

# 15-319 / 15-619

# Cloud Computing

Recitation 09

October 22, 2019

# Overview

- **Last week's reflection**
  - Project 3.3
  - Query 1, Team Project
  - OLI Unit 4 - Module 14 (Storage)
  - Quiz 7
- **This week's schedule**
  - OLI - Modules 15, 16 & 17
    - Quiz 8 – due on Friday, October 25<sup>th</sup>
- **Team Project, Phase 1**
  - Query 2 is due on Sunday, October 27
  - Report is due on Tuesday, October 29

# Last week: Week 8

- **Module 14: Cloud Storage**
  - Quiz 7
- **Project 3.3**
  - Multi-threading Programming and Consistency
- **Team Project, Phase 1**
  - Query 1 was due on Sunday

# This Week: Conceptual Content

## OLI UNIT 4: Cloud Storage

- **Module 15: Case Studies: Distributed File System**
  - HDFS
  - Ceph
- **Module 16: Case Studies: NoSQL Databases**
- **Module 17: Case Studies: Cloud Object Storage**
- **Quiz 8**
  - **DUE on Friday, October 25th**

# Project 3.3 Reflection

- You have explored
  - Replication and sharding
  - Multithreaded programming
  - Strong consistency model
    - Use PRECOMMIT to keep proper order on all datastores
  - Bonus Task: Eventual Consistency
    - No guarantee of ordering for incoming requests
    - Compare timestamp with last timestamp for the key

# Project 3.3 Reflection

- Most common issues:
  - Incorrect implementation of locking using a Priority Queue
  - Incorrect use of wait() and notifyAll()
  - Improper implementation of synchronization block
  - Exception in threads, which caused the threads to exit prematurely without closing the connection.
- Best ways to debug:
  - Logging to keep track of what exactly happened with a request

# TEAM PROJECT

## Twitter Data Analytics



+



=



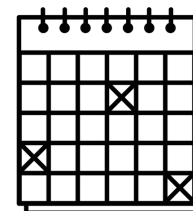
# Team Project

## Twitter Analytics Web Service

- Given ~1TB of Twitter data
- Build a performant web service to analyze tweets
- Explore web frameworks
- Explore and optimize database systems







# Team Project Timeline

Phase	Deadline ( <u>11:59PM ET</u> )
<b>Phase 1 (20%)</b> <ul style="list-style-type: none"><li>- Query 1</li><li>- Query 2</li></ul>	<ul style="list-style-type: none"><li>● Q1 CKPT (5%): Sun, 10/13</li><li>● Report1 (5%): Sun, 10/13</li><li>● Q1 FINAL (10%): Sun, 10/20</li><li>● Q2 CKPT (10%): Sun, 10/20</li><li>● <b>Q2M &amp; Q2H FINAL (50%): Sun, 10/27</b></li><li>● <b>Report2 (20%): Tue, 10/29</b></li></ul>
<b>Phase 2 (30%)</b> <ul style="list-style-type: none"><li>- Add Query 3</li></ul>	<ul style="list-style-type: none"><li>● Live Test on Sun, 11/11</li></ul>
<b>Phase 3 (50%)</b> <ul style="list-style-type: none"><li>- Managed Services</li></ul>	<ul style="list-style-type: none"><li>● Live Test on Sun, 12/02</li></ul>



# Q2 Hints

- Be careful about encoding
  - Non-english characters
  - Emojis 😄
- Optimization is time-consuming
  - Think about your schema design
  - Think about your database configuration

# Performance Tuning Tips

- To do performance tuning, you first need to identify which part of your system is the bottleneck.
  - Profile and monitor your system
    - Write an LG yourself to test your system performance
    - Use CloudWatch for resource utilization such as CPU, Network, Disk, etc.
    - Read the profiling and tuning primer

# Performance Tuning Tips

- Web Tier
  - Did you put too much computation at the Web Tier?
  - If you have multiple Web Tier servers, is the workload distributed evenly?
  - Have you optimized your code?

# Performance Tuning Tips

- Database Tier
  - Try to reduce the number of rows or the size of the data retrieved in each request.
  - Remember that Q2 is read-only.
    - You can choose schema that are specifically optimized for Q2.
  - Normalized tables versus compact tables

# Performance Tuning Tips

- Database Tier - MySQL
  - Tune the parameters
    - Check the official documentation
    - Search for MySQL performance tuning best practices
- Database Tier - HBase
  - Tune the parameters
    - compaction, region server split etc.
    - Scan can be really slow, try to avoid it when possible  
If not, try to scan as few rows as possible

# Performance Tuning Tips

- Review what we have learned in previous project modules
  - Scaling out
  - Load balancing
  - Replication and Sharding
- Ask on Piazza or go to office hours if you are stuck for too long!


# Q2 Budgets


- Your team has a total AWS budget of \$45 for Phase 1
- Your web service should not cost more than \$0.83 per hour, this includes (see write-ups for details):
  - Use EC2 on-demand instance cost
    - even if you use spot instances, we will calculate your cost using the on-demand instance price
  - EBS cost
  - ELB cost
  - We will not consider data transfer and EMR cost
- Targets:
  - Query 2 - target is for both MySQL and Hbase




# Upcoming Deadlines



- 
- Conceptual Topics: OLI (Module 15, 16, 17)  
Quiz 8 due: **Friday, 10/25/2019 11:59 PM Pittsburgh**

- 
- Team Project: Phase 1 - Query 2  
Due: **Sunday, 10/27/2019 11:59 PM Pittsburgh**

- 
- Team Project: Phase 1 - Report and Code  
Due: **Tuesday, 10/29/2019 11:59 PM Pittsburgh**