Scalable Distributed Computing - Appunti

Francesco Lorenzoni

September 2024

Contents

Part I

Introduction to SDC

Chapter 1

Basic Concepts

1.1   Introduction

Definition 1.1 (Distributed) *Spreading tasks and resources across multiple machines or locations*

Example 1.1.1   ◇ *Google Search*
⋄ *Facebook*
⋄ *Amazon*

Definition 1.2 (Scalable) *Ability to grow and handle increasing workload without compressing performance*

1.2   Motivation for Scalable Systems

⋄ **Growing data** - large datasets from applications like social media, IoT, AI, etc.
⋄ **Global Users** - Billions of users worldwide
⋄ **Performance** - Reducing latency, increasing throughput, and improving reliability
  Sometimes latency is *not* a priority: in some systems it is okay to have high latency to guarantee high throughput and reliability

The challenges are mostly to **manage resources** across geographically distributed systems, and ensuring **low latency** and **high availability**.

1.2.1   Target Architectures

Some architectures which require scalable distributed systems are IoT networks, High-Performance Computing (HPC), and Cloud/Edge Computing.

Example - Cameras in a district

What if I send all the data gathered from cameras to a *single cloud*?

Pros
⋄ Unlimited storage and processing power
⋄ Centralized management
⋄ Simplicity

Cons
⋄ High latency
⋄ High bandwidth usage
⋄ Single point of failure (Scalability and Reliability)

But also simpler applications may considerably benefit from scalable and distributed architectures.
⋄ Large graph analysis
⋄ Stream processing
⋄ Streaming services
⋄ Machine Learning
⋄ Big Data
⋄ Computational Fluid Dynamics
⋄ Web and online services

1.2.2 Distribution is cool, but. . .

Consider that local computation is always faster than remote computation. (*Waaay faster*)
From the CPU perspective, time passes *very slowly* when the data travels outside the machine.

If one CPU cycle happened every second, sending a packet in a data center would take 20 hours. Sending it from NY to San Francisco would take 7 years.

1.3 Assessment Method - Exam

There are three options, but note that **in every case an oral exam will follow.**

1. *Writing a Survey or a Report*
2. *Individual or Group Project* (leq3 members) Designing, implementing and presenting a solution or prototype related to scalable distributed computing.
3. *Traditional Written Exam* "Questions, answers. . . you know the drill." Very sad option, in my opinion, but prof. Dazzi did not completely discourage it.

Prof. Dazzi is very open to proposals for the exam, he'd like to stimulate our creativity and curiosity.

Prof. Dazzi says that usually its oral examinations last from 30 to 35 minutes, even though there may be exceptions. Clearly, if the student chooses the report or the project, part of the oral will be about the proposed work, but also questions about the course will be asked.