sighpc

ACM Special Interest Group on High Performance Computing

EDUCATING AND TRAINING STUDENTS FOR ROLES AS HPC SYSTEMS PROFESSIONALS

Stephen Lien Harrell

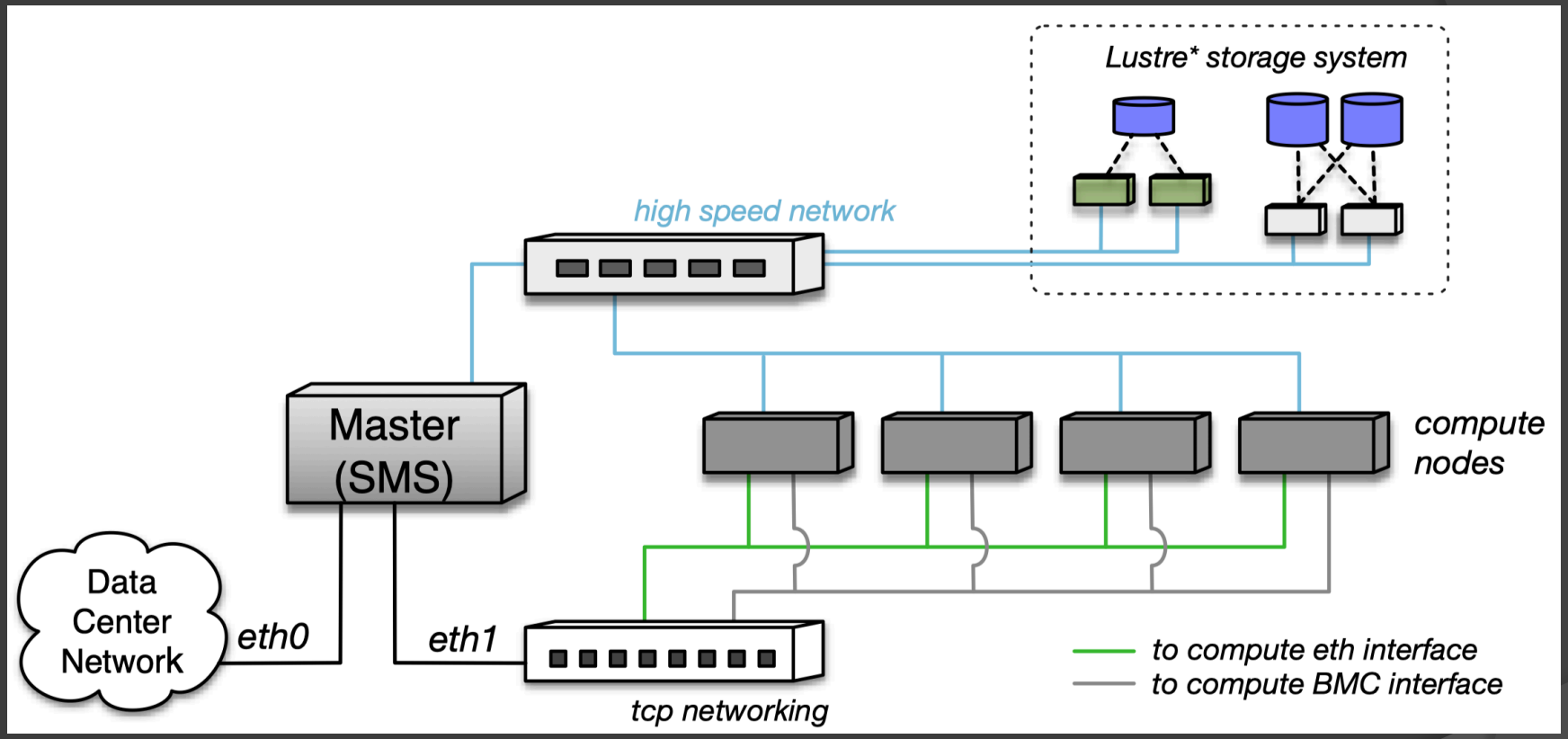
Purdue University

Secretary, SIGHPC System Professionals Virtual Chapter

Stephen Lien Harrell - Background

- Currently
  - Senior Research Analyst at Purdue (Support Staff in Research Computing)
- Previously
  - Senior HPC System Administrator at Purdue
  - System Administrator at many companies including Google and Shutterfly
- Published work on HPC System Admin training and education
  - Mentoring Undergraduates into Cyber-Facilitator Roles – PEARC 18
  - Linux Clusters Institute Workshops: Building the HPC and Research Computing Systems Professionals Workforce – HPCSYSPROS 17
  - Student Cluster Competition: A Multi-disciplinary Undergraduate HPC Educational Tool – EDUHPC 15
  - Methods of Creating Student Cluster Competition Teams – Terragrid 11

Typical Architecture of a Modern HPC Cluster



Schulz, K. et al, Cluster computing with OpenHPC. HPCSYSPROS16

HPC Systems Professional Roles

- 4 Major Roles
  - System Administration
    - Responsible for node OS software, cluster deployment, configuration management, etc
  - Network Administration
    - responsible for fast networks - Inifiniband, top end Ethernet, etc
  - Storage Administration
    - Responsible for high speed distributed file systems, usually multiple
  - Datacenter and Hardware Administration
    - Responsible for power, cooling, and computer, storage and network hardware
- Any given HPC System Professional will have at least 1 of these roles but usually multiple

HPC vs Non-HPC roles

- System Administration
  - Non-HPC
    - Linux Administration, understanding of basic network principles, POSIX standard, service specific details (databases, web servers, dhcp, dns, etc)
  - HPC
    - Additionally, high speed network driver configuration, distributed storage clients, schedulers, configuration management and deployment at scale
- Network Administration
  - Non-HPC
    - Router administration, routing protocols, cabling, network hardware maintenance, advanced networking concepts
  - HPC
    - Additionally, RDMA networks, Infiniband

HPC vs Non-HPC roles

- Storage Administration
  - Non-HPC
    - Storage Appliances, NFS, CIFS, etc
  - HPC
    - Parallel Filesystems (Lustre, GPFS), Filesystems over RDMA, often tape archival (HPSS)
- Datacenter and Hardware Administration
  - Largely the same concerns of any high-density datacenter at scale

HPC is Interconnected

- Modern HPC is about clustering and the connections between a myriad of technologies

- In order to be a proficient at HPC system administration
  - Understand each component (systems, network, storage)
  - Understand how each system interacts with the others
  - Understand how failures and partial failures affect the whole system
  - Understand how MPI and other parallel applications work

Finding System Administrators

- Common requirements for entry-level typical system administration roles
  - BS in CS, IT, Tech, etc or
  - Certifications such as RHCE or CCNA or
  - Demonstrated practical experience
- These requirements are typical for HPC System Administrators as well
  - For entry-level HPC Administrators it is difficult to find someone with HPC experience
  - Many organizations use an apprenticeship model

Current Efforts for Training HPC System Administrators

- Training/Educating Students
  - Homegrown training programs at universities
    - Typically exists within the HPC center and is run by center staff
  - Student Cluster Competitions
    - International: SC, ISC, ASC
    - Regional: OHPCC
- Training Existing Professionals
  - LCI (Linux Cluster Institute)
    - Includes beginner and advanced trainings
    - Many organizations use this to help train junior HPC admins
  - HPCSYSPROS
    - State-of-the-practice workshops for existing HPC Systems Professionals co-located with SC and PEARC

Homegrown Training at Purdue

- Reasons
  - Hard to get experienced HPC System Administrators to move to West Lafayette, IN
  - No HPC System Administration classes on campus
  - Growing HPC need
- Mentoring and Training on site
  - Primarily use the apprenticeship method
  - Target undergraduates
  - Basic skills for a student system administrator interview
    - Know and can apply knowledge about computer components (put together a desktop computer)
    - Know basic linux commands, have installed it at least once
  - Graduated responsibilities
    - Hardware breakfix -> Small admin tasks - > Scripting non-critical systems -> Modify config management on critical systems
    - If a student gets to the last step they are ready for hire as an HPC system administrator
- Student Cluster Competition (SCC) classes
  - Use the SCC as a carrot for staff to get exposure to engaged undergraduates and recruit them
  - Approaches that have been used
    - teaching scientific application usage alongside the system administration – different tracks at an undergraduate level
    - Survey class where all students build clusters from parts, then run scientific application on the cluster and visualize

Student Cluster Competitions

- Rules for SC flavor
  - Team of 6 undergraduates
  - 3000 watts
  - Students bring cluster with help of center staff/professor and vendor
  - Students compute 4 scientific problems on the exhibit floor of SC
  - Other SCCs are similar
- Outcomes
  - Typically one or two students are responsible administration of the cluster
  - Students learn about using HPC as well as simplistic cluster administration
  - Great introduction and preview of possible roles at HPC centers
  - Primarily for getting feet wet and inspiring students to pursue careers in HPC

Existing Professional Training

- LCI
  - Beginner training aimed at basic configuration management and services around clustering
  - Advanced training around important HPC technologies such as parallel filesystems and advanced scheduler configuration
- HPCSYSPROS
  - Accept state-of-the-practice papers and digital artifacts (scripts, config management, etc)
  - HPC technology moves quickly and this is a forum for finding out how new technologies are used at other sites

Developing Pathways

- HPC SYSPRO REU? (not exactly research)
- Putting it in the curriculum
  - Linux System Admin
  - Introduction to HPC
  - HPC SYSPRO class after the two above
- Invite system administrators for large systems to talk about interesting points

Questions and Discussion

- How does this fit into the overall HPC Education Ecosystem
- Should training/education be formalized and move beyond a primarily apprenticeship model?