Exam 2 in general

- You can bring a cheat sheet, 1 piece of printing-sized paper, double sided is fine
- No electronics, including calculator, watch, phone, etc., during the exam
- Exam 2 will only include materials we covered after Exam 1. But if some of these materials require background knowledge of those we covered before Exam 1, you still need to know.
- Exam 2 will tentatively last 2 hours, but the time may be updated during the exam.
- Exam 2 will take place in our classroom.
- After all exam papers are collected, if there is still time, TA will discuss the answers in the remaining time of the class.

Key points for Exam 2

- **MapReduce**
  - Able to write pseudo programs in MapReduce
  - How the Hadoop system supports the three phases of MapReduce?
  - There are improved versions of MapReduce platforms, from Hadoop to Spark. What features are improved? What each platform did to achieve the targeted improvements?
  - Special features in Spark
  - Some specifics in Storm

- **Virtualization**
  - Different CPU virtualization techniques and other techniques that are similar to but different from virtualization
    - You need to know what they are and what are their differences
    - You need to know what to use for a given scenario
  - CPU scheduling in virtualization
    - The impact of different types of clocks
  - Memory virtualization techniques
    - Some bullets were given on the whiteboard at the end of the virtualization topic
    - Basic techniques
    - Being able to do addressing and page faults
    - The works to be done by the OS and VMM for memory management
  - VM migration techniques
    - Combine VM migration, VM scaling, etc. with IaaS load balancing and consolidation techniques
    - Consider DHT based file systems for load balancing and the load balancing algorithms we covered recently

- **IaaS resource management**
  - Know the three algorithms for VM placements: Volume based, Dot Vector based, Hexagon based
  - Know other schemes we have introduced (basic greedy algorithm, error margin and p-percentile considerations, etc.)
  - The methods in VMware to monitor workloads
  - The methods for better workload prediction
  - The methods in VMware for evaluating gains due to migration in order to make proper migration decisions

- **PaaS and SaaS**
  - Multi-tenancy
    - What is SaaS and what is multi-tenancy?
    - Techniques for Database, UI, Business logic, Access control to support multi-tenancy

- **Erasure code**
  - Understand the tradeoffs between Erasure code and Replication techniques
  - For the Even-Odd code
- How to construct the code
- How to recover from failures (exclude one-data + horizontal-parity failures, which we did not cover), pay attention to the adjustment bit

For the Microsoft Azure Reed-Solomon code scheme
- 12 nodes, divided into two groups, each with 6 nodes
- 2 parities, one for each group
- 2 nodes for Reed Solomon codes

- How to recover from different failure patterns
  - Failures within a group, failures cross multiple groups