# Lecture 12: Final Review

CSCI4180 (Fall 2013)

Patrick P. C. Lee

# Cloud Computing Basics

- Key characteristics
- Service models: SaaS, PaaS, IaaS
- Public/private clouds
- Security issues
- **Try to remember all terminologies being introduced**

# MapReduce and HDFS

- Design features of MapReduce (e.g., map, reduce, combiner, partitioner, jobtracker, tasktracker)
- How does MapReduce handle fault tolerance?
- Design features of HDFS (e.g., namenode, datanode)?
- How does HDFS handle fault tolerance?

# MapReduce Programming

- MapReduce programming basics
  - Execution flow
- MapReduce algorithms (design patterns)
  - In-mapper combining, pairs/stripes, order inversion, value-to-key conversion
  - Can you write pseudo-code for a simple MapReduce program?
- Inverted index
- Shortest-path algorithms

# Key-Value Stores

- BigTable, HBase
  - Main design features
- How Amazon Dynamo balances performance and availability?

# Cloud Storage

- Cost model
- Deduplication
  - Fingerprinting
  - Chunking (Rabin Fingerprint)
  - Indexing
- Case studies
  - Dropbox: security implications
  - Facebook: minimizing I/Os in photo search

# Miscellaneous

- Zookeeper
  - Coordination mechanism for a distributed system
- Virtualization
  - Remember all terminologies we taught

# Final Exam

- 2-hour written exam
- Cover all lecture notes, tutorials, assignments
- Open notes, open books, open hardcopies
  - No notebooks, smartphones
- Approved calculators allowed

- Make assumptions if needed, and provide justifications

# Final Exam

- How to prepare?
  - Understand everything in class notes and assignments
  - Not required to read all readings, so long as you understand what the concepts mean
- Do past exams
  - http://library.cuhk.edu.hk/
  - Ignore questions that we didn't cover
- Ask via facebook, or find me in office (Dec 9 after 2pm)
  - Please don't ask me about solutions of past exams