

Languages & Runtimes for Big Data

Oliver Kennedy

Logistics

- Course website & forum
 - <http://odin.cse.buffalo.edu/teaching/cse-662/>
 - Disqus threads for each paper
- Grading
 - Group Project - 3 Reports (15% / 15% / 50%)
 - ~Weekly Papers & Discussion (20%)
- Office Hours
 - Oliver: Weds 1:00-3:00

Email

- Always add [CSE662] to the title of emails
 - (or use Disqus)
- This will ensure a faster reply as we will prioritize class related emails
 - This tag is mandatory for assignments
 - Emails should be sent to BOTH Oliver and Luke

Academic Integrity

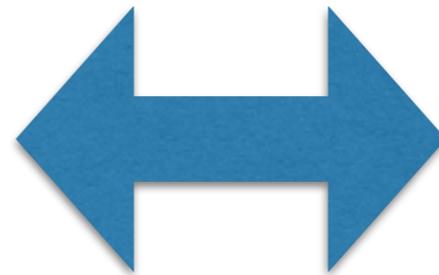
- All homework must be done by yourself
 - You may ask your classmates questions, but you must acknowledge who you talked to in your submissions
- Each group will have a separate project
 - you are free to help each other out, but you must acknowledge who you talked to in your submission

DB ~ PL

- Indexes
- Transactions & Logging
- Incremental View Maintenance
- Query Rewriting & Performance Prediction
- Probabilistic Databases
- Data Structures
- Concurrency & STM
- Self-Adapting Computation
- Compiler Optimization & Program Analysis
- Probabilistic Programming

DB ~ PL

Data-Centric
Programs



Turing Complete
Programs

Course Schedule

- Data Structures, Indexes, Adaptive Indexing
- Coping with Data Uncertainty
- Transactions & Synchrony
- High Throughput Data Processing

Course Structure

Monday

Wednesday

Friday

Classical Lecture (Paper of the Week)	Group Presentations / Meetings
--	--------------------------------

Group Presentations and Q&A

- Everyone should attend
- Present design choices, developed algorithms, background information, code, performance metrics and analysis
- Defend ideas and design choices in a public setting
- Discuss work in progress

Grade Break Down

Final Project 50%

Class Participation and Homework 20%

Project Checkpoint 1 15%

Project checkpoint 2 15%

Homework Grading

- 3 point System

0 points – nothing turned in / poorly done assignment

2 points – correctly completed assignment

1 point – everything else

Suggested Projects

- Query Processing
 - Sampling-Based Query Evaluation
 - Mimir on SparkSQL
- Data Quality
 - Deferring Manual Constraint Repair
 - Explaining Outliers
- Indexing
 - Adaptive Multidimensional Indexing
 - Data “Branching”
- Pocket-Scale Data
 - Garbage Collection in Embedded Databases

Homework Assignment 1

- Reading and Response to “Database Cracking”
- Due 9/1/2017 at 11:59pm

In-Class Assignment

- Form a group of 4 as a project group for the duration of the semester
- Come up with a clever group name
- Challenge: form a group with people you do not know or do not know well

Class Introductions

What is your name?

What did you do over the summer?

Why did you pick this class?

Favorite Editor (Emacs, Vim, Atom, Eclipse, Sublime, ...)?