UNIVERSITEIT VAN AMSTERDAM

Large Systems:

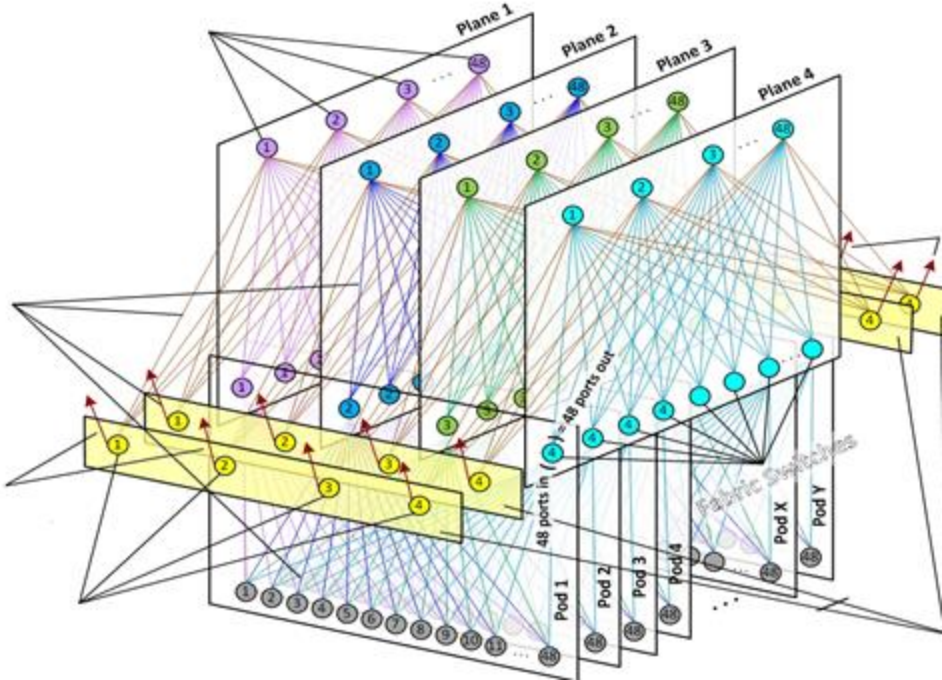
Design + Implementation + Administration

2024-2025

- Admin Intro

Shashikant Ilager
shashikantilager.com

28 oktober 2024



Teaching Staff

- Lecturer: Shashikant (Shashi) Ilager
- Consulting Time:
  - After the formal lecture, I will be here for 10 minutes for any of you to ask to clarify discuss or talk to me.
  - Forum (Lecture-Discussion on Canvas)
  - Course Email: ls2024@list.uva.nl
  - Email: s.s.ilager@uva.nl
  - For any further communication, I am available at L5.26 (Lab 42)

Teaching Assistants (TAs)

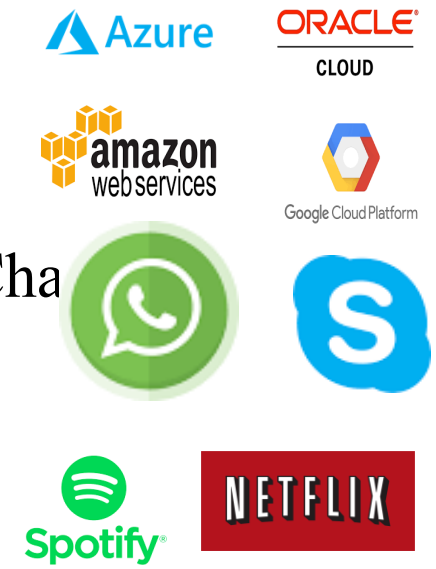
Handles lab sessions and assists with your assignments

- TA1: Paul Daniëlse , p.e.g.danielse@uva.nl
- TA2: Koen Greuell, koen.greuell@lifewatch.eu
- TA3: Maurice Mouw, maurice.mouw@sue.nl

- Consulting Time:
  - TAs will be available during lab sessions for direct interactions. For further communication use:
    - Lab Discussion Forum (Canvas):
    - Mailing list: ls2024@list.uva.nl

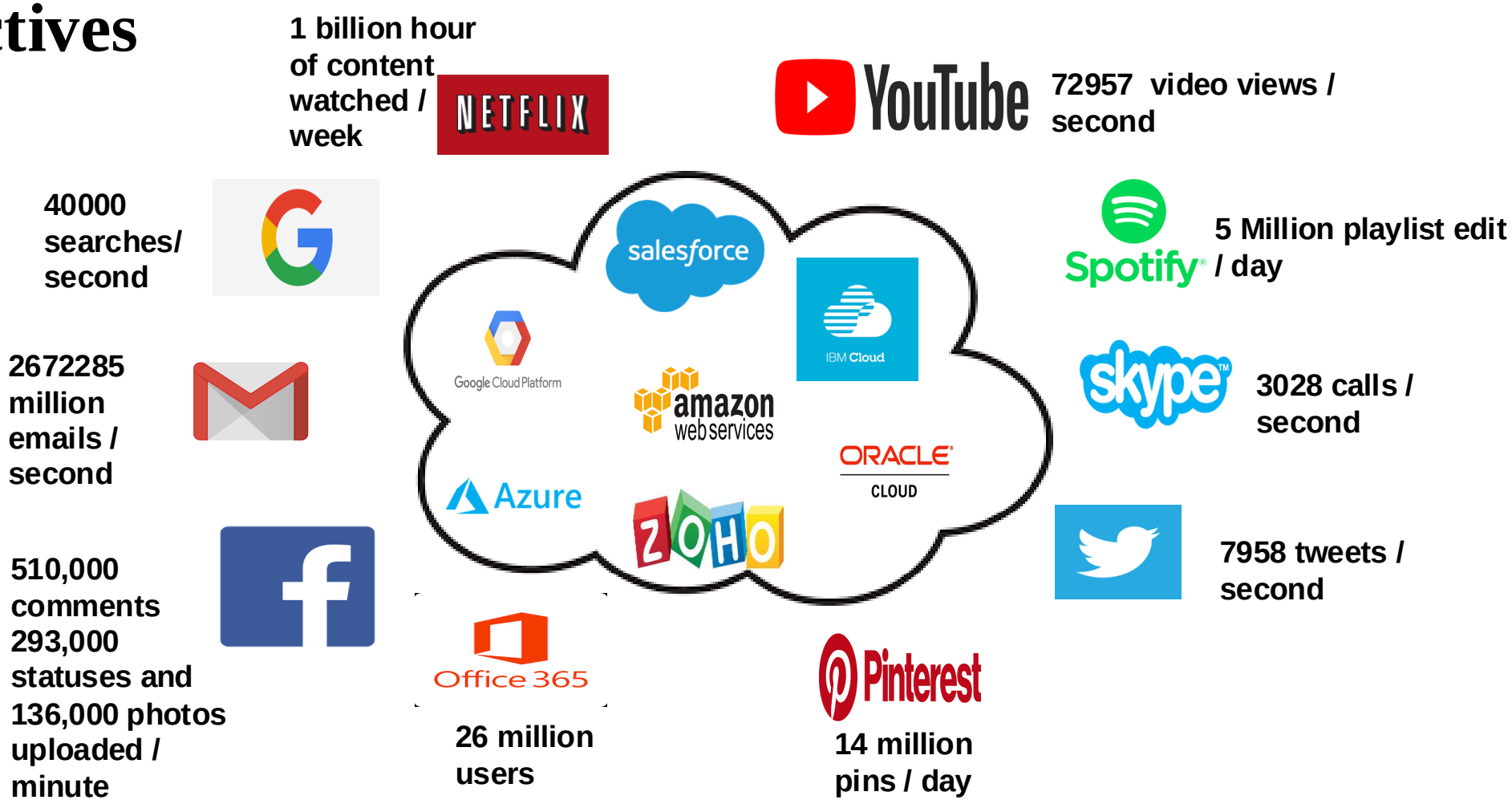
Why study Large Systems now?

- The Web and Internet applications and services are of unprecedented interest and importance.
  - SAP /Salesforce– enterprise management software
  - Banking and government services
  - Social Networks and Streaming Services: Facebook, WhatsApp, Skype, WeCha
  - Academic R&D worldwide: Service computing, e-Science, etc.
  - AI and its applications
  - Cloud Computing: Amazon EC2, Microsoft Azure, Google Cloud, Oracle



Large Systems: E.g., Cloud computing

User perspectives



*These numbers were before COVID

Large Systems….

Infrastructure perspectives

- BitTorrent transferring 9200 petabytes/month (2022) hosting many, many petabytes
- Spotify Hadoop cluster has over 3,000 nodes and over 100,000 terabytes of capacity
- Bank of America recently migrated to a private cloud: 200,000 servers to 70,000
- AI pipelines require massive accelerated computing/ GPU Clusters
  - Training and Inference platforms for LLMs now require 100's Megawatts of power infrastructure (computational resources)
- Google, Azure and many cloud service providers are hiring nuclear scientists to develop modular energy sources to power their data centres
  - Google recently procured nuclear energy in this direction

This Course

- What is their design?
- How is this design implemented?
- How do you administer such systems?

The SNE degree and this subject, in particular, aim to convey insight and principles into the design, implementation and administration of such large systems (from an infrastructure perspective).

Lecture Timelines

Time: Monday & Thursday 10-12 **Location:** SP B1.23

Week	Date	Topic	Personnel
1a	Oct 28	Course Intro & Distribution Intro & Virtualization	Shashi
1b	Oct 31	Virtualization	Shashi
2a	Nov 4	Virtualization	Shashi
2b	Nov 7	Virtualization	Shashi
3a	Nov 11	Communication & Coordination	Shashi
3b	Nov 14	Replication & Partitioning	Shashi
4a	Nov 18	Fault Tolerance	Shashi
Colloquia	Nov 20	Ansible 101	Industry Maurice [SUE]
4b	Nov 21	Admin: Servers, Desktops, Data Centers	Shashi
5a	Nov 25	Cloud Computing Intro, Cloud Application Development and Management	Shashi
5b	Nov 28	Advanced Large System Use cases: Decentralized blockchain systems	Zhiming
6a	Dec 2	DevOps [Guest lecture]	Gabriel
6b	Dec 5	Best Industrial practices for IaC and Automation	Industry Maurice [SUE]
7a	Dec 9	LS Presentation 1 [First slot]	
7b	Dec 12	LS Presentation 2 [Second slot]	
8a	Dec 16	Recap and Summary	Shashi
8b	Dec 19	Exam	

* Please note that each week's topic might be slightly different based on topic progress

Assignment Timelines

Lab Sessions: Time: Monday & Thursday 13-17 **Location:** B1.24CDE

There will be five assignments to be conducted in Week 1-7

1) Virtualization
2) VM migration
3) Infrastructure as code
4) Public cloud/AWS
5) Scientific measurements

Actual assignment deadlines will be informed in the lab sessions

* Assignments must be graded sufficient/completed to pass the course/attend the final exam!

Guest Speakers (planned)

- Maurice Mouw [SUE]
  - Configuration Management Basics: Ansible 101
- M. Mouw [SUE]
  - Best Industrial practices for IaC and Automation
- Zhiming Zhao(MNS, UvA)
  - Large System Use Cases: A decentralized blockchain network
- Gabriel Pelouze (UvA)

Site Visit!

- SURFSara Visit
- Data Center on Science Park
- Preliminary date: In late November
- I will create a separate section on Canvas to register for the site visit.
- Group size 10-12 people (Will create 2-3 slots)
- Bring the ID on the day
- ID must have a photo

Computational Resources for Labs

- Each of you will be assigned a physical server
- You would have received your credentials and information about the server assigned to you
- If not, please contact TAs asap…

Grading

- Closed-book Exam 85 %
- Presentation 15%
- Each element at least 4.0
- Average above 5.5
- Labs sufficient [All assignments must be passed to consider for the final exam]

Presentation

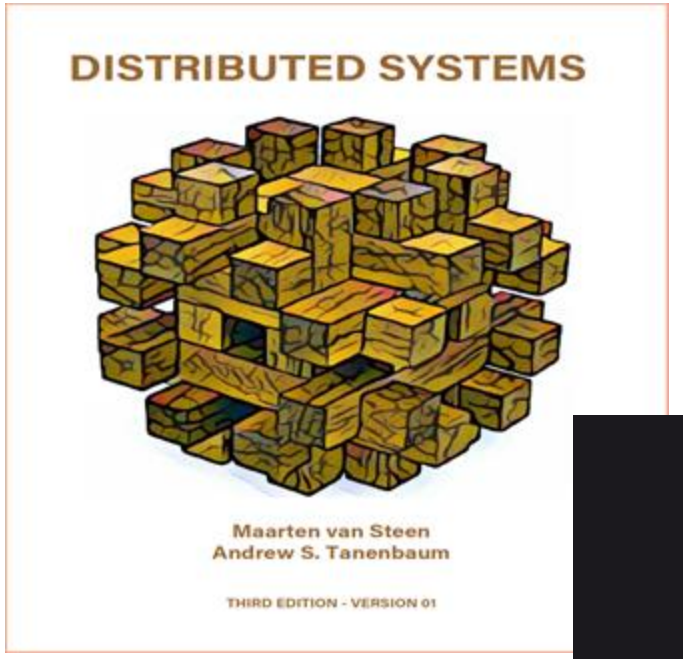
- Pick a recent paper about “Large Scale IT”: Academic: e.g. ACM / IEEE journal or conference from the field:
  - e.g. LISA, ATC, NSDI, ASPLOS, TPDS, EuroSys, SIGMETRICS
- Present this paper to the group in Week 7
- Presentation 10 minutes + 5 minutes questions
- Prepare in Week 6
- Group Size 2, different partner than SSN
- Can start looking for paper now.
- A separate submission link will be created on Canvas where you can register groups and upload files (Need to upload your selected paper and, later your PPT)
- After group registration, send email me your selected paper for approval.

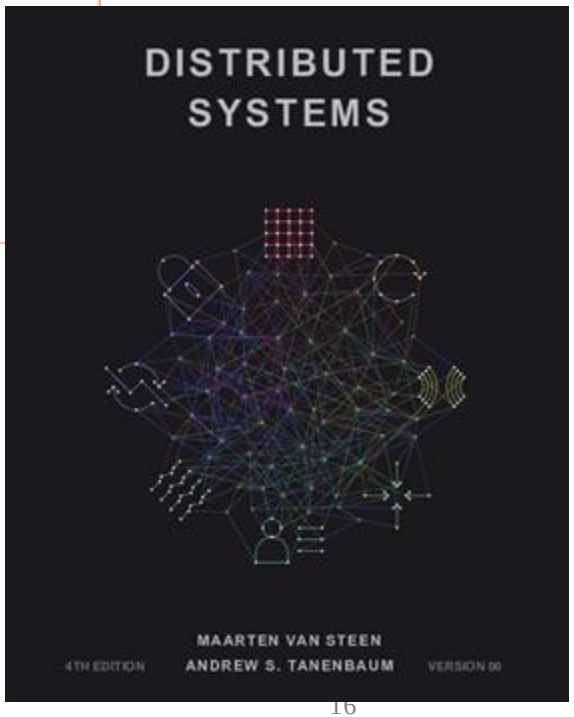
Lecture Presentation Slides

- Usually on the canvas just before the lecture
- Mostly derived from the text book.
- Good ideas and figures from alternative text book or reference may also be used.

Book 1

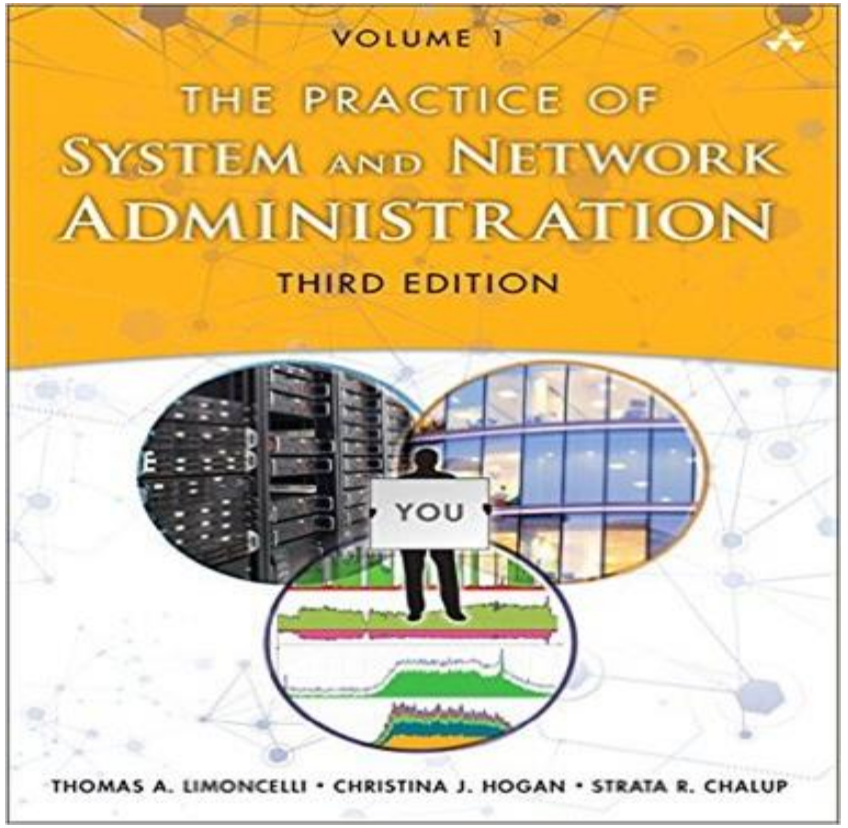
- Covers Design + Implementation
- Third Edition, 2017
- Selected chapters are basis for Lectures, Exam
- Free download: https://www.distributed-systems.net/index.php/books/ds3/ds3-ebook/





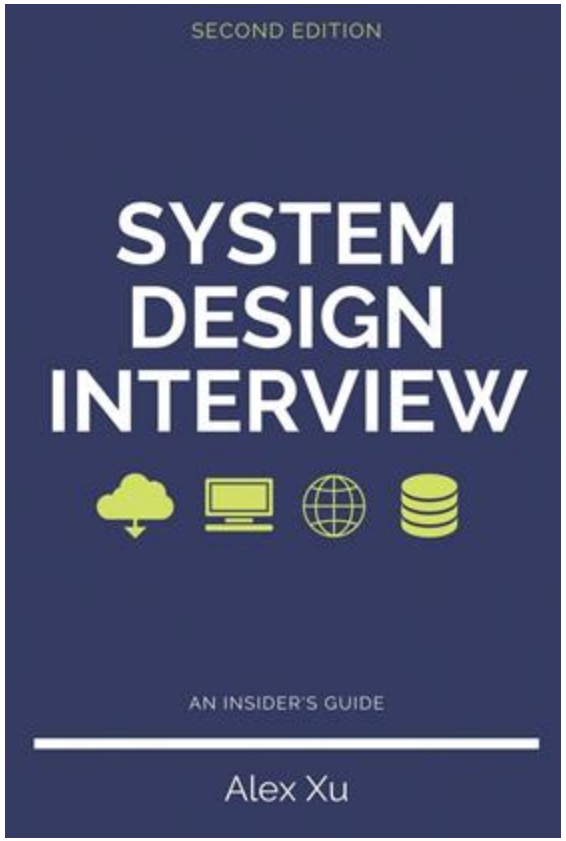
Book 2

- Covers: Administration
- Third Edition, 2016
- Selected chapters are basis for Lectures, Exam



Recommended Book

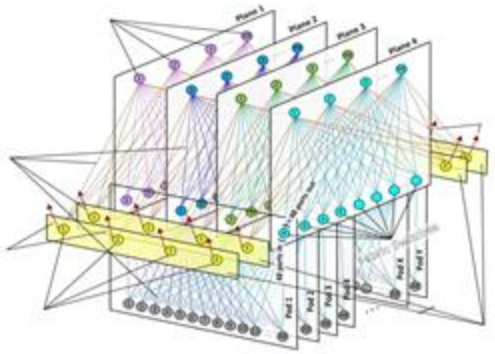
- Alex Xu
- “System Design Interview”
- Second edition, 2020
- **Not** part of the exam
- Easy guide to designing modern distributed systems
- Also: Volume 2, many designs



*We will also provide reading materials as required…

What do we expect from you?

- Regular attendance of lectures
  - Pay full attention, fully committed to learning new things, ask questions during the class (in labs as well), and participate in discussions.
- Review previous lecture material before coming to the class. – read material from the Textbook
- Start working on assignments right from the day they are announced and submit them on time.
- One deadline extension /person is allowed during the entire semester
- If you have any issues about the lectures/subject/??, please discuss it with us **early**.



Large Systems:

Design +
Implementation +
Administration