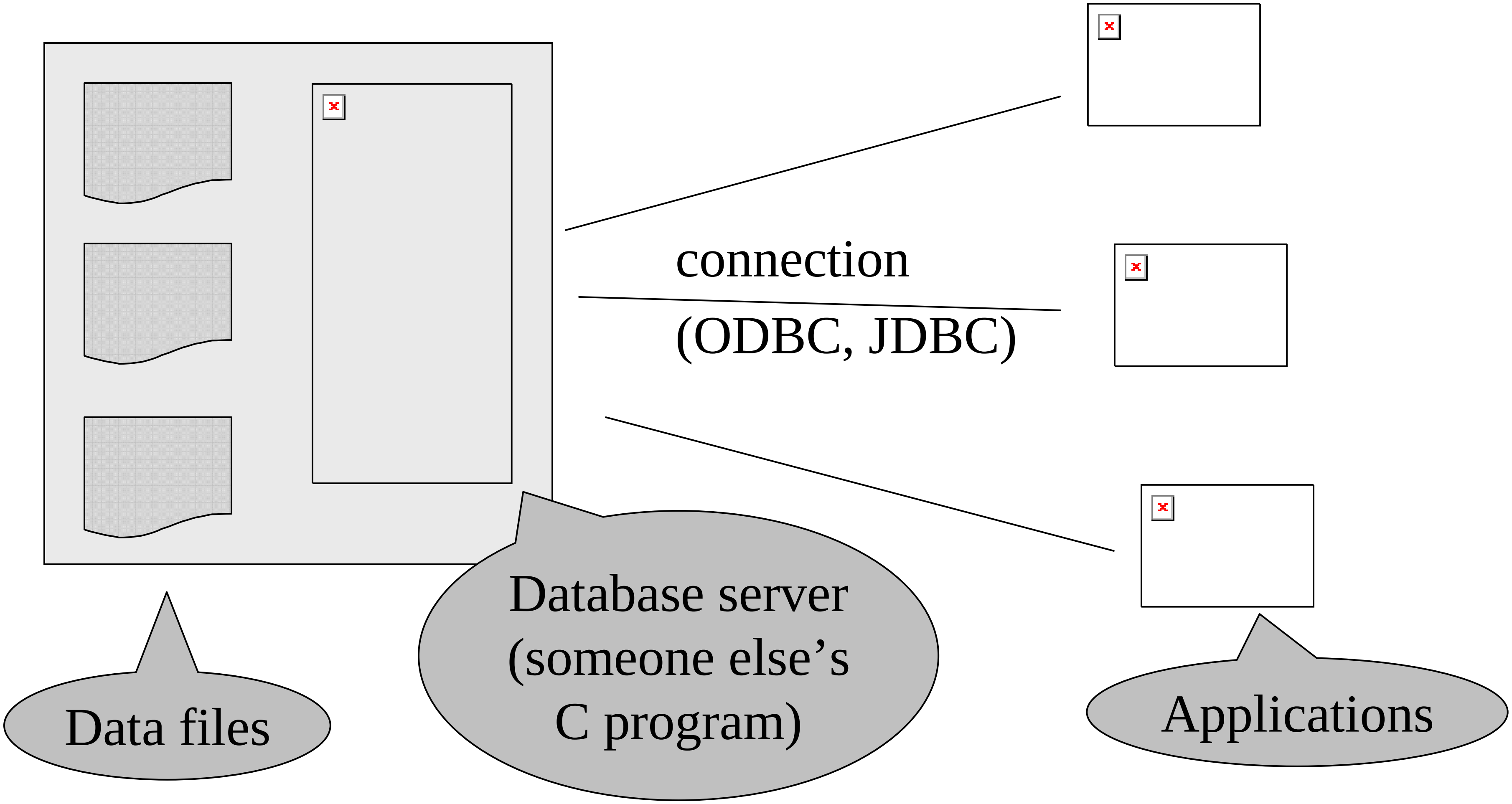
Same [?]

- DBMS

Table/attribute-level access control

Enters a DMBS

"Two tier system" or "client-server"



Data Independence

Logical view

Directors:

id	**fName**	**lName**
15901	Francis Ford	Coppola
. . .		

Movie_Directors:

id	**mid**
15901	130128
. . .	

Movies:

mid	**Title**	**Year**
130128	The Godfather	1972
. . .		

Directors_file

Moviews_title_index_file

Directors_fname_index_file

Movies_file

Physical view

What the Database Systems Does

1. Create/store large datasets
2. Search/query/update
3. Change the structure
4. Concurrent access to many user
5. Recover from crashes
6. Security

SQL DML

SQL DDL

Transactions
ACID

Grant, Revoke, Roles

Course Outline - TENTATIVE !!

1. 3/26: SQL
2. 4/2: SQL in C#; Database Design: E/R, NF
3. 4/9: Views, Constraints, Security
4. 4/16: XML/XPath/XQuery
5. 4/23: Transactions
6. 4/30: Database storage, indexes
7. 5/7: Physical operators, optimization
8. 5/14: Statistics, Database tuning
9. 5/21: Advanced topics (or slack)

Grading (TENTATIVE)

- Homework 30%
- Project 25%
- Midterm 15%
- Final 25%
- Intangibles: 5%

If we get 2nd TA

Reading Assignment

- Reading assignment for Wed, March 28
  - **Introduction** from **SQL for Web Nerds,** by Philip Greenspun, http://philip.greenspun.com/sql/

- This is a one-time assignment, no grading, BUT *very* instructive and lots of fun reading

Homework 1

- Homework 1:
  - SQL Queries
  - Due Friday, April 6
  - It is posted already!
- Homework 2:
  - Conceptual design: E/R diagrams, Normal Forms
  - Due Friday, April 20
- Homework 3:
  - XML/Xquery
  - Due Friday, May 4
- Homework 4:
  - Transactions: concurrency control and recovery
  - Due Friday, May 18

The Project:
Boutique Online Store

- Phase 1:
  - Design a Database Schema, Build Related Data Logic
  - Due Friday, April 13
- Phase 2:
  - Import data, Web Inventory Data Logic
  - Due Friday, April 27
- Phase 3:
  - Checkout Logic
  - Due Friday, May 11
- Phase 4:
  - Database Tuning
  - Due Friday, May 25

Project

SQL Server, C#, ASP.NET

- Supported
- Will provide starter code in C#, ASP.NET
- The import data is in SQL/XML on SQL Server

Alternative technologies: MySQL, postgres, PHPs

- Not supported (you are on your own)
- Worry about the SQL/XML part…

Accessing SQL Server

SQL Server Management Studio

- Server Type = Database Engine
- Server Name = IPROJSRV
- Authentication = SQL Server Authentication
  - Login = your UW email address (*not* the CSE email)
  - Password = 12345

Change your password !!

Then play with IMDB, start thinking about HW1