



CS425 – Fall 2016

Boris Glavic

Course Information

Modified from:

Database System Concepts, 6th Ed.

©Silberschatz, Korth and Sudarshan
See www.db-book.com for conditions on re-use



Hi, I am **Boris Glavic**,
Assistant Professor in
CS





Hi, I am **Boris Glavic**,
Assistant Professor in
CS

I am a **database** guy!





Hi, I am **Boris Glavic**,
Assistant Professor in
CS

I am a **database** guy!



I will teach you:
database stuff



Why are Databases Important?

- **What do Databases do?**
 1. Provide persistent storage
 2. Efficient declarative access to data -> Querying
 3. Protection from hardware/software failures
 4. Safe concurrent access to data

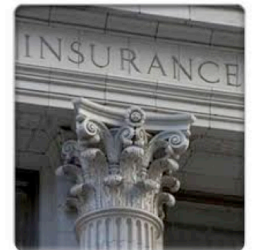


Who uses Databases?

- Most big software systems involve DBs!
 - Business Intelligence \Rightarrow e.g., IBM Cognos
 - Web based systems
 - ...
- **You!** (desktop software)
 - Your music player \Rightarrow e.g., Amarok
 - Your Web Content Management System
 - Your email client
 - ...
- **Every** big company
 - Banks
 - Insurance
 - Government
 - Google, ...
 - ...



Joomla!™





Who Produces Databases?

■ Traditional relational database systems is big business

- IBM ⇒ DB2
- Oracle ⇒ Oracle ☺
- Microsoft ⇒ SQLServer
- Open Source ⇒ MySQL, Postgres, ...

■ Emerging distributed systems with DB characteristics and Big Data

- Cloud storage and Key-value stores ⇒ Amazon S3, Google Big Table, ...
- Big Data Analytics ⇒ Hadoop, Google Map & Reduce, ...
- SQL over Distributed Platforms ⇒ Hive, Tenzing, ...





Why are Database Interesting (for Students)?

■ The pragmatic perspective

- Background in databases make you competitive in the job market ;-)

■ Systems and theoretical research

- Database research has a strong systems aspect
 - ▶ Hacking complex and large systems
 - ▶ Low-level optimization
 - cache-conscious algorithms
 - Exploit modern hardware
- Databases have a strong theoretical foundation
 - ▶ Complexity of query answering
 - ▶ Expressiveness of query languages
 - ▶ Concurrency theory
 - ▶ ...



Why are Database Interesting (for Students)?

- Connection to many CS fields
 - Distributed systems
 - ▶ Getting more and more important
 - Compilers
 - Modeling
 - AI and machine learning
 - ▶ Data mining
 - Operating and file systems
 - Hardware
 - ▶ Hardware-software co-design



Webpage and Faculty

■ Course Info

- **Course Webpage:** <http://cs.iit.edu/~cs425>
- **Google Group:** <https://groups.google.com/d/forum/cs425-2016-fall-group>
 - ▶ Used for announcements
 - ▶ Use it to discuss with me, TA, and fellow students
- **Syllabus:** <http://cs.iit.edu/~cs425/files/syllabus.pdf>

■ Faculty

- **Boris Glavic** (<http://cs.iit.edu/~glavic>)
- **Email:** bglavic@iit.edu
- **Phone:** 312.567.5205
- **Office:** Stuart Building, room 226C
- **Office Hours:** Mondays, 12pm-1pm (and by appointment)



TAs

- Tas
- TBA



Workload and Grading

■ Exams

- Midterm (25%)
- Final (35%)

■ Homework Assignments (preparation for exams!) – 20%

- HW1 (Relational algebra)
- HW2 (SQL)
- HW3 (Database modeling)

■ Course Project (20%)

- In groups of 3 students
- Given an example application (e.g., ticketing system)
 - ▶ Develop a database model
 - ▶ Derive a database schema from the model
 - ▶ Implement the application accessing the database



Course Objectives

- Understand the underlying ideas of database systems
- Understand the **relational data model**
- Be able to write and understand **SQL** queries and data definition statements
- Understand **relational algebra** and its connection to SQL
- Understand how to **write programs that access a database server**
- Understand the **ER model** used in database design
- Understand **normalization** of database schemata
- Be able to **create a database design** from a requirement analysis for a specific domain
- Know basic **index structures** and understand their importance
- Have a basic understanding of relational database concepts such as **concurrency control, recovery, query processing, and access control**



Course Project

- Forming groups
 - Your responsibility!
 - Inform me + TA
 - Deadline: TBA
- Oracle Server Accounts
- Git repositories
 - Create an account on Bitbucket.org (<https://bitbucket.org/>)
 - We will create a repository for each student
 - Use it to exchange code with your fellow group members
 - The project has to be submitted via the group repository
- Timeline:
 - Brainstorming on application (by Sep 11th)
 - Design database model (by Nov 12th)
 - Derive relational model (by Nov 25th)
 - Implement application (by end of the semester)



Fraud and Late Assignments

- All work has to be original!
 - Cheating = 0 points for assignment/exam
 - Possibly E in course and further administrative sanctions
 - Every dishonesty will be reported to office of academic honesty
- Late policy:
 - -20% per day
 - No exceptions!
- Course projects:
 - Every student has to contribute in **every** phase of the project!
 - **Don't let others freeload on you hard work!**
 - ▶ Inform me or TA immediately



Reading and Prerequisites

- **Textbook:** Silberschatz, Korth and Sudarshan
 - ***Database System Concepts, 6th edition***
 - McGraw Hill
 - publication date:2006,
 - ISBN 0-13-0-13-142938-8.
- Prerequisites:
 - CS 331 or CS401 or CS403



Outline

- Introduction
- Relational Data Model
- Formal Relational Languages (relational algebra)
- SQL
- Database Design
- Transaction Processing, Recovery, and Concurrency Control
- Storage and File Structures
- Indexing and Hashing
- Query Processing and Optimization