UC San Diego

CSE 291: Operating Systems in Datacenters

Amy Ousterhout

Sept. 27, 2022

UC San Diego

Agenda for Today

- Announcements
- CloudLab Overview
- Introduction to multicore and heterogeneity
- Multikernel discussion

Announcements

- Office hours
  - Amy: Tuesday after class or by appointment in CSE 3130
  - Anil: Friday 2-3 pm in CSE 3109
- The course is up on Canvas
- Sign up to lead a discussion
  - Due Thursday 9/29 at 11:59 pm
- Warm-up assignment
  - Will be posted on the website this evening
  - Due ~~Thursday 10/6~~Tuesday 10/11 at 11:59 pm
- First Day Survey #FinAid
  - New UCSD requirement!
  - Due Friday 10/7

UC San Diego

CloudLab Overview

Platforms for Experimentation

- Private compute resources
- Public clouds

- PlanetLab (2002-2020)
- Emulab
- Geni
- Mass Open Cloud (since 2013)
  - Enables experimentation with real users
- CloudLab (since 2014)

focus on networks and distributed systems

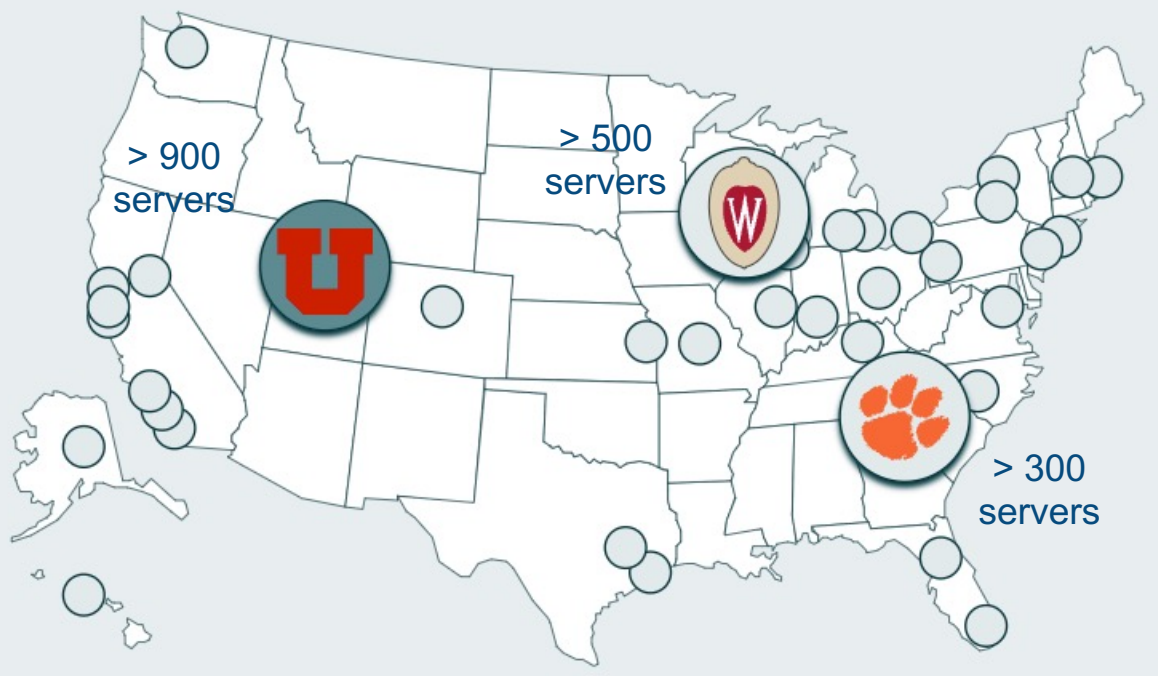
impacted

CloudLab Goals

- Customization
  - Modify storage, virtualization, networking
- Repeatable research
  - Bare metal
  - Uniform performance

What is CloudLab?

A testbed for research on cloud computing



https://www.usenix.org/sites/default/files/conference/protected-files/atc19_slides_duplyakin.pdf

Hardware in CloudLab

- Standard CPUs, memory, storage, NICs
- Specialized hardware: Intel Optane, GPUs, 100 Gbit/s NICs, SmartNICs
- Details at: `https://docs.cloudlab.us/hardware.html`

Warm-up assignment:

m510	270 nodes (Intel Xeon-D)
CPU	Eight-core Intel Xeon D-1548 at 2.0 GHz
RAM	64GB ECC Memory (4x 16 GB DDR4-2133 SO-DIMMs)
Disk	256 GB NVMe flash storage
NIC	Dual-port Mellanox ConnectX-3 10 GB NIC (PCIe v3.0, 8 lanes

Other options:

r7525	15 nodes (AMD EPYC Rome, 64 core, 512GB RAM, 2 x GPU)
CPU	Two 32-core AMD 7542 at 2.9GHz
RAM	512GB ECC Memory (16x 32 GB 3200MHz DDR4)
Disk	One 2TB 7200 RPM 6G SATA HDD
NIC	Dual-port Mellanox ConnectX-5 25 Gb NIC (PCIe v4.0)
NIC	Dual-port Mellanox BlueField2 100 Gb SmartNIC
GPU	Two NVIDIA GV100GL (Tesla V100S PCIe 32GB)

Who pays for CloudLab?

- National Science Foundation
- Free to use for research and educational purposes

How to Use CloudLab

- Create a ”Project Profile” which specifies:
  - The configuration of 1 or more servers
  - Network connectivity between them
  - Software to run on servers
- Instantiate your Project Profile to create an experiment
  - Start immediately or make a reservation
  - Stop your experiment when you’re done
  - It will terminate after 16 hours

Why use CloudLab?

- Cost – cheaper than buying and maintaining your own resources or using public clouds
- Flexibility – can try out different computing resources for short periods of time
- Customization – tune the hardware
- Reproduceable research

runs on CloudLab

AIFM: High-Performance, Application-Integrated Far Memory

Zhenyuan Ruan Malte Schwarzkopf † Marcos K. Aguilera ‡ Adam Belay
MIT CSAIL *†Brown University* *‡VMware Research*

Abstract. Memory is the most contended and least elastic resource in datacenter servers today. Applications can use only local memory—which may be scarce—even though

Throughput [accesses/sec]	*64B object*	*4KB object*
Paging-based (Fastswap [6])	582K	582K
AIFM	3,975K	1,059K

What kind of research is CloudLab not ideal for?

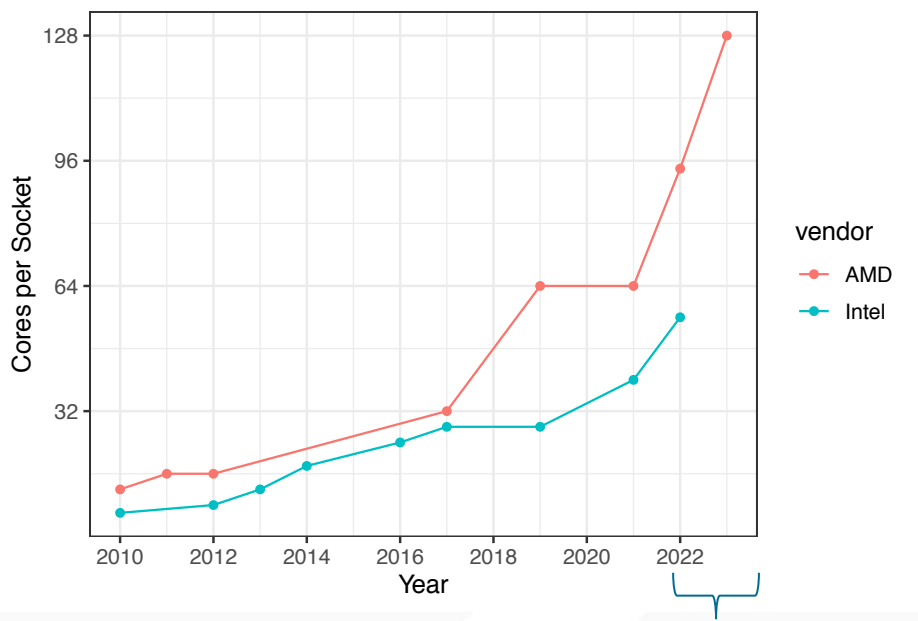
- Large scale – requiring hundreds or thousands of nodes
- Locations
  - More than a few locations
  - Specific locations
- Real cloud users

UC San Diego

Introduction to Multicore and Heterogeneity

The Rise of Multicore

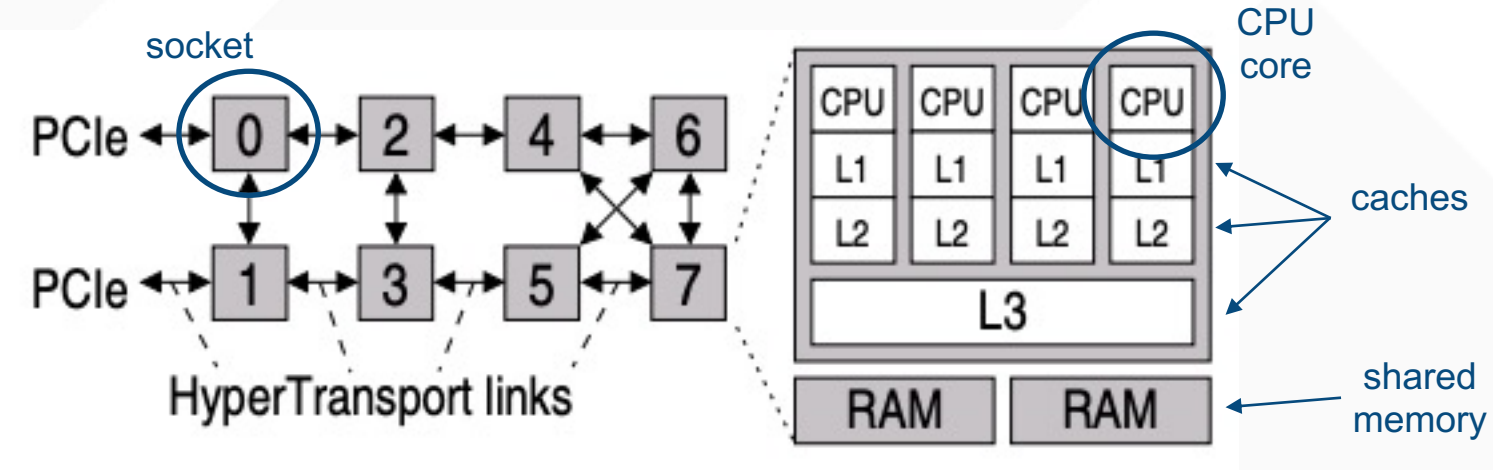
- First multicore CPUs?
  - IBM's Power4 in 2001 had 2 cores
- Driven by the end of Moore's Law



Data from: https://www.servethehome.com/amd-and-intel-2p-server-core-count-growth-2010-2022/

Multicore Architectures

- Cores grouped into sockets
  - Also referred to as “multi-core processors” or “NUMA nodes”
  - Interconnect in between them
- Cache coherence
  - Keeping data in separate caches consistent



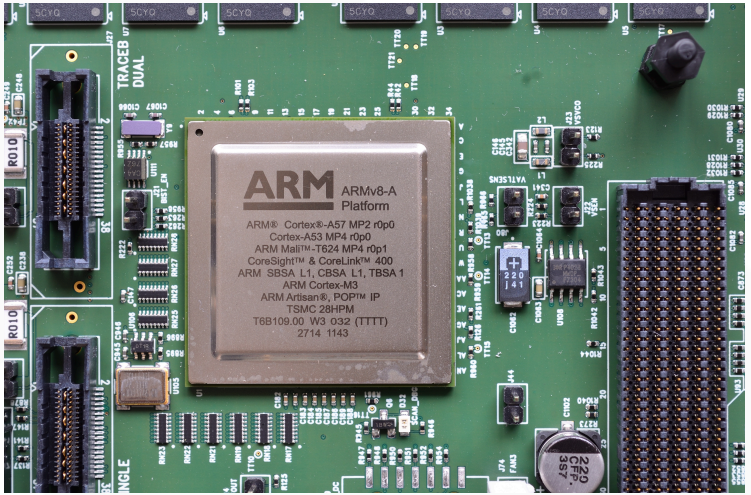
Challenges of multicore

- How do you design **applications** that scale well across many cores?
- How do you build an **OS** that scales well? ← today
  - How do you design the application-OS **interface** to enable scalability? ← "The Scalable Commutativity Rule" [SOSP '13]
  - How do you build a scalable **network stack**? ← Thursday
  - How do you make **CPU scheduling** scalable? ← later this quarter
  - How do you scale **memory management**? ← later this quarter

Heterogeneity

- Heterogeneous processors
  - Different power consumption ← common in mobile and tablets
    - E.g., ARM's big.LITTLE announced in 2011
  - Different ISAs
- Other types of heterogeneity:
  - GPUs
  - SmartNICs
  - FPGAs
  - Accelerators

common in datacenters



Challenges of heterogeneity

- May not be cache coherent
- Different memory layouts
- Different models of computation (e.g., GPU vs. CPU)

UC San Diego

Multikernel Discussion