

15-319 / 15-619

Cloud Computing

Recitation 10

March 26, 2019

Overview

- **Last week's reflection**
 - Project 3.3 done!
- **Team Project, Twitter Analytics**
 - The team project phase 1 is due this week!

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



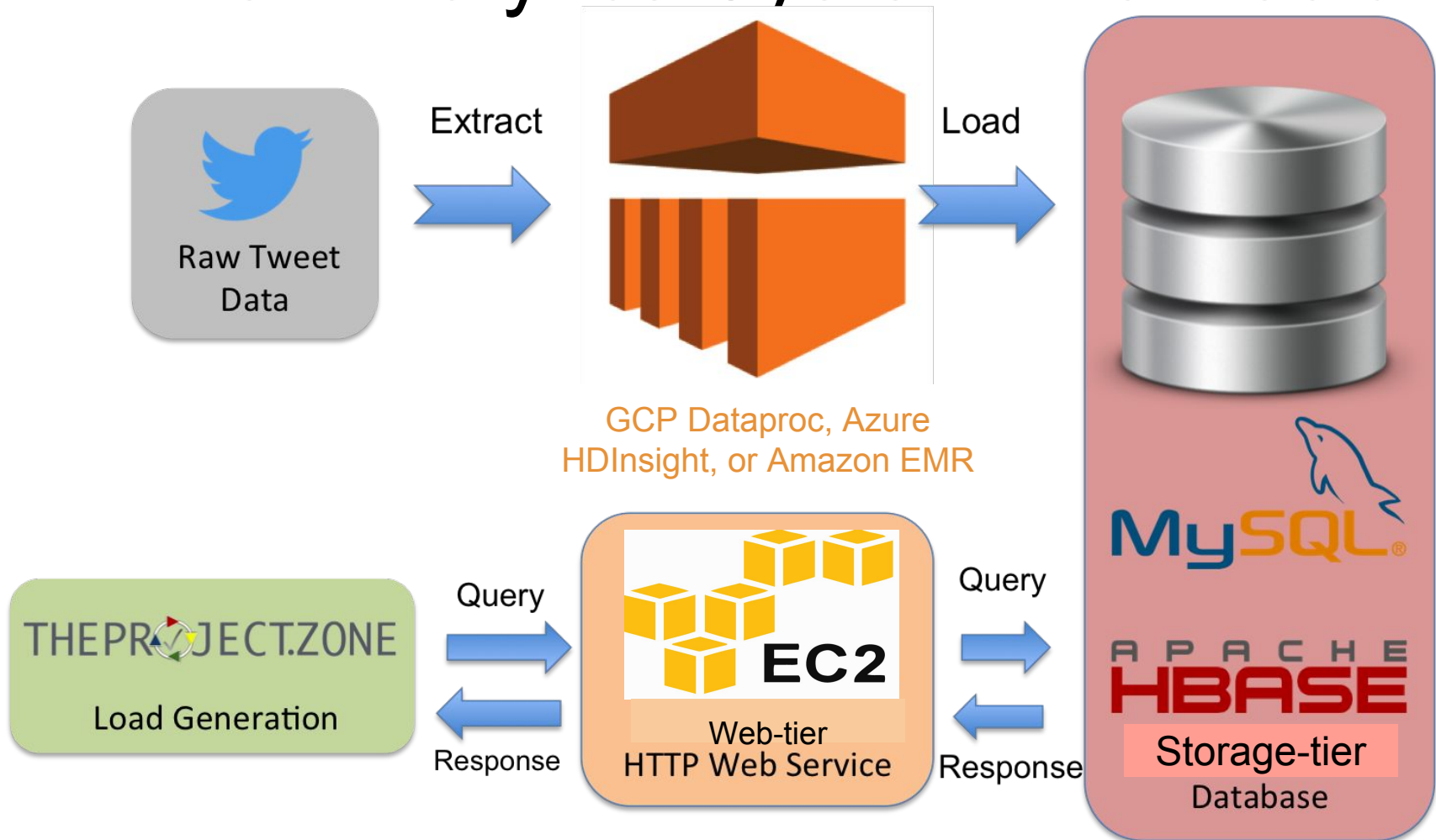
Team Project

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



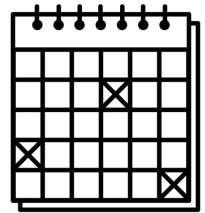
Twitter Analytics System Architecture



- Web server architectures
- Dealing with large scale real world tweet data
- HBase and MySQL optimization



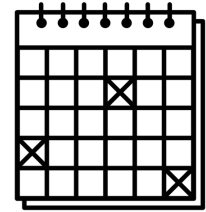
Suggested Tasks for Phase 1



Phase 1 weeks	Tasks	Deadline
Week 1 <ul style="list-style-type: none"> 2/25 	<ul style="list-style-type: none"> Team meeting Writeup Complete Q1 code & achieve correctness Q2 Schema, think about ETL 	<ul style="list-style-type: none"> Q1 Checkpoint due on 3/3 Checkpoint Report due on 3/3
Week 2 <ul style="list-style-type: none"> 3/4 	<ul style="list-style-type: none"> Q1 target reached Q2 ETL & Initial schema design completed 	<ul style="list-style-type: none"> Q1 final target due on 3/10
Week 3 <ul style="list-style-type: none"> Spring Break 	<ul style="list-style-type: none"> Take a break or make progress (up to your team) 	
Week 4 <ul style="list-style-type: none"> 3/18 	<ul style="list-style-type: none"> Achieve correctness for both Q2 MySQL, Q2 HBase & basic throughput 	<ul style="list-style-type: none"> Q2 MySQL Checkpoint due on 3/24 Q2 HBase Checkpoint due on 3/24
Week 5 <ul style="list-style-type: none"> 3/25 	<ul style="list-style-type: none"> Optimizations to achieve target throughputs for Q2 MySQL and Q2 HBase 	<ul style="list-style-type: none"> Q2 MySQL final target due on 3/31 Q2 HBase final target due on 3/31 Phase 1 report due on 4/2



Team Project Time Table



Phase (and query due)	Start	Deadlines	Code and Report Due
Phase 1 <ul style="list-style-type: none"> Q1, Q2 	Monday 02/25/2019 00:00:00 ET	Checkpoint 1, Report: Sunday 03/11/2019 23:59:59 ET Checkpoint 2, Q1: Sunday 03/25/2019 23:59:59 ET Phase 1, Q2: Sunday 03/31/2019 23:59:59 ET	Phase 1: Tuesday 04/02/2019 23:59:59 ET
Phase 2 <ul style="list-style-type: none"> Q1, Q2, Q3 	Monday 04/01/2019 00:00:00 ET	Sunday 04/14/2019 15:59:59 ET	
Phase 2 Live Test (Hbase AND MySQL) <ul style="list-style-type: none"> Q1, Q2, Q3 	Sunday 04/14/2019 17:00:00 ET	Sunday 04/14/2019 23:59:59 ET	Tuesday 04/16/2019 23:59:59 ET
Phase 3 <ul style="list-style-type: none"> Q1, Q2, Q3 	Monday 04/16/2019 00:00:00 ET	Sunday 04/28/2019 15:59:59 ET	
Phase 3 Live Test <ul style="list-style-type: none"> Q1, Q2, Q3 	Sunday 04/28/2019 17:00:00 ET	Sunday 04/28/2019 23:59:59 ET	Tuesday 04/30/2019 23:59:59 ET

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!

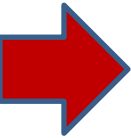
Q2 Budgets

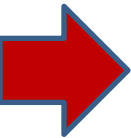
- Your team has a total AWS budget of \$45 for Phase 1
- Your web service should not cost more than \$0.83 per hour, this includes (see write-ups for details):
 - Use EC2 on-demand instance cost
 - even if you use spot instances, we will calculate your cost using the on-demand instance price
 - EBS cost
 - 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 18)
Quiz 9 due: **Friday, 03/29/2019 11:59 PM Pittsburgh**

- 
- Team Project: Phase 1 - Query 2
Due: **Sunday, 03/31/2019 11:59 PM Pittsburgh**

- 
- Team Project: Phase 1 - Report and Code
Due: **Tuesday, 04/02/2019 11:59 PM Pittsburgh**