Data Centric Systems and Networking (DCSN)

Session 1: Introduction to R212

Eiko Yoneki

Systems Research Group
University of Cambridge Computer Laboratory

My Trajectory
My Research Interests

- Spanning over Distributed Systems, Networking and Database
- Current Focus: Large-Scale Graph Processing
- MPhil project Suggestions
  http://www.cl.cam.ac.uk/~ey204/teaching/Projects/2014_2015

My Group: Data-Centric Systems

**Digital Epidemiology**
- Real world mobility data collection in Africa
- Analyse network structure to understand infectious disease spread
  - Multiple modes of spread in time

**Content Distribution Networks**
- Build self-adaptive CDN to understand behaviour in content networks
  - Use cognitive science (e.g. EEG, Eye Tracking)
- Enhanced content distribution with social diffusion information

**Graph Specific Data Parallel**
- Fast, flexible, and programmable graph processing
- Cost effective but efficient storage
  - Move to SSDs from RAM
- Reduce latency
  - Runtime prefetching
  - Graph algorithm specific runtime
  - Dynamic CPU/GPU scheduling
- Reduce storage requirements
  - Compressed adjacency lists
- Build efficient data analytic framework without huge computing resources
  - Search/update real time (Graph DB)
Introduction to R212

- Welcome to R212
- First introduce yourselves
  - Tell about yourself
    - Your name and where you studied before ACS
    - What is your research interests (topics)
    - What is potential your ACS project
    - Why are you interested in R212

R212 Course Objectives

- Understand key concepts of data centric approaches
- Understand how to build distributed systems in data driven approach
- Research skills
  - Establish basic research domain knowledge in data centric systems
  - Obtain your view of research area for thinking forward
Course Structure

- Reading Club
  - ~3 or 4 Paper review presentations and discussion per session (~=20 minutes presentation + discussion)
  - Each of you will present about 2 reviews during the course
    - Revised (if necessary) presentation slides needs to be emailed on the following day
  - Review_Log: minimum 1 per session
    - Email me by noon on Sunday before
    - Prepare a couple of questions
  - Active participation to review discussion!

Review_Log

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Name and (crsid):</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Paper Title and Authors</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1. Paper Summary (&lt;100 words)</td>
</tr>
<tr>
<td>Describe a brief summary (extract essentials).</td>
</tr>
<tr>
<td>2. List other papers you read or skimmed for understanding this paper</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3. Punch-line of the Paper (&lt;250 words):</td>
</tr>
<tr>
<td>What is the significant contribution?</td>
</tr>
<tr>
<td>What is the difference from the existing work?</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4. What didn't you understand?</td>
</tr>
<tr>
<td>Crystallise what you did not get from the paper and describe your potential questions to the presentation/discussion</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>5. Any major criticism to the authors?</td>
</tr>
<tr>
<td>Any criticism and suggestions to the authors?</td>
</tr>
</tbody>
</table>
**Review_Log**

1. **Paper summary (<100 words)**
   - Describe a brief summary
   - Aim: you have read and extracted essentials

2. **List other papers you read or skimmed**

3. **Punch-line of the Paper (<250 words)**
   - What is the significant contribution?
   - What is the difference from the existing works?
   - What is the novel idea?
   - What is required to complete the work?

4. **What didn’t you understand? (<100 words)**
   - Crystallise what you did not get from the paper and describe your potential questions to the presentation/discussion

5. **Any major criticism to the authors?**

---

**Course Work: Reports 1&2**

- **Review report** on full length of paper (1800 words ~3 pages)
  - Describe the contribution of paper in depth with criticism
  - Crystallise the significant novelty in contrast to the other related work
  - Suggestion for future work

- **Survey report** on sub-topic in data centric networking (<2000 words)
  - Pick up to 5 papers as core papers in your survey scope
  - Read them and expand your reading through related work
  - Comprehend your view and finish as your survey paper

- **Hand in reports**
  - Report 1: November 14 noon
  - Report 2: November 28 noon
  - No particular order
Study of Open Source Project

- Open Source project normally comes with new proposal of system/networking architecture
- Understand the prototype of proposed architecture, algorithms, and systems through running an actual prototype
- Any additional work
  - Writing applications
  - Extending prototype to another platform
  - Benchmarking using online large dataset
- Present/explain how prototype runs
- Some projects are rather large and may require extensive environment and time; make sure you are able to complete this assignment

Course Work: Reports 3

- Report on project study and exploration of a prototype (<2500 words)
  - Project selection by October 31, 2014
  - Title and brief description (100 words) by email
  - Project presentation on December 1, 2014
  - Final report on the project study on December 19, 2014
Candidates of Open Source Project


- List is not exhausted and discuss with me if you find more interesting one for you
- Expectation of workload on open source project study is about intensive 3 full days work except writing up report
- One approach: pick one in the session topic, which you are interested in along your survey report
- Apache Giraph, Naiad, Spark, GraphLab, CIEL...

Important Dates

- October 31 (Friday)
  - Project selection
- November 14 (Friday)
  - Review report or Survey report
- November 28 (Friday)
  - Review report or Survey report
- December 19 (Friday)
  - Open source project study report
Assessment

- The final grade for the course will be provided as a letter grade or percentage and the assessment will consist of two parts:

- 20%: for a reading club (presentation, participation and review_log)

- 80%: for the three reports
  - 20%: Intensive review report
  - 25%: Survey report
  - 35%: Project study

How to Read a Paper?
**How to Read a Paper?**

- Scope of DCSN is wide

- …includes distributed systems, OS, networking, programming language, database…

- Type of papers
  - Building a real system
  - Proposing algorithm/logic on architecture design
  - New idea

---

**Critical Thinking**

- Reading a research paper is not like reading a text book

- But the most important one is that the paper is not necessary the *truth*
  - there is no right and wrong, just good and bad
  - There are inherently subjective qualities…but you can’t get away with just your opinion: must argue

- Critical thinking is the skill of marrying subjective and objective judgment of a piece of work

S. Hand’10
**First Let’s Argue for...**

- What is the problem?
- What is important?

- Why isn’t it solved in previous work?
  - Why graph specific parallel processing? MapReduce is not good enough?

- What is the approach?
  - Graph specific MapReduce

- Why is this novel/innovative?
  - Iterative operation for graph parallel

---

**And Now against...**

- Problem is overstated (or oversold)
- Problem does not exist
- Approach is broken
  - It does not work for all the algorithms...
- Solution is insufficient
  - Only works when data is in memory...
- Evaluation is unfair/biased
  - Use HPC for experiment
So Which is RIGHT Answer?

- There isn’t one!
  - Most of arguments are mostly correct...

- Your judge on what is valuable on topic

- In this course, we’ll be reviewing a selection of ~15 papers (3-4 per week)
  - All of these papers were peer-reviewed and published
  - However you can pick your opinion on papers!

Reviewing Tips & Tricks

- Identify a core/major idea of the topic

- Read related work and/or background section and read key other papers on the topic

- Capture the author’s claim of contribution in introduction section and judge if it is delivered

- Understand the methodology that demonstrates paper’s approach

- Capture what authors evaluate and judge if that is a good way to evaluate the proposed idea

- For theory/algorithm paper, capture what it produces as a result (rather than how)
Key in Review Comments

- What do YOU think?
  - Where you finally get to explain your opinion!
  - You should aim to give a judgement on the work
  - Your judgement should be backed by your argument

- Questions for the authors

How to Review a Paper Aid...


- Simon Peyton-Jones: How to write a great paper and give a great talk about it, Microsoft Research Cambridge.

- David A. Patterson: How to Have a Bad Career in Research/Academia, 2001.

See course web page for the paper links.
Structure of Presentation

- Cover 3 things in your presentation

1. Background/context
   - What motivated the authors?
   - What else was going on in the research community?
   - How have things changed since?

2. What is problem to be tackled?
   - What is the problem they tried to solve?
   - What are the key ideas?
   - What did the authors actually do?
   - What were the results?

3. Your opinion of the paper
   - What you agree and what you disagree?
   - What is the strength and weakness of their approach?
   - What are the key takeaway?
   - What was the impact (possible impact)?

Preparing...

- Not too much basics: remember, others will have read the paper
  - Brief overview
  - Do not make exact repeat of the paper

- Aim: generate discussion – spit your straight opinion about the paper to stir the discussion
  - Explore the arguments they make and the conclusions they draw. What is your opinion on it?
  - When you argue, state clearly the point of argument
Presenting...

- Practice beforehand to ensure length of your presentation

- Getting nervous is normal!
  - We are in the same boat and we help each other to understand the paper
  - Presentation is a tool to provide a discussion forum

- Try not to get defensive or angry at questions
  - It is not your paper!

Listening Presentation...

- You need to get involved

- Ask questions from your review – bring your review_log copy

- Always be respectful of the speaker
How to Write Reviews (Report 1)

- **Paper Summary**
  - Provide a brief summary of the paper
  - At this stage you should try to be objective
- **Problem**
  - What is the problem? Why is it important? Why is previous work insufficient?
- **Solution or Approach**
  - What is their approach?
  - How does it solve the problem?
  - How is the solution unique and/or innovative?
  - What are the details?
- **Evaluation is unfair/biased**
  - How do they evaluate their solution?
  - What questions do they answer?
  - What are the strength/weakness of the system and evaluation itself?

How to write Survey paper (Report 2)

- Demonstrate a summary of recent research results in a novel way that integrates and adds understanding to work in the research area
- Must expose relevant details associated, but it is important to keep a consistent level of details and to avoid simply listing the different works
- For example:
  - Define the scope of your survey
  - Classify and organize the trend
  - Critical evaluation of approaches (pros/cons)
  - Add your analysis or explanation (e.g. table, figure)
  - Add reference and pointer to further in-depth information
Summary

- R212 course web page:

  http://www.cl.cam.ac.uk/~ey204/teaching/ACS/R212_2014_2015
  
  Email: eiko.yoneki@cl.cam.ac.uk

- Slides of presentation, forms, other information will be on the web

Topic Areas

Session 1: Introduction

Session 2: Programming in Data Centric Environment

Session 3: Processing Models of Large-Scale Graph Data

Session 4: Map/Reduce Hands-on Tutorial with EC2

Session 5: Optimisation in Graph Data Processing + Guest lecture

Session 6: Stream Data Processing + Guest lecture

Session 7: Scheduling Irregular Tasks

Session 8: Project study presentation