Languages & Runtimes for Big Data

Lukasz Ziarek & Oliver Kennedy

Logistics

- Course website & forum
  - http://odin.cse.buffalo.edu/teaching/cse-662/
  - https://piazza.com/buffalo/fall2016/cse662

- Grading
  - Group Project - 3 Reports (15% / 15% / 50%)
  - ~Weekly Papers & Discussion (20%)

- Office Hours
  - Luke: Mon 1:00 - 2:00 additional TBD
  - Oliver: Weds 1:00-2:30

Email

- Always add [CSE662] to the title of emails, or post on Piazza
- 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

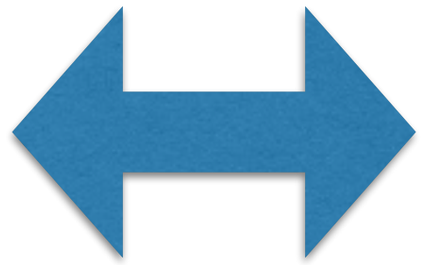
DB ~ PL

- 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
- Emerging Workload Challenges
- Probabilistic Languages & Data
- Transactions & Synchrony
- Incremental Computation
- Program Analysis & Optimization (Time Permitting)

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

- Just-in-Time Data Structures
  - Policies for Specific Workloads
- Pocket-Data
  - Get to Know an App
- Data Cleaning & Exploration in Mimir
  - Natural Language Queries
  - Performance through User-Defined Functions

Homework Assignment 1

- Reading and Response to “Database Cracking”

- Due 9/4/2015 at 11:59pm

In-Class Assignment

- Form a group of 4 as a project group for the duration of the semester
- Come up will 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?

Star Wars, Star Trek, Bab 5, BSG, none?