CSE 544
Principles of Database Management Systems

Magdalena Balazinska (magda)

Fall 2007

Lecture 1 - Class Introduction

Outline

- Introductions
- Class overview
- What is the point of a db management system (DBMS)?
- Main DBMS features and DBMS architecture overview

Course Staff

- Instructor: Magda
  - `magda@cs.washington.edu`
  - Office hours by appointment
  - Location: cse 550

- TA: Nodira Khoussainova
  - Graduate student in the database group
  - `nodira@cs.washington.edu`
  - Office hours: Wednesday 11am-12pm
  - Location: CSE 218

Who is Magda?

- Assistant Professor since January 2006
- PhD from MIT, February 2006

- Areas of interest: databases and systems
- Current research focus
  - Stream processing & real-time monitoring
  - RFID data managment

Current Research Projects

- RFID Ecosystem
  - Tracking people and objects in the Paul Allen Center
  - Data management sys. for RFID data & longitudinal studies
- StreamClean/PEEX
  - Probabilistic RFID data cleaning
  - Probabilistic event extraction
- Moirae
  - Integrating history into monitoring systems
- Mobisensors
  - Data management system for mobile sensor networks
- FlowDB
  - Forensic analysis in network intrusion detection systems

Potential Class Disruption

- Yes, I am pregnant
- Due date: December 12th... 1 week after last lecture
  - Last lecture: hard deadline for project presentations and reports
- So, hopefully, there should be no disruptions
- In case we don't quite make it
  - If I don't show up for lecture, please wait 15 minutes, then assume the class is canceled. Check email for details.
  - Nodira can lead the last few lectures
  - We might have to reschedule the final presentations

Goals of the Class

- Study principles of data management
  - Data models, data independence, normalization
  - Data integrity, availability, consistency, etc.
- Study key DBMS design issues
  - Storage, query execution and optimization, transactions
  - Replication, distribution, streaming & sensor data
  - Data warehousing, (automatic) database tuning, etc.
- Ensure that
  - You are comfortable using a DBMS
  - You can write (web-based) applications that use a DBMS as a back-end

Class Format

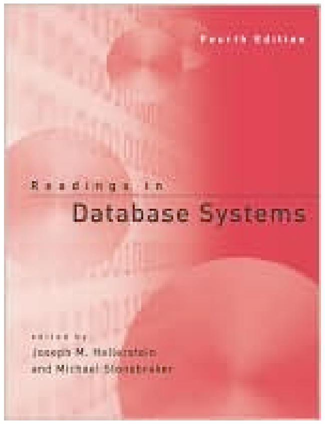
- Two lectures per week: MW @ 1:30pm

- Mix of lecture and discussion
  - Mostly based on papers
  - Must read papers before lecture
  - Come prepared to discuss them

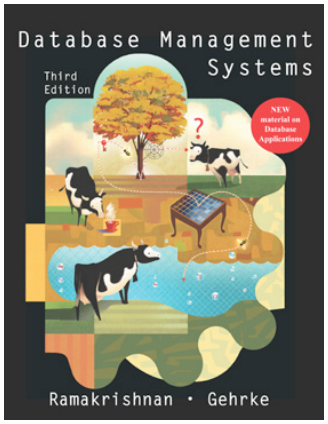
- One guest lecture: Surajit Chaudhuri
  - From Microsoft Research
  - Will talk about automatic database tuning

Readings and Notes

- Readings are based on papers
  - Mix of old seminal papers and new papers
  - Papers available online
  - Many come from the “red book”



- Background readings from the following book
  - Database Management Systems. **Third Ed**. Ramakrishnan and Gehrke. McGraw-Hill.


Database Management Systems
Third Edition
NEW material on Database Applications
Ramakrishnan · Gehrke

- Lecture notes (the ppt slides)
  - Posted on the class website
  - Available after each lecture

Class Resources

- Website: lectures, assignments, projects

  http://www.cs.washington.edu/544

  List of all the **deadlines**

- Mailing list:

  cse544@cs.washington.edu

  Make sure you register!

Evaluation

- **Class participation 15%**
  - Paper readings and discussions
- **Paper reviews 20%**
  - Due by 10pm the night before the lecture
  - Check class website for reading questions
- **Assignments 20%**
  - HW1: Using a database (SQL, views, indexes, etc.)
  - HW2: Writing a Web app with a db back-end
- **Project 45%**
  - Small research project but you must start it now!

Class Participation

- An important part of your grade

- Because
  - We would like you to read and think about papers throughout the quarter
  - Important to learn to discuss papers

- Expectations
  - Ask questions, raise issues, think critically
  - Learn to express your opinion
  - Respect other people's opinions

Paper reviews

- Between 1/2 page and 1 page in length
- Summary of the main points of the paper
- You can also add a discussion
  - (your thoughts about the paper)
- Reading questions
  - For some papers, we will post reading questions to help you figure out what to focus on when reading the paper
  - Please address these questions in your reviews

Assignments

- Goal: hands-on experience using a DBMS

- HW1. **Due October 17th.** Will soon be posted on website
  - Setup a db from scratch
  - Practice writing SQL queries
  - Browse the system catalog
  - Get experience with integrity constraints & triggers
  - Play with indexes and views

- HW2. Due October 31st.
  - Writing an application that uses a db as a back-end

Project Overview

- Choose from a list of mini-research topics
- Or come up with your own
- Can be related to your ongoing research
- Must be related to databases
- Must contain some element of research or significant engineering (there must be something novel)
- Open ended

- Final deliverables
  - Short research paper (8 pages)
  - Conference-style presentation

Project Goals

- Apply database principles to a new problem
  - Understand and model the problem
  - Research and understand related work (2-3 papers)
  - Propose some new approach
    - Creativity will be evaluated
  - Implement some parts
  - Evaluate your solution
  - Write-up and present your results

- Amount of work may vary widely between groups

Project Milestones

- Oct 1st: teams formed
- Oct 10th: project proposal
- Nov 12th: milestone report
- Dec 5th: final report and project presentations

- More details on the website, including project ideas

- We will meet with you regularly throughout the quarter

Let’s get started

- What is a database?

- Give examples of databases

Let’s get started

- 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

Data Management

- Data is valuable but hard and costly to manage

- Example: Store database
  - **Entities**: employees, positions (ceo, manager, cashier), stores, products, sells, customers.
  - **Relationships**: employee positions, staff of each store, inventory of each store.

- What operations do we want to perform on this data?
- What functionality do we need to manage this data?

Required Functionality

1. Describe real-world entities in terms of stored data
2. Create & persistently store large datasets
3. Efficiently query & update
    1. Must handle complex questions about data
    2. Must handle sophisticated updates
    3. Performance matters
4. Change structure (e.g., add attributes)
5. Concurrency control: enable simultaneous updates
6. Crash recovery
7. Access control, security, integrity

Difficult and costly to implement all these features

Database Management System

- A DBMS is a software system designed to provide data management services

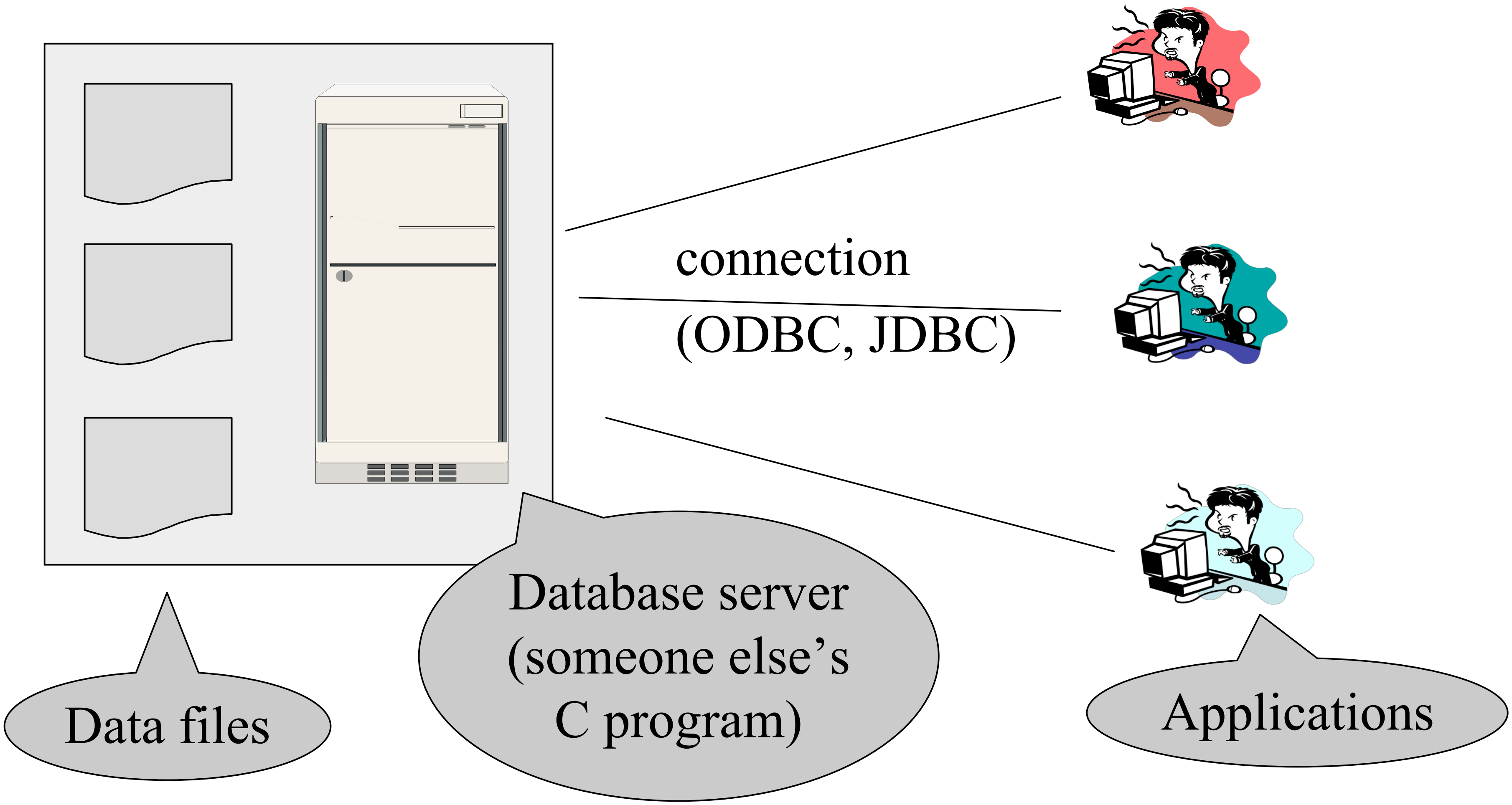
- Examples of DBMS
  - Oracle, DB2 (IBM), SQL Server (Microsoft),
  - PostgreSQL, MySQL,…

Market Shares

- In 2004 (from www.computerworld.com)
  - IBM, 35% market with $2.5 billion in sales
  - Oracle, 33% market with $2.3 billion in sales
  - Microsoft, 19% market with $1.3 billion in sales

Typical System Architecture

“Two tier system” or “client-server”



connection

(ODBC, JDBC)

Data files

Database server (someone else’s C program)

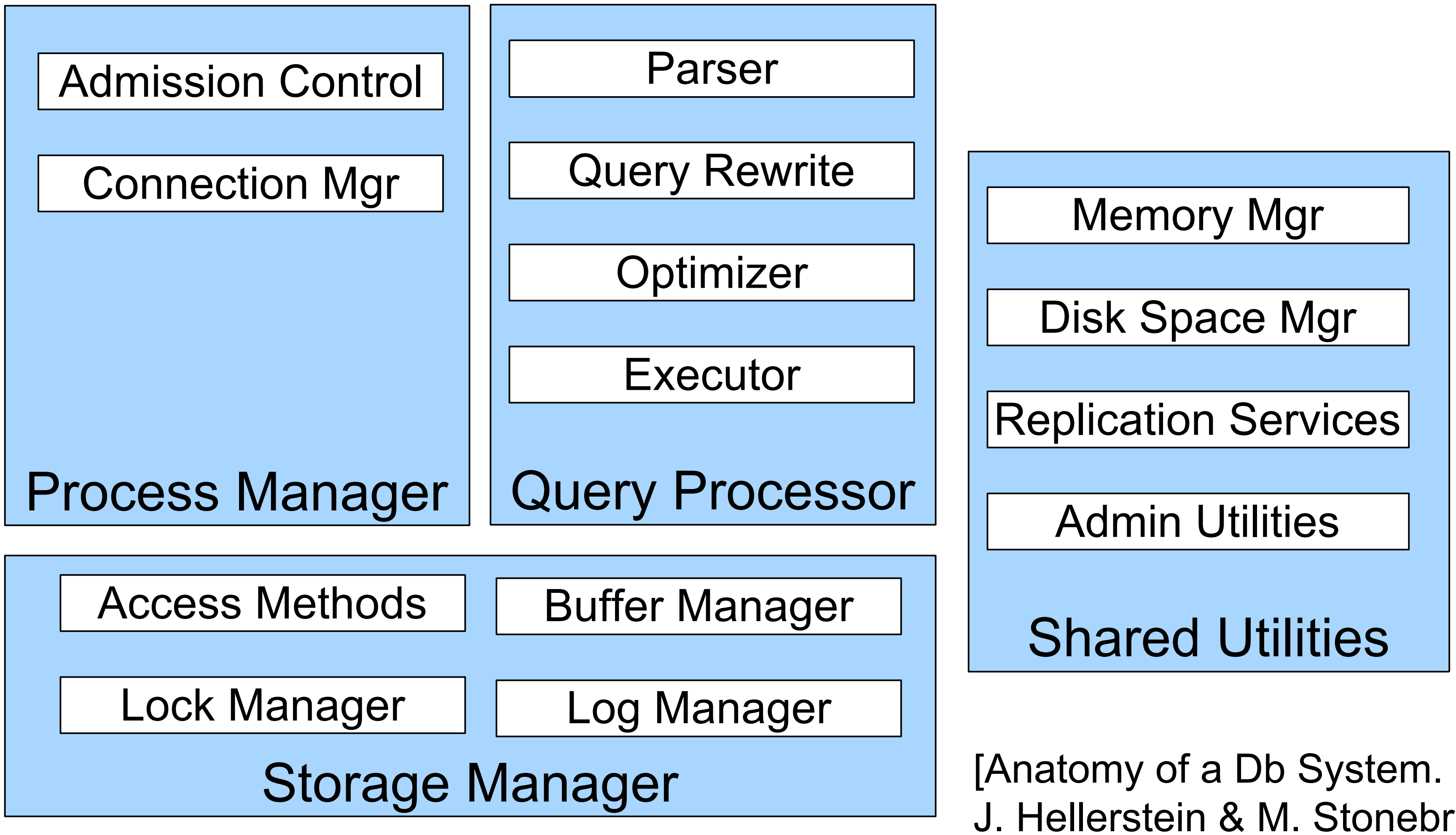
Applications

Main DBMS Features

- Data independence
  - Data model
  - Data definition language
  - Data manipulation language
- Efficient data access
- Data integrity and security
- Data administration
- Concurrency control
- Crash recovery
- Reduced application development time

How to decide what features should go into the DBMS?

A Quick Look Inside a DBMS



[Anatomy of a Db System.
J. Hellerstein & M. Stonebraker.
Red Book. 4ed.]

When not to use a DBMS?

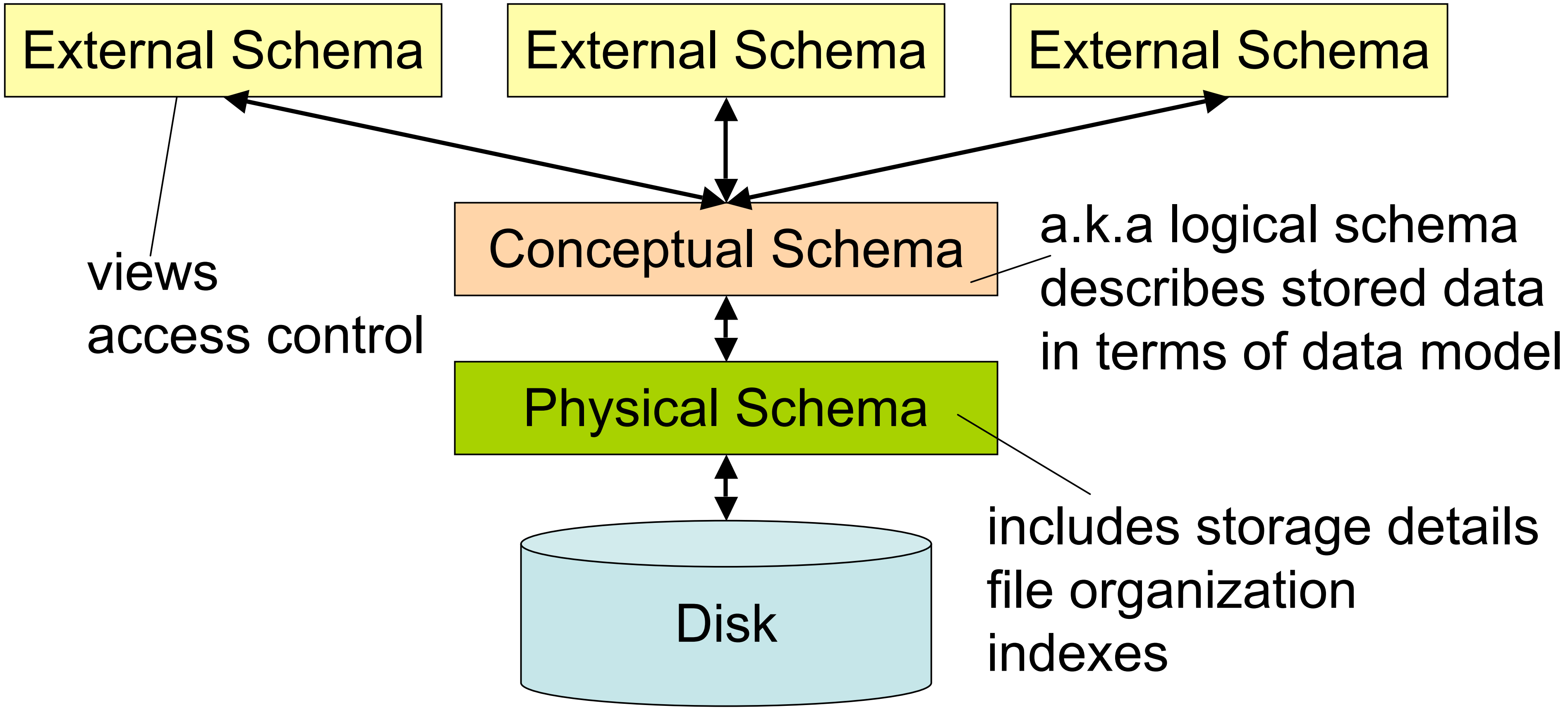
- DBMS is optimized for a certain workload

- Some applications may need
  - A completely different data model
  - Completely different operations
  - A few time-critical operations

- Examples
  - Text processing
  - Scientific analysis

Preview for Next Week

Levels of abstraction in a DBMS



External Schema

External Schema

External Schema

Conceptual Schema

Physical Schema

Disk

views
access control

a.k.a logical schema
describes stored data
in terms of data model

includes storage details
file organization
indexes