Concepts of Distributed Systems

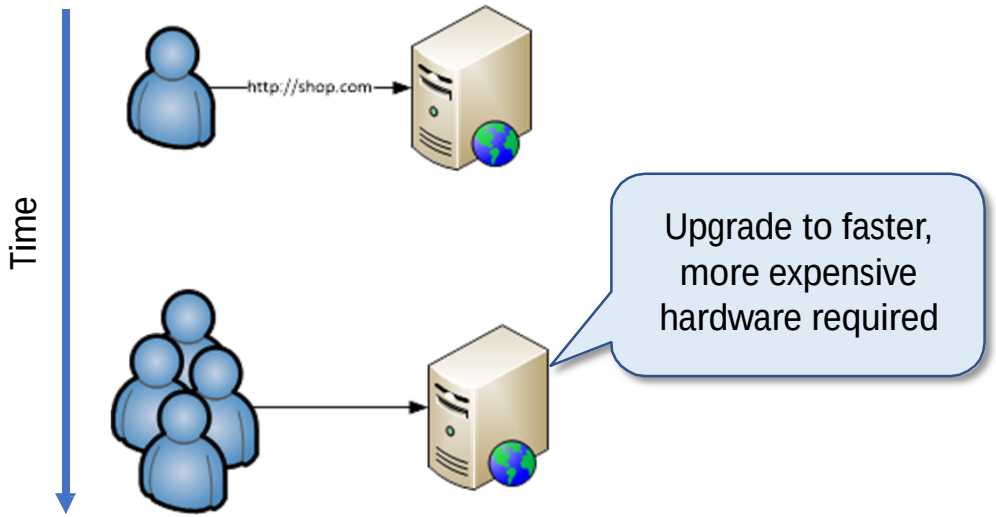
Agenda

- Evolution of computer system's architecture
  - Centralized system
  - Distributed system
- Definition of distributed computing and its challenges
- Describe use-case of course application

Centralized System

Definition

Centralized systems leverage client-server architecture, where multiple clients are directly connected to a single server, which handles all their requests.



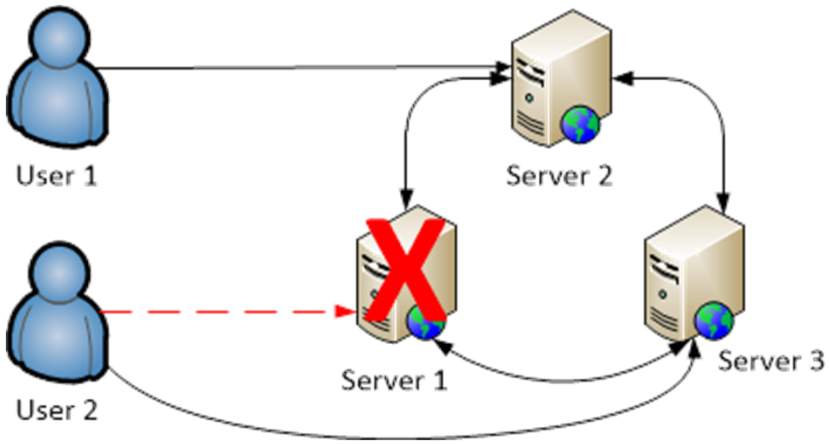
Disadvantages:

- Performance improvement limited to vertical scalability
- Single point of failure
- Hard to achieve 99,999% availability

Distributed System

Definition

Distributed system is a computing environment, where multiple processes running on different machines, communicate through the network and coordinate actions in order to appear to the end-user as a single coherent system.



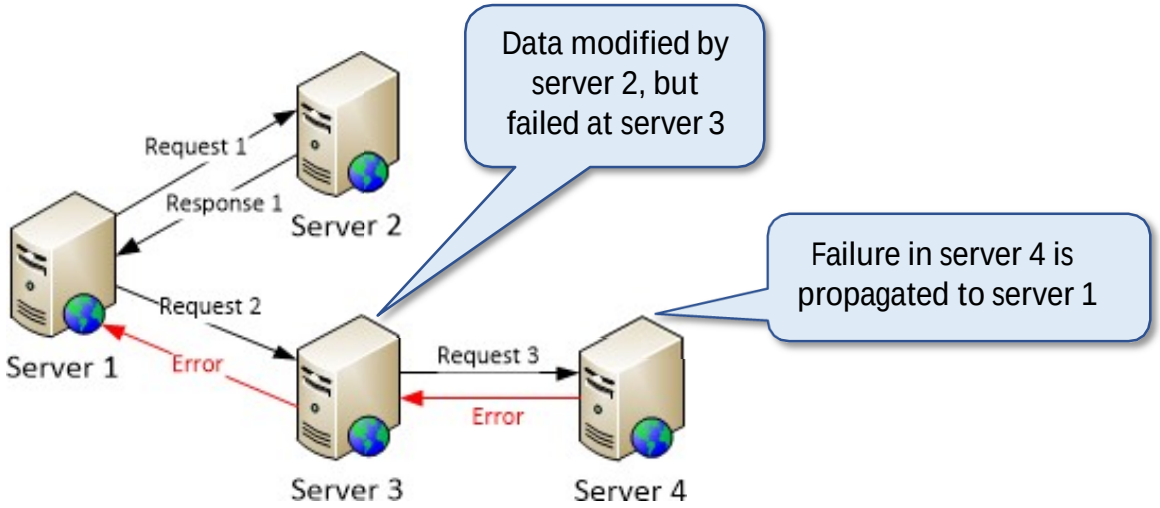
Advantages (when “done right”):

- Infinite performance due to horizontal scalability
- Resilient and fault-tolerant
- 24/7/365 availability

Resilient means able to withstand or recover quickly from difficult conditions.

Distributed System's Challenges

- Failure handling
  - Any remote procedure call can fail at runtime for different reasons
  - Failure propagation from distant parts of the system
  - Partial failure vs. data consistency



Local function call returns response or exception. Remote function call can timeout

```
sayHello("Amit");

private String sayHello(String name) {  return
        String.format(
                "Greetings      %s!", name
        );
}
```

Distributed System's Challenges

- Concurrent resource access
  - Synchronize local threads and remote processes
  - Inconsistent state across servers

- Development and bug reproduction
  - Time as a variable factor
  - Number of software and hardware components involved

"You have to design distributed systems with the expectation of failure."

K. Arnold

Task

Design and implement URL shorten service.

https://github.com/apache/kafka/blob/trunk/raft/README.md → https://tiny.com/qglj

https://tiny.com/qglj → https://github.com/apache/kafka/blob/trunk/raft/README.md

$1234567_{base\ 10} = qglj_{base36}$

$26 * 36^3 + 16 * 36^2 + 21 * 36^1 + 19 * 36^0 = 1234567_{10}$

0, 1, 2, ... , 9, A, B, C, ... Z => 36 Digits, so base = 36

Single-Threaded vs. Multi-Threaded vs. Distributed System

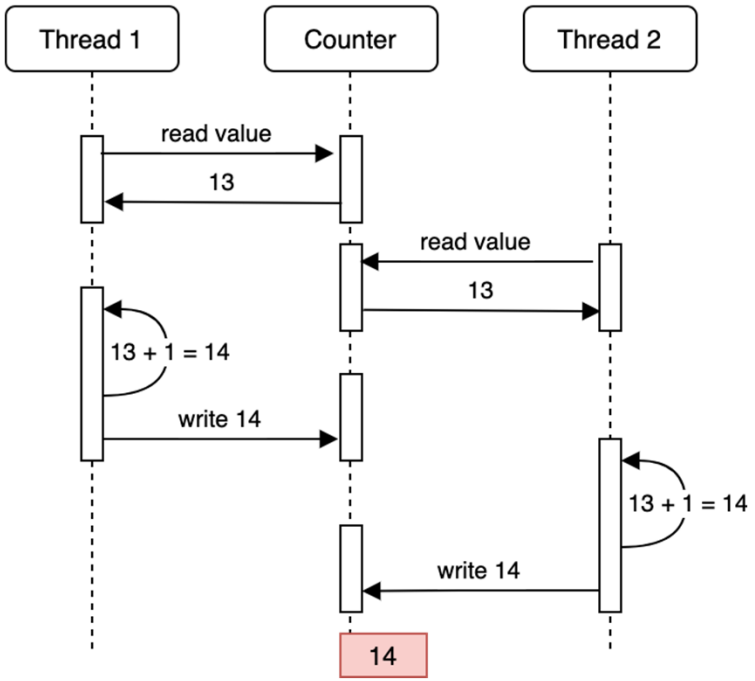
```
private long lastGeneratedId = 0;

public String shortenUrl(String longUrl) {
        long id = generateNextId();  String
        shortUrl = String.format(
                "http://tiny.com:8080/ s",   %
                Long.toString(id, 36));
        insertMapping(shortUrl, longUrl);  return
        shortUrl;
}

private synchronized long generateNextId() {
        long id = ++lastGeneratedId;
        return id;
}
```

Generate next ID and format short URL

Insert mapping to database and return result



Scale up:

- Single-threaded → Multi-threaded
- Multi-threaded → Distributed System

Summary

- Disadvantages of Centralized System
- Definition of distributed computing
- Challenges of Distributed System

"You know you have [a distributed system] when the crash of a computer you've never heard of stops you from getting any work done."

L. Lamport