Basics of Cluster Computing

Pranjal Rawat

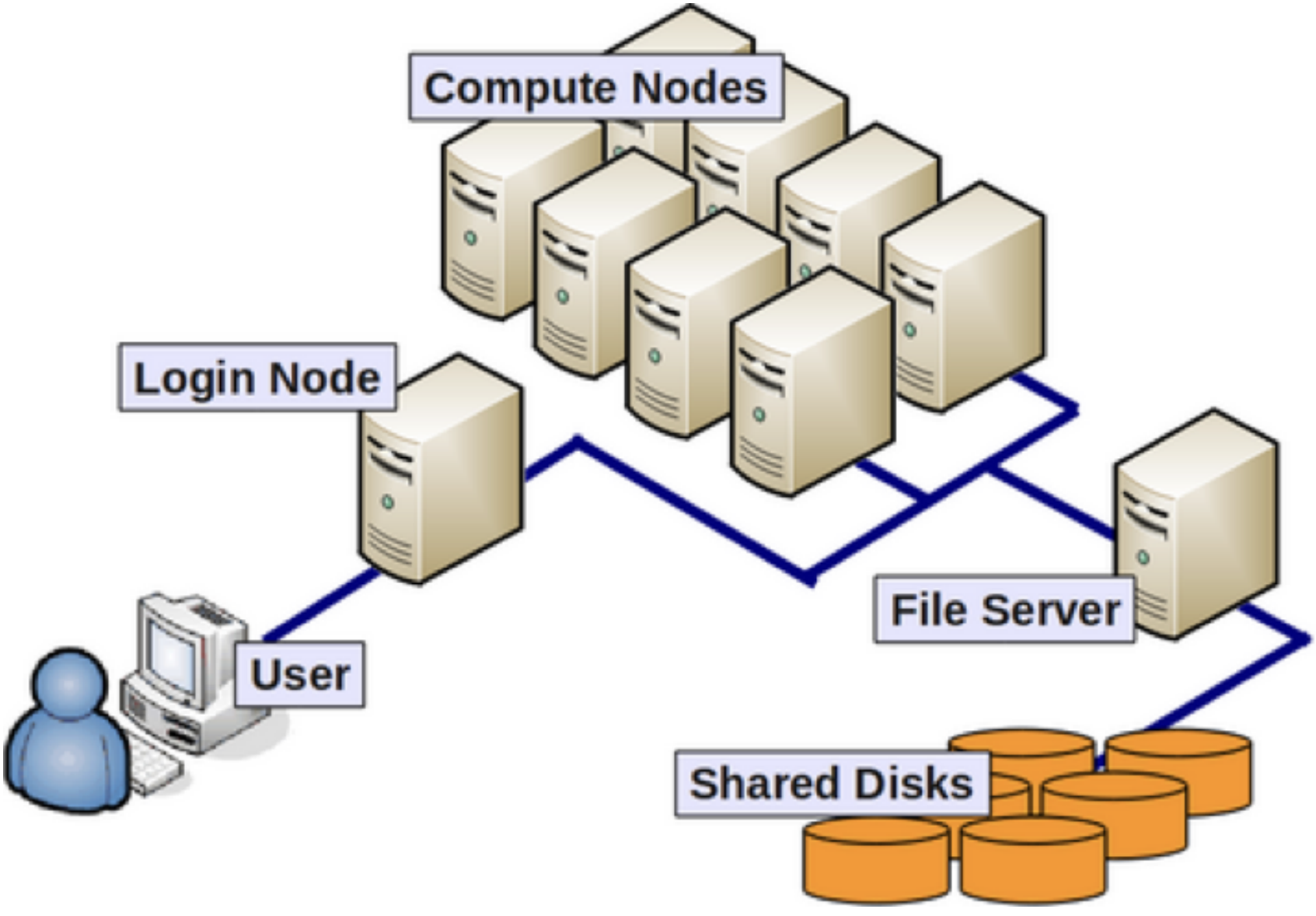
December 26, 2023

Georgetown University

Definitions - I

- Cluster: A network of computers.
- Node: An individual computer within a cluster.
- Storage: Disks in network where data is stored.
- OS: An operating system (e.g., Linux, Windows)
- Admin: Oversees access, software installation and permissions.

Cluster



Definitions - II

- High-Performance Computing (HPC) Cluster: Cluster with better processing units (GPUs), high-speed interconnects, and optimization for parallelization.
- Graphics Processing Unit (GPU) - specialized electronic circuit designed to accelerate the rendering of graphics and perform complex mathematical calculations. (e.g. NVIDIA GPU)
- Virtual Machine (VM): isolated virtualized computing environment managed in a cluster.
- Cloud Computing: use of clusters through internet only.
- Cloud Service Provider (CSP): cloud computing environment - e.g. Amazon AWS, Microsoft Azure, Google Gcloud.

Why use clusters?

- It's much faster than your laptop and can store more data.
- You can do the following easily:
  - Estimation of complex models - machine learning, structural IO, Bayesian models, HANK
  - Monte Carlo Simulations and Agent-Based Models
  - Big data processing - image, text, graphs.
- You can automate the boring stuff.
- You can parallelize tasks.
- You can share data and code easily.

Automation

Reading and writing files

Consolidating tasks

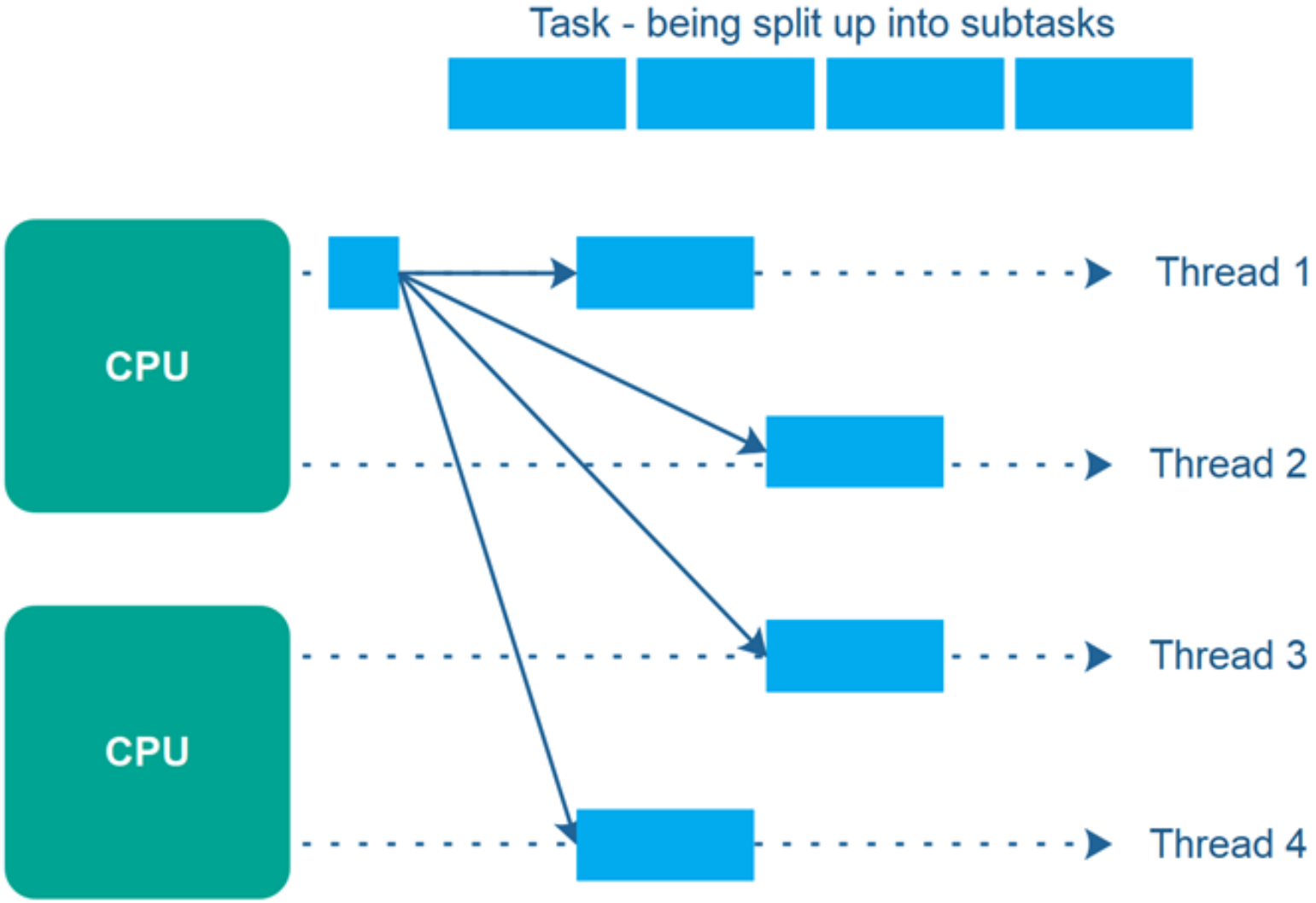
Interacting with APIs

Reformatting

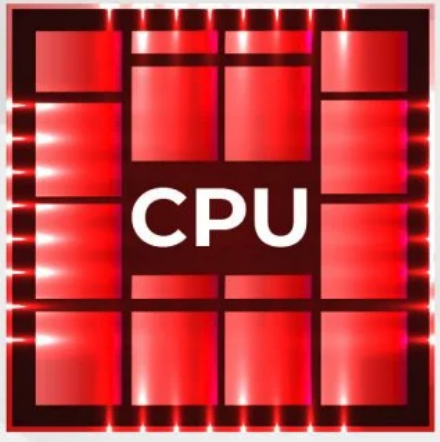
Organizing data

Web scraping

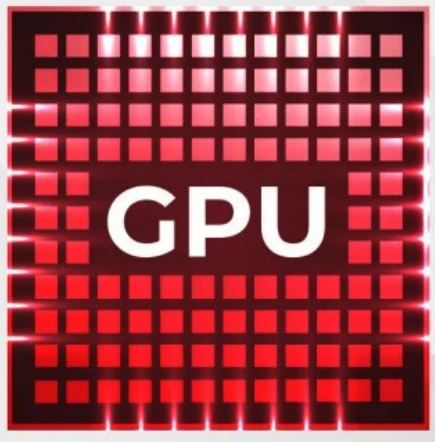
Multi-Threading



CPU vs GPU



- CPUs have few strong cores
- Suited for serial workloads
- Quick access to System Memory (RAM)



- GPUs have thousands of weaker cores
- Suited for parallel workloads
- Can only access vRAM quickly

Definitions - III

- Cores: These are individual processing units that can handle tasks independently.
- Threads: Number of processes each core can handle.
- Cache: high-speed storage in core for frequently accessed data
- RAM: high-speed storage in cluster for frequently accessed data

Resources

Internal:

- "Econ Server"
- HPC

External:

- Google collab, Pro
- Gcloud, AWS, etc.

UIS Econ Server

- Hostname: econ-prod-1.uis.georgetown.edu
- Owner: Prof John Rust
- Installation Steps:
  - Install GU VPN
  - Log in via Netid pwd and DUO code.
  - In terminal/command line do:
    ssh NETID@econ-prod-1.uis.georgetown.edu

Working in a cluster

- Bash - file and process management
- Vim - to edit files
- nohup - to schedule jobs
- venv - virtual environments
- git - repository management

HPC

- Virtual Machine on Gcloud Compute
- Login Node: figs-controller
- GPU Nodes: gpu, gpuspot, spot
- Installation Steps:
  - Get access
  - gcloud init
  - gcloud compute ssh figs-controller
- Slurm commands - sinfo, srun, sbatch, scancel
- Slurm batch files