

# 15-619

## Cloud Computing

### 15-619 Project Overview

Recitation 11

Apr 1<sup>st</sup> and Apr 3<sup>rd</sup>, 2014

# Some feedback on Phase 1

- ONLY **m1.small**, **m1.medium** and **m1.large** are allowed for your **front end** and **back end** system.
- For ETL you can use other instance types
- Always keep cost in mind

# Briefing on phase 2

- Phase 2 deadline is **11:59 PM EST, Apr 8**
- 4 teams have submitted testing requests for all 3 queries.
- Achieving full score is not easy

# 619 project Phase 2

- What we expect you to learn
  - Working on a large dataset
    - Approaches to process data
    - Code efficiency
    - Sensitivity on cost
  - System performance tuning
  - Database schema design and performance tuning
  - Comparison between MySQL and HBase
    - Small dataset vs. large dataset

# Performance Tuning

- What can be the bottleneck?
  - Database?
  - Front end framework?
  - Network bandwidth?
  - ELB?
  - Testing service?
  - ...

# System Performance Tuning

- How to find the system bottleneck?
  - CloudWatch (CPU, Instance Disk, Network)
    - Little impact on performance
    - No extra installation
  - vmstat, iostat (CPU, Memory, Disks, Network...)
    - More information
    - Arbitrary sampling frequency
    - Memory statistics (vmstat)

# Database Performance Tuning

- What contributes to the query processing time?
  - RTT?
  - Index lookup / Hash calculation?
  - Cache / Disk read?
    - How many records need to be read?
  - Network transmission?
  - ...

# Performance Tuning

- Synthesize the information from your System and the Database
- Identify your system's bottleneck(s)
- Find a solution to remove the bottleneck
  - System architecture
  - Database schema
  - Other system/database configuration parameters



Q & A