15-319 / 15-619
Cloud Computing

Recitation 10

March 24, 2020

Overview

- **Last week's reflection**
  - Project 3.3
  - Query 2 Checkpoint, Team Project
  - OLI Unit 4 - Module 15, 16 & 17
  - Quiz 8
- **This week's schedule**
  - OLI - Modules 18
    - Quiz 9 – due on Friday, March 27
- **Team Project, Phase 1**
  - Query 2 is due on Sunday, March 29
  - Report is due on Tuesday, March 31

Project 3.3 Reflection

- You have explored
  - Sharding and replication
  - 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
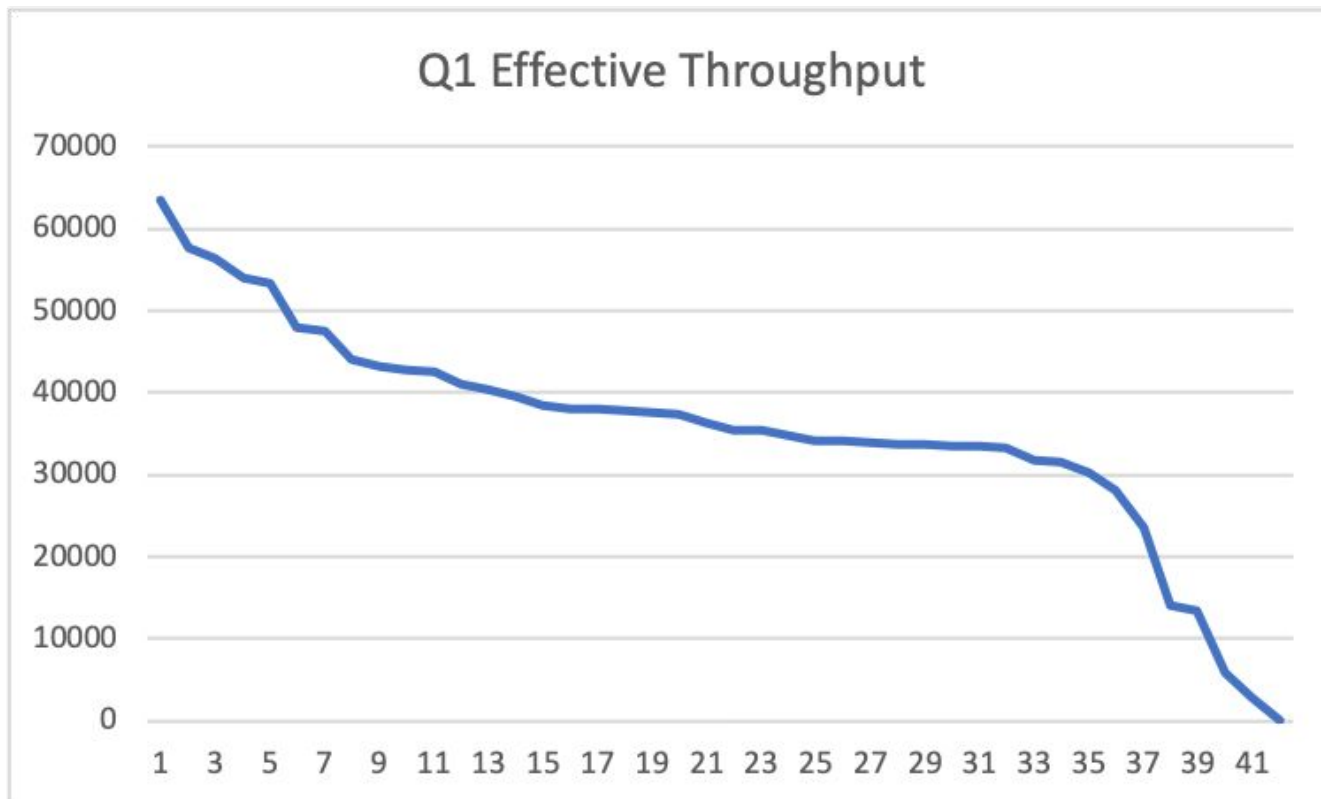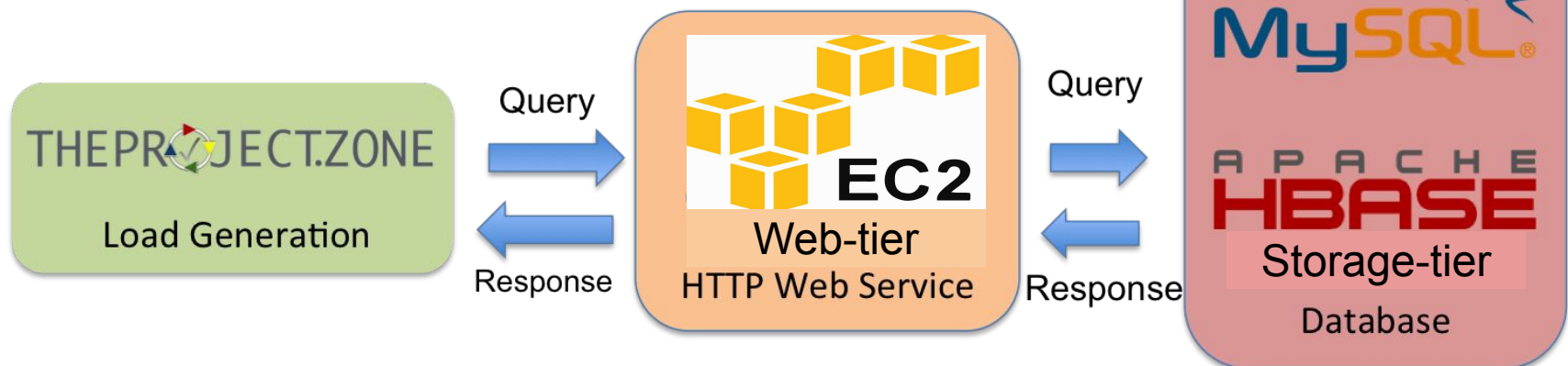
Team Project - Query 1

- 42 teams attempted a Query 1 submission.
- 41 teams made successful 10-minute submission.
- 32 teams reached 32,000 RPS.

Q1 Effective Throughput

Team Project - Query 2

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 - Query 2

Use Case: When you follow someone on twitter, recommend close friends.

Query: GET /q2?**user_id**=<ID>&**type**=<TYPE>&**phrase**=<PHRASE>&**hashtag**=<HASHTAG>

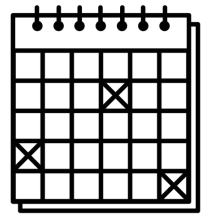**Response**:

    <TEAMNAME>,<AWSID>\n

    uid\tname\tdescription\ttweet\n

    uid\tname\tdescription\ttweet

Final Score: Interaction Score * Hashtag Score * Keywords Score

Q2 target throughput: 10,000 RPS **for both MySQL and HBase**

Team Project Timeline

Phase	Deadline (11:59PM ET)
Phase 1 (20%) - **Query 1** - **Query 2**	● Q1 CKPT (5%): Sun, 03/01 ● CKPT Report (5%): Sun, 03/11 ● Q1 FINAL (10%): Sun, 03/08 ● Q2 CKPT (10%): Mon, 03/23 ● Q2M & Q2H FINAL (50%): Sun, 03/29 ● Final Report (20%): Tue, 03/31
Phase 2 (30%) - **Add Query 3**	● Live Test on Sun, 04/11
Phase 3 (50%) - **Managed Services**	● Live Test on Sun, 04/26

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
    - Use CloudWatch for resource utilization such as CPU, Network, Disk, etc.

Performance Tuning Tips

- Web Tier
  - Why not use multithreading?
  - Did you put too much computation in the web tier?
  - Is the workload distributed evenly on multiple web servers?
  - 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

Performance Tuning Tips

- Database Tier - MySQL
  - Different MySQL engines

- Database Tier - HBase
  - Locality and compaction, region server split, etc
  - Scan can be really slow, try to avoid it if possible
    If you can't, try to scan as few rows as possible
- Tune parameters
  - Check the official documentation
  - Search for performance tuning best practices
- Ask on Piazza or go to office hours if you are stuck for too long!

Upcoming Deadlines

- Conceptual Topics: OLI (Module 18)
  Quiz 9 due: **Friday, 03/27/2020 11:59 PM Pittsburgh**

- Team Project: Phase 1 - Query 2
  Due: **Sunday, 03/27/2020 11:59 PM Pittsburgh**

- Team Project: Phase 1 - Report and Code
  Due: **Tuesday, 03/31/2020 11:59 PM Pittsburgh**