

Lets talk HPC / AI Environments

AI means a lot of different things to a lot of different people, depending on who you ask. It could mean just automation, chat bots / LLMs, any system with GPUs, or whatever else to be defined. I have nearly 20 years of enterprise data center infrastructure systems engineering experience and nearly 20 years of hands on linux engineering experience, the heart and framework of any HPC / AI Infrastructure environments.

While I was at NC State University's ECE (Electrical and Computer Engineering) Department, I was responsible for:

- Ensuring backbone system stability of enterprise scale environments. Restoring AI/ML infrastructure workloads to university research department. Enabled thousands of students, from undergrad to graduate to doctorate dissertations deadlines, research professors to publish papers, research labs to run experiments, and faculties to secure multi million dollar grants.
- Resolving kernel panics, administration of computer labs with Nvidia gpu workstations for teaching AI/ML. Repaired hardware systems from vendor lineup marketed for AI/ML compute workloads from dell, lenovo, supermicro, red barn HPC, and other lesser known vendors.
- Ensuring optimal performance of linux-based enterprise infrastructure. Administration, system maintenance, OS upgrades, patches, and security enhancements, backups, monitoring of enterprise linux infrastructure. Management of system backups and data archivals of systems' life cycle. Management user accounts, permissions, and access to linux systems. Technical support to end-users, and resolving hardware and software-related issues.

I performed a bare metal restore and rebuild of a \$70K+ deep learning AI/ML supercomputer, Nvidia DGX workstation out of vendor contract support, reinstalled vanilla base os, system drivers and firmware, gpu cuda and toolkit drivers, management and utility layer api's, restoring the system back to the DGX platform configuration, verified and addressed all system issues and placed back into production use.

I also redesigned and rebuilt the cisco/arista networking teaching lab for the dept of electrical and computer engineering at NCSU.

While I was at Cisco, I supported and maintained the high speed network aspect of AI/ML infrastructure. A networking lab of high-speed switches, network security equipment and appliances. Supported the staging and validation of engineering POC Testbed's. Management of Cisco UCS C-series rack mount servers and Cisco UCS Blade Centers chassis, 8k chassis encompassing both servers, network modules, and line cards, ASR5000/5500 chassis, IOS, IOS XE,

I design, integrate, implement, deploy, optimize, and resolve issues in data center environments. I live and breathe enterprise data center infrastructure environments, revived server farms/AI/ML compute nodes, engineered SAN/NAS environments, managed enterprise linux infrastructure, ensured stability in AI research infrastructure, rebuilt a AI/ML deep learning supercomputer (Nvidia DGX workstation), handled data center operations for nearly 20 years, supported thousands of users, thousands of enterprise racks and cabinets of compute, storage, networking, virtualization equipments, environments, and frameworks, seen, survived, and done nearly everything in the enterprise data center infrastructure environment.

I'm infrastructure first-battle tested and hardened, knowledgeable in multiple domains, architecturally mature, comfortable with at scale operations, reliable under pressure, and certainly HPC/AI capable.