CSE 132A
Database System Principles

The slides are partially based on Prof. Victor Vianu’s slides for past offerings of CSE132A

Alin Deutsch

UC San Diego
introduction
“Databases”

What comes to your mind?
Data. Why data?

- Data play already an important role in our lives
Data in everyday life

Let’s go have some coffee!

Find coffee shop

Yelp

Maps

Stores

Reviews

Maps

Traffic
Data in everyday life

Let’s go have some coffee!

Order coffee

Coffee House

Bank

Fraud Detection

Inventory

Sales

Accounts

Statistics
Data in everyday life

Let’s go have some coffee!
Data. Why data?

- Data play already an important role in our lives…
- …and computer science is becoming more and more data-centric
Data in Computer Science

- Web knowledge harvesting
- Crowd sourcing
- Cloud computing
- Scientific databases
- Networks
- Streaming sensor monitoring
- Social networks
- Geographic information systems
- Big data…
Big Data in the news

- Why Big Data Will Drive the Next Big Boost in Productivity
  BY DAVID STEINBERG, XL MARKETING 06.28.13
  Wired

- The Big Data Opportunity in Travel
  The Wall Street Journal

- Big Data is a Big Deal
  Forbes

- Four Big Data Trends That Change Everything
  Gigaom

- Goldman Sachs invests $100M in 13-year-old big data firm
  Gigaom

- To launch the initiative, six Federal departments and agencies will announce more than $200 million in new commitments...
  The White House
How to manage data?

- Store data
- Query/retrieve data
- Update data

This class:
Learn basic concepts behind data management systems
Database System

• Tailored to a particular application
  Very time-consuming to design, implement and optimize
Database Management System

• A generalized database system
• Used in a variety of application environments
• Provides common approach to:
  Data organization
  Data storage
  Data access
  Data control

• Examples:
  Ingres/PostgreSQL, DB2, Oracle, SQL Server, MySQL, etc

Leverage years of research gone into the design of the DBMS
History of Database Systems

1950 - early 1960s

• Data processing based on magnetic tapes
  Magnetic tapes for data storage
  Punched cards for data input

• Need to process data in specific order
  Magnetic tapes allow only sequential order to data
History

late 1960s - 1970s

- Advent of disks
  - No need for sequential processing
  - Design of data structures for data storage and processing

- Relational model
  - Proposed by Codd
  - (who won the Turing Award for his work)
  - Non-procedural way of querying data
1980s

• Research relational prototypes…
  System R (by IBM Research)
  Ingres (by UC Berkeley)

• …lead to commercial relational systems
  IBM DB2, Oracle, Ingres, DEC Rdb
History

1990s

- The WWW era
- New requirements for DBMSs
  - High availability
  - Support for web interfaces to data
- From transaction (update-intensive) processing to query-intensive applications (decision support and data mining)
History

2000s - 2010s

• XML & XQuery standards
• Automated DB administration
  Auto-admin
• Open-source DBMSs
  PostgreSQL, MySQL
• Specialized DBMSs for big data management
  Column stores, highly parallel DBMSs, NoSQL
• Tremendous expansion across computer science
class info
Goal of the class

- Learn basic principles of database management systems (DBMSs) with an emphasis on the relational model
- Get hand-on experience using a real DBMS through programming assignments
- Acquire the necessary background for follow-up database courses
  - CSE 132B: Database System Applications
  - CSE 135: Online Analytics Applications
Content

• Database Management System Overview

• Relational databases
  The relational model
  Commercial query languages: SQL (& QBE)
  Formal query languages: relational algebra & calculus
  Indexing (sequential files, B-trees)
  Schema design: normal forms & E-R model

• Other topics as time allows
  Column Store, MapReduce, NoSQL
Course Composition

Tentative (to be adjusted based on time allowance):

- Homeworks: 15%
- Programming Assignments (SQL + JDBC): 30%
- Midterm: 25%
- Final: 30%
Textbook

• Main textbook (recommended, not required):
  "Database System Concepts"
  *by Silberschatz, Korth and Sudarshan, 6th Edition*

• Alternative textbook:
  "Database Systems: The Complete Book"
  *by H. Garcia-Molina, J.Ullman and J. Widom, Prentice Hall*

• For those interested in database theory:
  "Foundations of Databases"
  *by S.Abiteboul, R.Hull and V.Vianu, Addison-Wesley, 1995*
Other resources

- **Class web-site**
  http://cseweb.ucsd.edu/classes/fa17/cse132A-a

- **Piazza discussion forum**

Announcements will be made to piazza. Check both *daily* for updates!
Course Policies

• No plagiarism!
• Re-grade
  Request at TA within 1 week from the day of return
• Makeups or extensions only in true hardship cases
  Contact me as soon as possible
• E-mail
  Prefer office hours & Piazza. If you use it, conciseness is a virtue!

Read policies on the course web-site!
Contact me for questions.
Databases@UCSanDiego

• Database group
  Prof. Alin Deutsch
  Prof. Arun Kumar
  Prof. Yannis Papakonstantinou
  Prof. Victor Vianu

• Group web-site
  http://db.ucsd.edu

• Interested in database research?
  Contact any of us!