



15-799 — Advanced Topics in Database Systems

Today's Agenda

- Course Logistics
- Reading Assignments
- Lightning Talks
- Programming Projects
- Course Motivation

Course Logistics

- Course Policies
 - *Refer to course web page.*
- Academic Honesty
 - *Refer to [CMU policy page](#).*
 - *If you're not sure, ask me.*

Office Hours

- Immediately after class:
 - *Mon/Wed: 1:30 – 2:30*
- Things we can talk about:
 - *Issues on implementing projects, paper clarifications/discussion, relationship advice.*

Waitlist

- Add your name to the sign-up sheet and I will add you to the course roster.

Class Structure

- Seminar Course
 - *We read papers and talk about our feelings...*

Class Structure

- 2+ papers assigned per class.
- Everybody reads them before.
- One person presents them to the class for an hour.
- Extra time for discussion, lightning talks, & projects.

Paper Reviews

- If you're not presenting, then you must turn in reviews for each paper *before* class starts.
- Late submissions will not be accepted.
- You can miss two submissions.

Paper Reviews

- One page per paper.
- Standard conf review format:
 - *Overview.*
 - *Three strong points.*
 - *Three weak points.*
 - *Technical questions for class.*

Paper Presentations

- 60 minutes.
- In depth description & analysis of the primary papers.
- May need to incorporate info from supplemental sources.

Paper Presentations

- Avoid boring us.
- If you're not sure what parts of the paper to present, ask me.
- If you borrow from other presentations, be sure to provide attribution.

Lightning Talk

- Everyone will give a **5 minute** talk about a “Big Data” system or technology.
 - *Topic is up to you.*
 - *Cannot be something that is covered in paper readings.*
 - *No demos (they never go well).*

All Presentations

- You must email me a PDF copy of your slides immediately after presenting in class.
- Be sure to include your name in the meta-data.

Projects

- Semester-long programming group project.
- Max two people per group.
- Must open source all code.

Projects

- Project Proposal (3 pages)
- Final Report (10 pages)
- End of semester presentation.
- Final source code deliverable.

Project Topics

- Must be related to “Big Data.”
- Must be significant effort by both partners.
- Ask me if you’re looking for ideas or a partner.

Good Example

- Write an application that uses a NoSQL system to discover fake accounts on Wikipedia from +300 million revision history records.

Bad Example

- Use Facebook's LinkBench benchmark suite on **MySQL** and **Postgres** and show a bunch of graphs.

Cool Story, Bro.

- It's ok to incorporate your research into this course.
- Try to pick papers and lightning talk topics related to your project.

Computing Resources

- Everyone is going to get \$100 from Amazon.
- Additional computing resources are available on a case-by-case basis.

Final Grade

- Paper Reviews + Class Participation (20%)
- Paper Presentations (25%)
- Lightning Talk (5%)
- Programming Project (50%)

CMU Database Group

- Weekly meetings start today.
- Mondays @ 4:30
- GHC 8115
- <http://db.cs.cmu.edu>

Big Data Seminar Series

- High-profile database speakers are coming to CMU.
- Thursdays @ 12:00
- <http://www.pdl.cmu.edu/SDI>

Big Data Seminar Series

- **Oct 10:** Dwight Merriman (MongoDB)
- **Oct 17:** Tyson Condie (UCLA / Microsoft)
- **Oct 24:** Jonathan Ellis (Cassandra)
- **Nov 14:** Dmitriy Ryaboy (Twitter)
- **Dec 5:** Rich Hickey (Datomic / Clojure)