

HPC SEMICONDUCTOR LAB SYSTEM OVERVIEW

Nguyen Quang

TABLE OF CONTENTS

- Introduction
- System Overview
- Network Overview
- Main Features
- Node Roles & Services
- Flow
- High Availability

INTRODUCTION

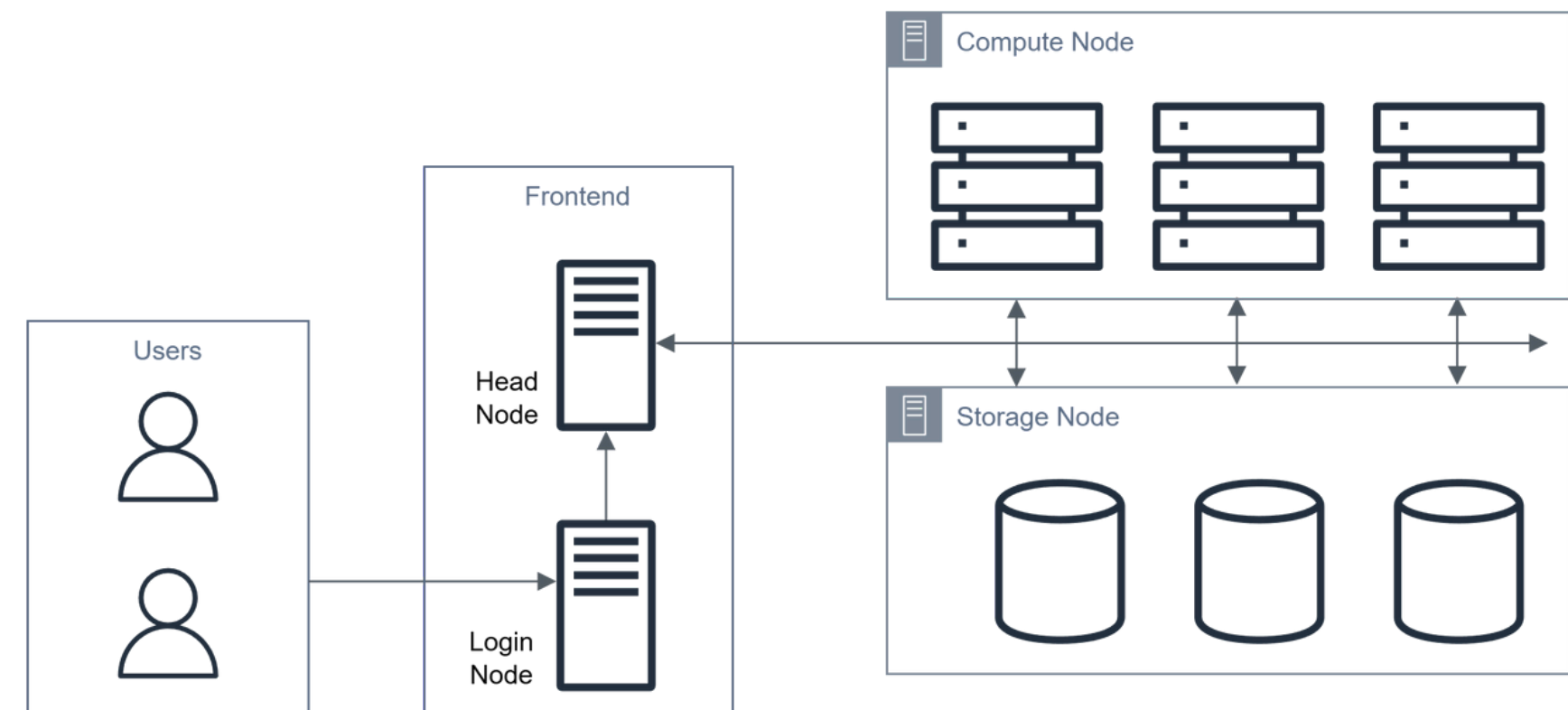
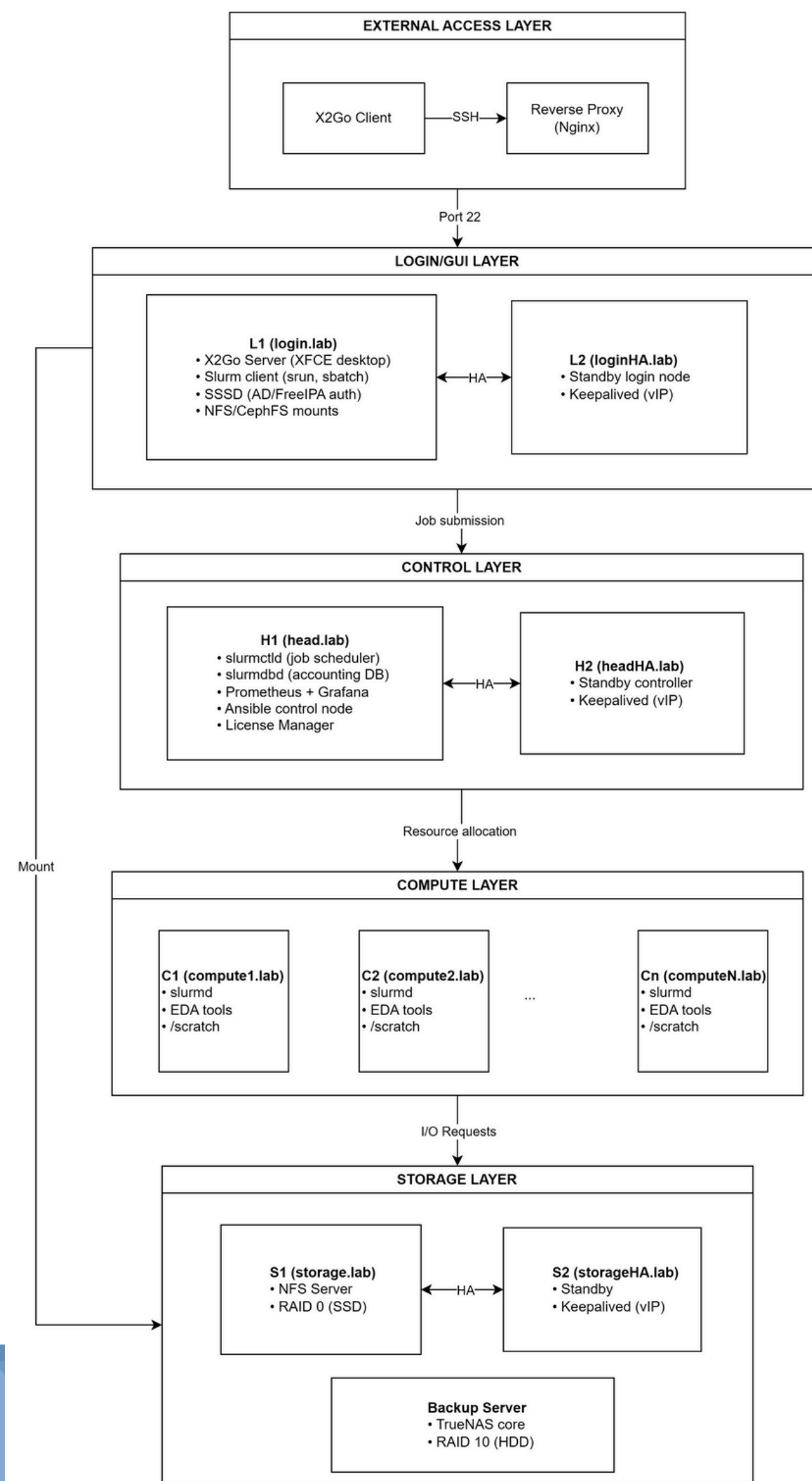
- Cung cấp môi trường mô phỏng và thiết kế mạch bán dẫn tập trung, hiệu năng cao.
- Cho phép sinh viên, giảng viên và trợ giảng đăng nhập, chạy các phần mềm EDA/TCAD/COMSOL thông qua giao diện đồ họa (GUI) của Rocky Linux 8.
- Đảm bảo tính sẵn sàng, khả năng mở rộng cao và bảo trì lâu dài, đồng thời giữ chi phí hợp lý cho quy mô đào tạo.

Minimal Requirements

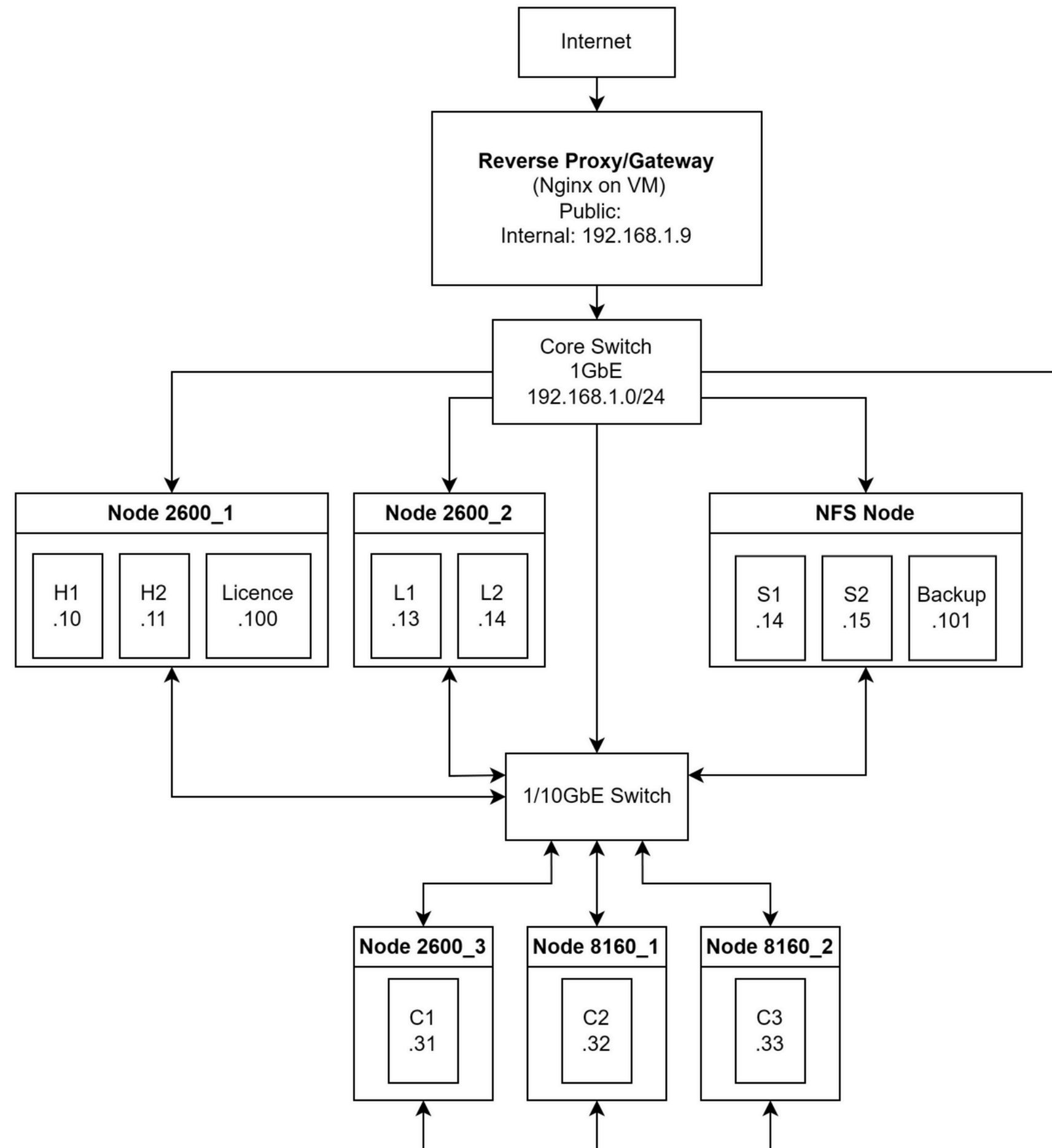
		Minimal Requirement
Head Node	1 Node	4 vCPU, 8 GB RAM, 200 GB SSD
Login / GUI Node	1 Node	8 vCPU, 16 GB RAM
Compute Node	≥ 2 Node	≥ 8 vCPU, ≥ 32 GB RAM
Storage Node	1 Node	SSD, ≥ 1 TB usable
Network		1 Gbps Ethernet

SYSTEM OVERVIEW

- **H1 Head Node:** slurmctld + slurmdbd + Prometheus + Granfana + MariaDB.
- **L1 Login/GUI Node:** X2Go, XFCE, SSSD.
- **C1–Cn Compute Nodes:** slurmd, tools.
- **S1 Storage Node:** RAID10, SSD cache, NFS.
- Nginx Reverse Proxy, Backup Server



NETWORK OVERVIEW



MAIN FEATURES

User Management & Authentication

Công cụ sử dụng:

1. FreeIPA (Identity & Access Management)

Tại sao lại chọn FreeIPA mà không phải chọn LDAP thuần, Microsoft AD, hay OpenLDAP?

- FreeIPA = LDAP + Kerberos + DNS + Certificate Authority + Policy → all-in-one.
- FreeIPA tích hợp gần như đầy đủ, chuẩn quốc tế, triển khai nhanh, ổn định hơn OpenLDAP
- FreeIPA opensource, tích hợp tốt với Linux, có tích hợp webUI quản trị
- Ms AD tốn phí, dư tính năng

2. SSSD (Authentication client)

- Tích hợp tốt với FreeIPA
- Cache được credentials khi offline, giúp job và session vẫn có thể chạy được khi mất mạng.
- Bảo mật cao, tích hợp đồng nhất trên toàn cluster

MAIN FEATURES

Job Scheduling

Công cụ sử dụng:

1. Slurm (slurmctld, slurmd, slurmdbd)

- Open source, dễ cấu hình, phổ biến dùng cho HPC trên thế giới
- Hệ thống plugins, community mạnh
- Tích hợp fair-share
- Có GUI monitoring

2. MariaDB (Logs, Accounting)

- Open source, ổn định cao
- Hiện tại slurm chỉ chính thức hỗ trợ MySQL và MariaDB
- MariaDB là nhánh của MySQL và tương thích tốt hơn

MAIN FEATURES

GUI Access

Công cụ sử dụng:

1. X2Go, XFCE

Vì sao chọn X2Go thay vì VNC, NoMachine, GNOME hay X11 forwarding thuần?

- X2Go sử dụng NX protocol – một giao thức thiết kế tối ưu cho môi trường băng thông thấp và độ trễ cao.
- Lựa chọn phổ biến trong các lab, HPC và trường đại học.
- Kết nối an toàn qua SSH, không expose X11 native ra Internet.
- X2Go render server-side, chỉ gửi hình ảnh đã nén, mượt hơn VNC nhiều lần.
- X2Go hỗ trợ đa phiên (multi-session), X11 không hỗ trợ.
- GNOME quá nặng, NoMachine cùng giao thức, tốt hơn X2Go nhưng có chi phí cao.
- Phù hợp hệ thống only-CPU

2. Nginx (reverse proxy)

3. cgroups (Giới hạn tài nguyên)

MAIN FEATURES

GUI Access

Công cụ sử dụng:

4. Open OnDemand (OOD)

Open OnDemand làm được gì?

- Web UI xem Slurm Queue
 - Xem job details, output files, resources
 - Submit job bằng giao diện
 - File Manager (xem file home/proj giống như Google Drive)
 - Terminal Web
- Rất phù hợp cho sinh viên vì dễ dùng và không yêu cầu cài bất kỳ phần mềm nào.
- Admin dễ theo dõi lớp. Thấy được ai chạy nhiều job, ai pending quá lâu.

MAIN FEATURES

Storage System

Công cụ sử dụng:

1. NFS (Network File System)

- Setup nhanh, bảo trì dễ, đã có sẵn trên hệ thống hiện tại.
- Tương thích tốt với autofs, freeIPA
- Performance ổn định, đồ trễ thấp

2. Quota, ACL (Resource Limit)

- Quota giới hạn bộ nhớ sử dụng tránh đầy hệ thống
- ACL hỗ trợ phân quyền linh hoạt hơn chmod 755/775
- Phân quyền k ảnh hưởng file user khác

3. restic (Backup & Restore)

- Dùng để backup hàng tuần và restore
- Tốc độ thực hiện nhanh, hầu như chỉ phụ thuộc vào tốc độ của disk và đường truyền mạng
- Open source

MAIN FEATURES

Security

Công cụ sử dụng:

1. openSSH

- Vô hiệu hoá đăng nhập root trực tiếp → Tránh nguy cơ brute-force
- Giới hạn MaxAuthTries, MaxSessions → ngăn tấn công đoán mật khẩu và lạm dụng SSH multiplexing.
- Chỉ cho phép các user hợp lệ truy cập

2. FreeIPA + integrated Kerberos

- Áp đặt chính sách mật khẩu tập trung: độ dài tối thiểu, số lớp ký tự, thời gian hết hạn.
- Bắt buộc reset mật khẩu lần đầu.
- Kerberos ticket giảm rủi ro password exposure khi sử dụng nhiều dịch vụ (Slurm, NFSv4).
- Đồng bộ với SSSD trên tất cả node.

MAIN FEATURES

Security

Công cụ sử dụng:

3. SELinux (Enforcing Mode)

- Ngăn tiến trình trái phép truy cập file hệ thống (prevent privilege escalation).
- Kiểm soát nghiêm mô hình DAC + MAC → hạn chế user chạy phần mềm tùy ý.
- Tạo sandbox tự động cho dịch vụ nhạy cảm: nfsd, slurmd, x2goserver.
- Quan trọng với login node (L1), nơi user chạy GUI + scripts.

4. Fail2Ban

- Chặn tự động IP brute-force password SSH.
- Giảm thiểu rủi ro từ user nội bộ cố ý hoặc vô tình thử nhiều mật khẩu.
- Tự động unban sau thời gian định trước.
- Login node (L1/L2) là điểm duy nhất lộ ra internet qua reverse proxy → Fail2Ban giữ an toàn.

MAIN FEATURES

Monitoring & Observability

Công cụ sử dụng:

1. Auditd (Linux auditing system)

- Ghi lại toàn bộ sự kiện nhạy cảm. VD: sudoers, login/logout, slurm, permission change.
- Xuất log cho Prometheus/Grafana dashboards phục vụ phân tích bảo mật.

2. Prometheus

- Sử dụng Node Exporter trên toàn bộ các node.
- Thu thập định kỳ (5–15s) toàn bộ metrics quan trọng.
- Phát hiện và cảnh báo sớm các sự cố
- Hạn chế overhead (Node Exporter chỉ <1% CPU)
- Gửi cảnh báo qua email, Telegram, Discord hoặc dashboard.

3. Grafana

- Sử dụng datasource từ Prometheus tạo dashboard trực quan hóa.
- Dễ dàng phát hiện nút nghẽn (I/O hotspot, compute imbalance, memory leak).

MAIN FEATURES

Others

Công cụ sử dụng:

1. Ansible (Automation)

- Quản lý cấu hình đồng nhất, dễ dàng mở rộng cụm HPC.
- Triển khai package, dịch vụ, user, quota bằng playbook.
- Cấu hình version-controlled, dễ vận hành.

2. chronyd

- Đồng bộ thời gian cho toàn bộ hệ thống, tránh lỗi bảo mật, xác thực.

3. iperf3

- Benchmark NFS, phát hiện bottleneck mạng.

NODE ROLES & SERVICES

1. Head Node – Quản lý & điều phối

Vai trò

- Trung tâm điều phối toàn bộ HPC cluster.
- Chạy Slurm Controller (slurmctld) và SlurmDBD.
- Quản lý trạng thái job, queue, tài nguyên.
- Thu thập và lưu trữ accounting, logs.
- Không cho user login trực tiếp.

Cách hoạt động

1. Nhận yêu cầu submit job từ Login Node / Open OnDemand
2. Phân tích tài nguyên (CPU, RAM, time, policy)
3. Quyết định node thực thi job
4. Gửi lệnh xuống Compute Node
5. Ghi accounting vào MariaDB
6. Cung cấp metrics cho Prometheus / Grafana

NODE ROLES & SERVICES

1. Head Node – Quản lý & điều phối

Services

	Functions
slurmctld	Lập lịch job, áp dụng policy, phân bổ CPU/RAM/time cho từng job
Munge	Xác thực tin cậy giữa slurmctld ↔ slurmd
Slurmdbd + MariaDB	Thu thập log job (user, node, thời gian, tài nguyên), lưu trữ accounting, lịch sử job
FreeIPA Server + Kerberos	Quản lý user, group, policy tập trung + Xác thực không truyền password, dùng ticket
Prometheus + Node Exporter	Thu thập metrics + Export metrics hệ thống của Head Node
Grafana	Hiển thị dashboard tải hệ thống, job running, resource usage
Ansible Controller	Tự động triển khai cấu hình, package, service trên các node
rsyslog (optional)	Thu thập log tập trung để debug, fix lỗi hệ thống
chronyd (server)	Đồng bộ thời gian, đảm bảo log & xác thực chính xác

NODE ROLES & SERVICES

2. Login/GUI Node – Cổng truy cập người dùng

Vai trò

- Điểm đăng nhập duy nhất của người dùng.
- Cung cấp môi trường GUI (X2Go, XFCE).
- Cho phép submit job, chỉnh sửa script.
- Không thực thi job nặng.

Cách hoạt động

1. User đăng nhập qua SSH / X2Go
2. Xác thực user qua FreeIPA + SSSD
3. Mount /home và /proj từ Storage Node (NFS)
4. User submit job → gửi lên Slurm Controller
5. Chuyển tiếp GUI, terminal, file manager cho user

NODE ROLES & SERVICES

2. Login/GUI Node – Cổng truy cập người dùng

Services

	Functions
X2Go Server	Cung cấp phiên GUI từ xa qua SSH
XFCE	Desktop chính cho sinh viên, nhẹ, tiêu thụ ít CPU/RAM
Open OnDemand	Web UI submit job, xem queue, file manager
SSSD	Nhận thông tin user/group từ FreeIPA
Slurm client	Gửi job tới Head Node
NFS client	Mount /home và /proj
Cgroup/PAM limits	Giới hạn tài nguyên mỗi phiên + Áp dụng policy đăng nhập, giới hạn truy cập
Fail2ban + SE Linux	Bảo vệ SSH/x2go, cô lập tiến trình GUI & SSH
Node Exporter	Gửi metrics CPU, RAM, disk, load
chronyd (client)	Đồng bộ thời gian với Head Node

NODE ROLES & SERVICES

3. Compute Node – Thực thi job

Vai trò

- Xử lý toàn bộ job.
- Có /slurmd nhận lệnh từ Head Node.
- Sử dụng /scratch cục bộ kéo từ NFS để tăng I/O cho job.
- Đồng bộ /home và /proj qua NFS.

Cách hoạt động

1. Nhận job từ Slurm Controller
2. Thực thi job
3. Truy cập dữ liệu từ Storage Node (NFS)
4. Gửi metrics về Prometheus
5. Không cho user SSH trực tiếp

NODE ROLES & SERVICES

3. Compute Node – Thực thi job

Services

	Functions
slurmd	Nhận job từ Head, chạy job theo yêu cầu đã cấu hình
Munge	Xác thực tin cậy giữa slurmctld ↔ slurmd
cgroups	Giới hạn CPU, RAM cho từng job
NFS Client	Truy cập input/output từ /home, /proj
Node Exporter	Export metrics hệ thống về Head Node
auditd	Ghi lại log hoạt động của job và user
SELinux	Cô lập job, ngăn truy cập trái phép
chronyd (client)	Đồng bộ thời gian, đảm bảo log & xác thực chính xác

NODE ROLES & SERVICES

4. NFS Storage Node – Lưu trữ & Backup

Vai trò

- Lưu trữ dữ liệu dùng chung cho toàn cluster.
- Cung cấp /home, /proj, dữ liệu job.
- Đảm bảo dữ liệu nhất quán giữa các node.
- Hỗ trợ quota và phân quyền.

Cách hoạt động

1. Export NFS cho Login & Compute Nodes
2. Cung cấp home directory cho user
3. Lưu input/output của job
4. Áp dụng quota & ACL
5. Thực hiện backup định kỳ (restic)

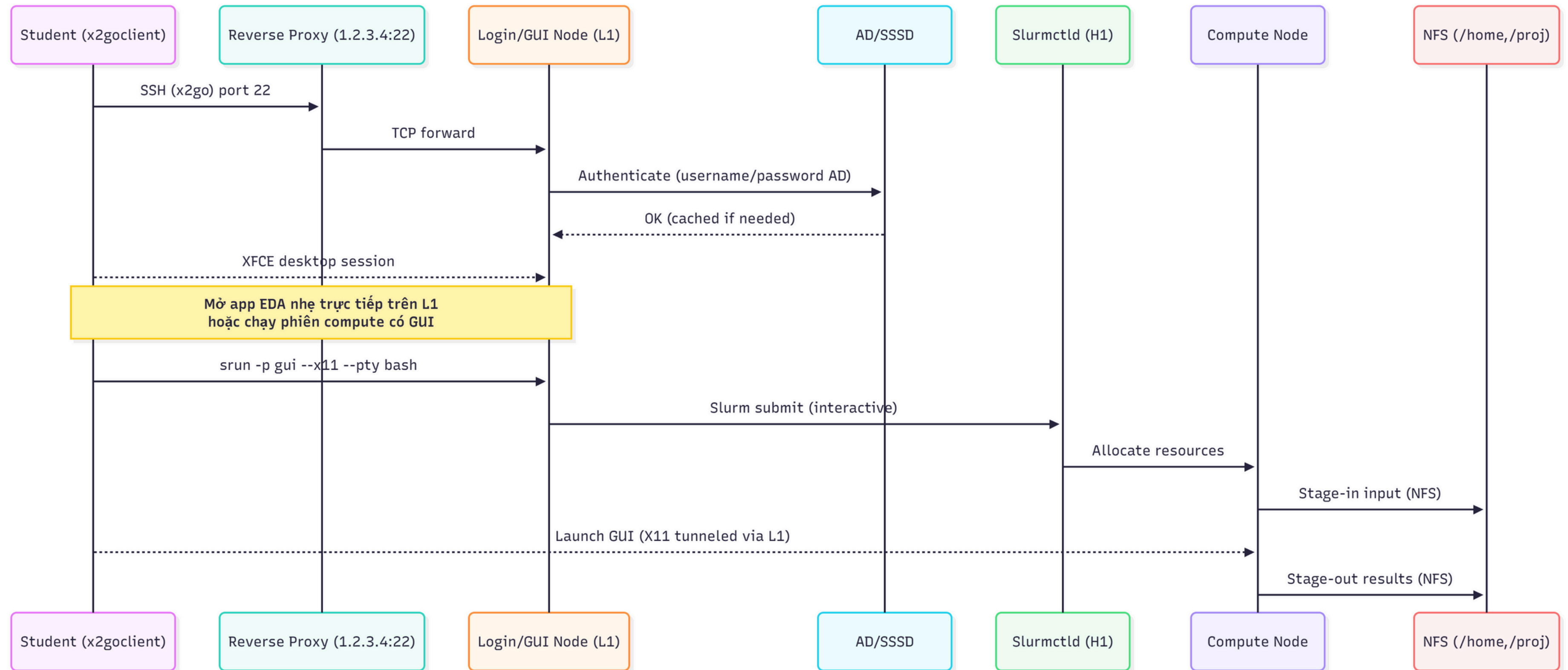
NODE ROLES & SERVICES

4. NFS Storage Node – Lưu trữ & Backup

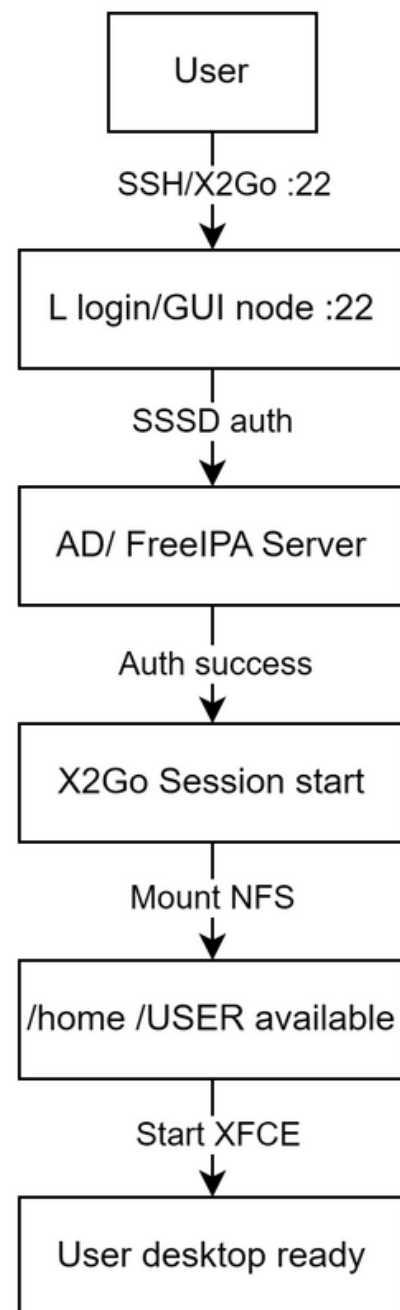
Services

	Funtions
NFS Server	Filesystem dùng chung cho cluster
Quota	Giới hạn tài nguyên dung lượng đĩa được phép sử dụng
ACL	Phân quyền chi tiết cho file và folder
autofs	Mount tự động trên client
restic	Backup định kỳ & Restore
Node Exporter	Export metrics hệ thống về Head Node
chronyd (client)	Đồng bộ thời gian

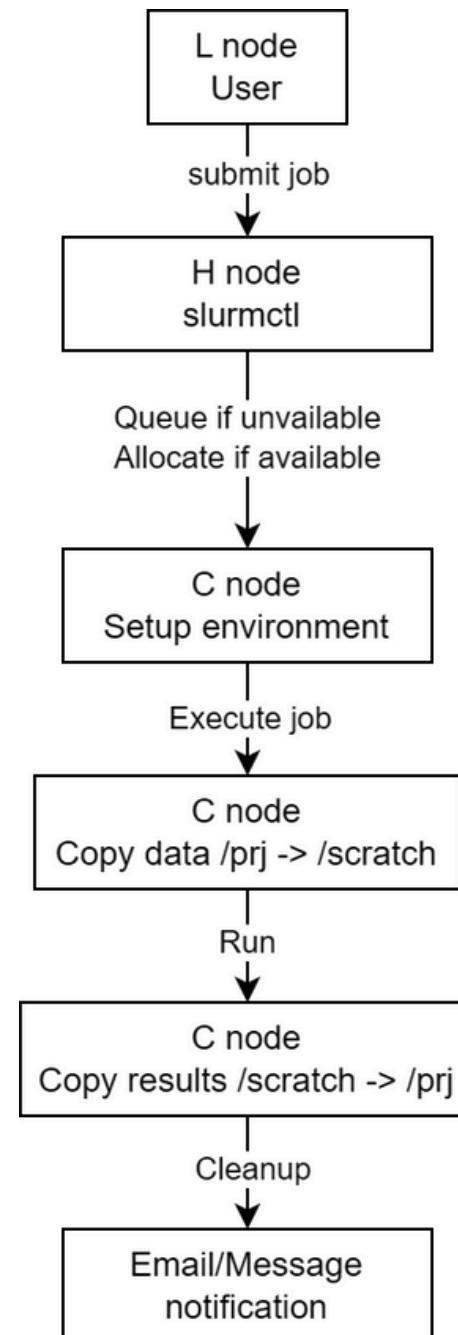
MAIN FLOW



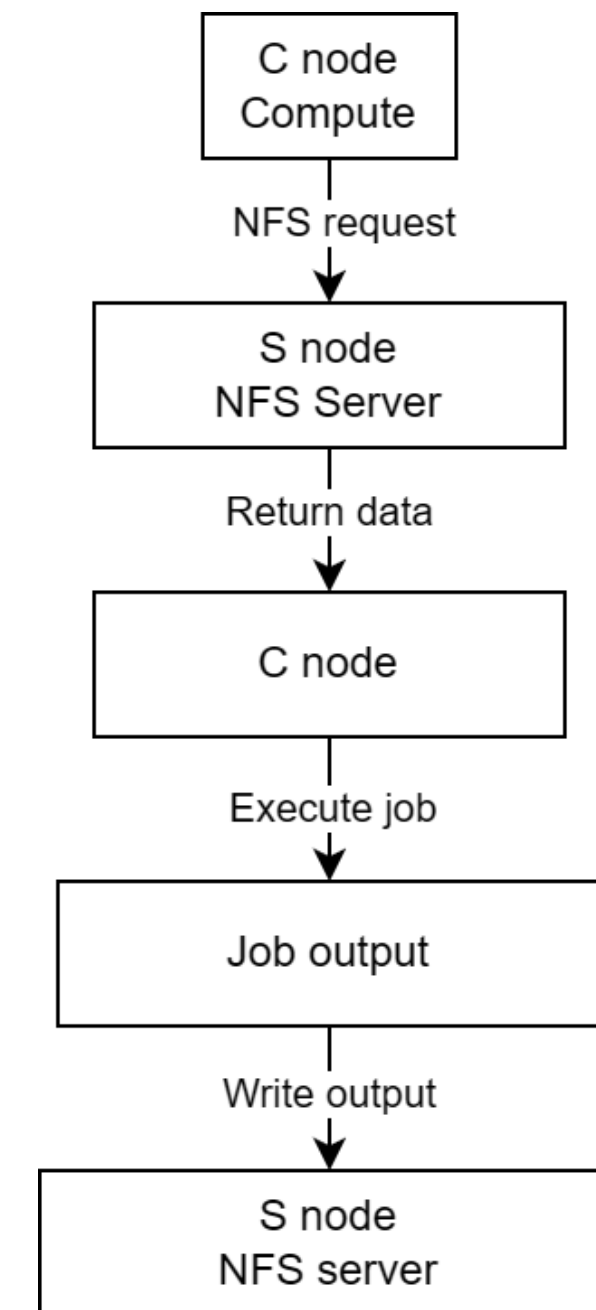
USER LOGIN FLOW



JOB SUBMISSION FLOW



STORAGE I/O FLOW



THROUGHPUT ESTIMATION

Login/GUI Remote

- Kerberos ticket (~1-5 KB)
- FreeIPA policy data (~10-50 KB)
- NFS mount metadata (~1-10 KB)
- User profile (~100 KB - 1 MB): .bashrc, .profile, settings
- X2Go session data (compressed images, ~1-5 Mbps streaming)

→ **>=200 Users sử dụng GUI đồng thời, tuy nhiên bị nghẽn I/O ở NFS**

Job Submission

- Job script file (~1-100 KB)
- Input data (variable: 1 MB - 100 GB) từ NFS
- Network: (Max 100MB/s)

→ **Thời gian tải phụ thuộc vào băng thông còn trống → 40 Jobs ~ 2.5MB/s**

HIGH AVAILABILITY (HA)

Head Node HA (Active-Backup + Failover IP)

- Dual Head Nodes (H1-H2) share chung thư mục StateSaveLocation. (Active - Backup)
- Keepalived Virtual IP bảo đảm Slurm controller tự động failover khi node chính gặp sự cố.
- MariaDB/SlurmDBD replication giúp dữ liệu accounting luôn nhất quán.
- VIP + replication giúp chuyển vai trò controller trong vài giây mà không gián đoạn công việc người dùng.

Login Node HA (Active-Active)

- Nhiều login nodes (L1-L2) chạy song song (Active-Active), tận dụng được hết tài nguyên
- Nginx TCP Load Balancer phân phối SSH/X2Go connections và tự loại node lỗi khỏi pool.
- LB giúp duy trì kết nối và phân tải trong giờ cao điểm.

Storage HA

- *NFS HA bằng Pacemaker + Corosync + DRBD (Open-Source, Active - Passive)*
 - + Không tốn license, failover vài giây.
 - + Tuy nhiên DRBD làm giảm I/O nếu workload nặng, chỉ là Active-Passive không tận dụng hết tài nguyên.
- *CephFS (100% open-source, phân tán, Active-Active)*
 - + HA tốt nhất, không cần license, tận dụng hết tài nguyên.
 - + Tuy nhiên cấu hình phức tạp, tốn nhiều máy.



THANK YOU