CS 838: Foundations of Data Management

Spring 2016
INTRODUCTION

• undergrad in Athens, Greece
• Ph.D. in University of Washington (the other UW)
• at UW-Madison since 2015!

Research Interests
• parallel processing of big data
• data pricing
• uncertainty in data management
COURSE LOGISTICS
COURSE FORMAT

• Lectures **T+Th 1:00-2:15 pm (ENGR HALL 3345)**

• Office Hours: **Th 2:30-3:30pm or by appointment**


• Mailing List: compsci838-1-s16@lists.wisc.edu
COURSE STRUCTURE

The course will have two parts:

1. Query Languages + Complexity
2. Advanced Topics: provenance, privacy, uncertainty, ...

The lectures will be on the blackboard. For some lectures I will post notes on the webpage, for others we will focus on specific papers.
Not any formal prerequisite. It will be helpful if you have good knowledge of:

- Databases, SQL
- Algorithms
- Complexity
GRADING

- Class participation: 15%
- Assignments (3): 30%
- Paper reviews (~5): 15%
- Research project: 40%
Assignments

- Individual assignments
- Submitted through learnuw (use Latex!)
- You can use up to 4 late days for all 3 assignments
Paper Reviews

- Read an assigned paper before the lecture
- Submit a brief review of the paper
- Answer a question related to the content of the paper
Research Project

• In groups of 1-3 members
• Independent research on any topic related to this course
• Deliverables:
  – 2/12: groups
  – 3/4: project proposal
  – 4/3: milestone
  – 5/3+5/5: project presentations (10% of grade)
  – 5/8: final report
What is This Class About?
What is This Class about?

- Data is everywhere!
- Managing data is critical:
  - scientific discoveries
  - online services (social networks, online retailers)
  - decision making
- **Databases** are the core technology
- In this class:
  - Foundations of databases
CLASSIC DATABASE THEORY

- Conjunctive Queries
- Query containment/equivalence
- Query complexity
  - how fast can we evaluate a join?
  - how big can the result of a join be?
  - are some join queries easier to compute than others?
DATALOG

• Declarative Language that allows us to express larger classes of queries!
PARALLEL QUERY EVALUATION

• How do we evaluate queries in parallel environments?
  – MapReduce
  – Spark
  – ...

CS 564 [Fall 2015] - Paris Koutris
How do we deal with uncertain data?

• probabilistic databases
• consistent query answering
• repairs
Other Topics

- Data Streaming
- Provenance
- Privacy