2020

CSCI 49378: Lecture 5: Distributed Web-based Applications

Bonan Liu
CUNY Hunter College

NYC Tech-in-Residence Corps

Distributed Web-based Applications

Bonan Liu
Tech-In-Residence Member, Hunter College, CUNY
Spring 2020

Disclaimer

The content of this presentation is being provided for educational and informational purposes only. The views, thoughts, and opinions expressed in this presentation belong solely to the author, and not necessarily to the author's employer.

The content of this presentation is not endorsed by the author's employer.

Agenda

- Single-box web-based applications
- Distributed web-based applications
  - Performance
  - Scalability
  - Extensibility
- CSCI 49378 Distributed Systems Module Recap
- Review Assignment 1

Single-box web-based applications

Generally, a web-based application is a program that is accessible through networks protocols. It is also called as web app as web service.

Basic components:

- program (accessible via an address and a port)
- files (unstructured data)
- databases (structured data)

Single-box web-based applications

Simple web-based application (MVC pattern):

- Model
- View
- Controller

Distributed web-based applications

Common techniques to grow single-box applications to distributed applications:

- Layered Architecture
- Splitting the business logic
- Sharding the data
- Distributed Infrastructure
- Caching
- Asynchronization and Batching
- Redundancy

Distributed web-based applications

Optimize the web applications for Performance:

- Front-end Optimization:
  - Reduce the number of requests
  - CDN: Content Distribution Network
- Cache
  - Cache Warmup
  - Cache Inconsistency
  - Cache Null Value

Distributed web-based applications

Optimize the web applications for Performance:

- Asynchronization
  - An example: workflow systems
- Batching
  - Chatting apps, movie/exam distribution, etc

Distributed web-based applications

Optimize the web applications for Scalability:

- Add load balancing component
  - Round Robin
  - Weighted Round Robin
  - Random
  - Least Connections
  - Source Hashing
- Distributed cache service: Memcached

Distributed web-based applications

Optimize the web applications for Extensibility:

- Decoupling: Distributed Messaging Queue
  - Event-driven development
- Decoupling: Data pipeline
- Decoupling: Split business logic

Distributed System Recap

- Distributed Systems
- Concepts/Techniques
  - Sharding
  - Primary/Replica
  - Master/Worker
  - Two-Phase Commit
- Synchronization
- Replication
- Consistency

Distributed System Recap

- Distributed File Systems
  - Distributed File Systems
  - Distributed Key-Value Systems
  - Distributed Databases
- Distributed Web-Based Applications
  - Performance
  - Scalability
  - Extensibility

Readings

TAO: Facebook's Distributed Data Store for the Social Graph.

https://research.fb.com/publications/tao-facebooks-distributed-data-store-for-the-social-graph-2/