Building the HPC Workforce: RMACC's Cohort Program for System Administrators

Shelley Knuth
Research Computing
University of Colorado
Boulder, CO, USA
shelley.knuth@colorado.edu

Craig Earley
University of Colorado
Boulder, CO, USA
Craig.Earley@colorado.edu

Brandon Reyes
University of Colorado
Boulder, CO, USA
Brandon.Reyes-2@colorado.edu

Kyle Reinholt
University of Colorado
Boulder, CO, USA
kyle@Colorado.EDU

Jan Mandel
University of Colorado
Denver, CO, USA
Jan.Mandel@ucdenver.edu

Mitchell McGlaughlin
University of Northern Colorado
Greeley, CO, USA
Mitchell.Mcglaughlin@unco.edu

Jarrod Schiffbauer
Colorado Mesa University and
University of Colorado Boulder
Grand Junction, CO, USA
jschiffbauer@coloradomesa.edu

Joel Sharbrough
New Mexico Tech
Socorro, NM, USA
joel.sharbrough@nmt.edu

Abstract

As demand for computational power grows, access to advanced cyberinfrastructure (CI) and the professionals needed to support it has become increasingly critical. However, hiring and retaining cyberinfrastructure professionals (CIPs) remains a significant challenge, as many trained in enterprise services lack the specialized skills required for advanced CI administration. To address this gap, a student cohort program was implemented via the Rocky Mountain Advanced Computing Consortium (RMACC), providing hands-on training in CI administration. The program included weekly virtual sessions, mentorship, and two in-person experiences where students participated in real-world system deployment and decommissioning tasks. Students gained practical experience in Slurm configuration, Linux proficiency, hardware procurement, and system troubleshooting. This initiative has proven highly successful, offering professional development opportunities and expanding the pipeline of skilled CIPs. The results emphasize the importance of integrating CI administration education into research computing programs to ensure the sustainability and growth of advanced CI support.

CCS Concepts

• **Social and professional topics** → **Computing occupations**; *Employment issues*; **Informal education**.

Keywords

workforce development, system administration, informal training

ACM ISBN 979-8-4007-1398-9/25/07
https://doi.org/10.1145/3708035.3736068

ACM Reference Format:
Shelley Knuth, Craig Earley, Brandon Reyes, Kyle Reinholt, Jan Mandel, Mitchell McGlaughlin, Jarrod Schiffbauer, and Joel Sharbrough. 2025. Building the HPC Workforce: RMACC's Cohort Program for System Administrators. In *Practice and Experience in Advanced Research Computing (PEARC '25), July 20–24, 2025, Columbus, OH, USA*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3708035.3736068

1 Introduction

The Rocky Mountain Advanced Computing Consortium (RMACC) is a group of thirty-nine institutions in the Rocky Mountain region that have a general interest in sharing information on high performance computing (HPC), large-scale data issues, and other topics associated with supporting research utilizing advanced cyberinfrastructure (CI). RMACC members are located at various institutions in Colorado, Arizona, Utah, Wyoming, Washington, Idaho, New Mexico, Nevada, and Montana. These institutions include R1 universities, institutions within Established Program to Stimulate Competitive Research (EPSCoR) jurisdictions, Primarily Undergraduate Institutions (PUIs), national laboratories, Hispanic Serving Institutions (HSIs), colleges serving primarily rural communities and two community colleges. The consortium is very active, with an annual symposium that attracts several hundred people, regular proposal collaborations, and working group participants. The barrier for membership is low, and RMACC continues to grow. A history of strong collaborations resulting in National Science Foundation (NSF)-funded grants has been a benefit of RMACC (#1659425, #19255766).

The University of Colorado Boulder (CU) Research Computing (CURC) group supports advanced CI in the form of a supercomputer, large-scale data storage, user-contributed condo cluster, a secure research compute enclave, the campus cloud broker service, trainings, consultations, and tools/gateways for users. In 2017, CURC launched "RMACC Summit", which was funded by NSF (#1532236) and provided 23% of its compute cycles to Colorado State University (CSU), who were partners on the project, and 10% of its cycles to members of RMACC [1]. While RMACC had been in existence for approximately seven years at that point, the use of a common system solidified the connection points within the consortium. The total CPU hours consumed by RMACC users on Summit increased by 150% between January 2020 and December 2022, with the total CPU hours used by RMACC users in 2022 was 6,813,084. The number of RMACC Summit users also nearly tripled during that time period, going from 26 active users in 2020 to 77 in 2022. The availability of advanced CI via Summit was critical for many institutions in RMACC who did not have on-premise computing to support their research or classes.

As RMACC Summit moved to its end-of-life stage, a new system, Alpine, was built through investments from CU Boulder, CSU (#2201538), and the University of Colorado Anschutz Medical Campus (AMC). A recent NSF funded project (#2322260) has provided compute and storage to members of RMACC, continuing to support research in the region. Since the "CC* Regional Computing: Enhancing Computing At Regional Schools in the Rocky Mountain Advanced Computing Consortium (RMACC)" project began in 2023, we have continued to see growth in usage. From January 1 through December 31, 2023, 206 new users got accounts. During that same time period in 2024 that number was 80. The number of CPU hours consumed by these users was 10,637,246 in 2023 and 22,002,666 in 2024.

Because of the availability of this infrastructure, significant research projects are able to continue. The research efforts of co-authors Mandel, McGlaughlin, Sharbrough, and Schiffbauer leverage high-performance computing resources for diverse scientific advancements. Mandel's team develops coupled atmosphere-fire models using NASA funding and supercomputers like NCCS Discover. McGlaughlin's lab applies genomics to plant conservation, identifying new species, advising federal agencies, and training graduate students in HPC research using RMACC resources. Sharbrough's lab studies cytonuclear co-evolution, utilizing Alpine's computational power to analyze the genomic architecture of photosynthetic and respiratory performance. Schiffbauer's group engages undergraduates in molecular dynamics simulations using RMACC resources, supporting microgravity experiments, interfacial conductivity studies, and machine learning applications, preparing students for STEM careers. These collaborations drive computational research forward while fostering student training and scientific innovation.

2 Workforce Development for HPC System Administrators

As the need for increased computational power grows, so does the ability for researchers to be able to access advanced CI to conduct their research. Simultaneously, the continued development and influx of cyberinfrastructure professionals (CIPs) to support the advanced CI being used for research is critical to keeping these systems active and available. One common struggle organizations that provide advanced CI have are the challenges faced when hiring CIPs. Network, storage, and systems administrators who have been trained on enterprise services typically have not been taught how to apply those skills to advanced CI. As a result, the number of CIPs who are available to support computational research utilizing advanced CI is limited, with positions remaining vacant and heavy turnover among teams. Often the CIPs who have training in advanced CI will change positions between peer groups, which further limits the expansion of the field. Maturation of professional knowledge from different disciplines, organizations, or institutions is therefore limited.

The need for more skilled CIPs was reflected in the results of a survey sent out to the Association for Computing Machinery (ACM) Special Interest Group on High Performance Computing (SIGHPC) Systems Professionals (SYSPROS). The survey gained 54 responses from a variety of institutions including academic, national labs, non-profits, and industry partners [2]. When asked the question, "How easy is it to find staff to hire?" 76% of the respondents indicated it was "hard" to "very hard". The teams need to grow these individuals within their group through on-the-job training, which is a sentiment reflected by many in the community.

Administration of advanced CI is often overlooked as an important component of research support at educational institutions. A large barrier often encountered by researchers who want to use CI is how to schedule and manage jobs. This barrier can prevent researchers from using CI effectively or provide a daunting task that distracts them from their research. To this end, it is important to have CIPs who help researchers utilize the provided resources. However, without CIPs who have an educational background in administering advanced CI, these systems will be underutilized or used inefficiently. In order to support research in science and engineering effectively, educational opportunities to train future CIPs on the specifics of administering advanced CI is critical.

To coincide with the timing of the equipment purchases, we also implemented a student cohort program where we invited several students from institutions across RMACC to shadow CURC system administrators and datacenter staff during the deployment and configuration of the new equipment. This hands-on cohort program was comprised of weekly virtual sessions during the academic year where the system administrators educated the cohort about what goes into the decision on what equipment to purchase, how to work with datacenter staff to deploy the equipment, and the basics of system administration, including Slurm configuration and a level-set of Linux skills. Two in-person experiences included hands-on work in the datacenter both in deployment of the new nodes and the decommissioning of the previous system (Summit). In addition, a significant emphasis of mentorship and networking has been made to ensure professional development for the cohort members. This cohort has proven to be very successful, with strong positive testaments from the cohort members.

2.1 Description of Program

Since the inception of this project, there were weekly remote educational and workshop sessions leading to two in-person experiences in January and August 2024. As part of the student educational sessions in the fall and winter of 2023-24, RC staff provided several angles on considering the node order being made as part of the equipment add-on, with an attempt to simulate the process as fully as possible. The students were provided a sample quote for high-memory nodes, abstracted and modified to avoid sharing confidential information privileged to CU and Dell. They were informed as to the parameters and process for submitting and quoting orders. Finally, as an exercise, they were instructed to compare the mock quote to the specs described in the award and to provide other feedback. In a follow-up session, students were tasked with building an example specification using the lessons from building the quote. They then presented their options to the group to consider and discuss the alternatives they produced. The order was then prepared and submitted by CURC staff following the discussions in the group setting.

The deployment of the equipment (2048 cores and 1.344 PB raw capacity expansion of scratch storage) involved shared responsibilities between the students and staff members due to existing university procedures and requirements. CU Boulder's Data Center Operations team performed physical installation, followed by the Network Engineering Operations team enabling network connectivity. Finally, an RC system administrator with privileged access performed the release of a base operating system to the hardware. Students took over in the January and August in-person sessions. Each was assigned a set of nodes with the base operating system. They were tasked with running an Ansible playbook to bring the nodes into alignment with the remainder of RC's hardware. They worked in groups of two or three. Each group was responsible for troubleshooting, identifying problems or inconsistencies, and reporting progress to the rest of the group. Nodes were put into production by RC staff with privileged access following the conclusion of the summer 2024 in-person experience.

Each cohort had a focus on learning hands-on skills for deploying and decommissioning equipment in an active datacenter. The cohorts worked closely with staff and each other to learn skills that they can take into their future careers. Besides a transfer of knowledge, cohort members met with other CURC staff to discuss career and networking opportunities. An outline of the topics they engaged in are as follows:

- Introduction to HPC and system administration through recurring student sessions
- Introduction to vocabulary and concepts in computing and system administration
- Establishment of general Linux proficiency (terminal)
- Running of jobs on a Slurm HPC cluster - (existing experience varied)
- Slurm concepts and implementation - database, controllers, daemons; commands relevant to each; reporting and some troubleshooting; node health checks and automated testing
- Build specifications for hardware requested for expansion of Alpine HPC cluster
- Evaluate example quotes for reference
- Sharing of HPC community resources and venues for professional networking (e.g. CaRCC, ACCESS)

January in-person experience

- Hands-on work in two production data centers - the High-Performance Computing Facility (HPCF) and the Space Science Center (SPSC)
- Decommissioning of the Summit HPC cluster, hands-on, included removing cables, removing from racks, stacking
- Reporting and metrics of the Slurm cluster; teamwork for information gathering and troubleshooting

July-August in-person experience:

- Meet and talk with RC staff about system administration and user support positions
- Provisioning of hardware and rollout of system libraries and configuration using common HPC tools (Ansible, pdsh, clush) in collaboration with system administration team lead (Earley) and user support analyst (Reyes)
- Review resumes and provide direction on seeking professional opportunities in HPC

3 Future Work

To date, this project has been very well received, with several cohort members returning for both cohorts. In total, we had a mixture of 17 graduate and undergraduate students from eight RMACC institutions. Feedback has been very positive from the cohort and already has had a positive impact as the RMACC cohort experience has assisted two participants with being selected for student positions in HPC. As a result, one final cohort experience is planned for late July/early August 2025. This session will look a bit different than previous sessions for a variety of reasons. First, as the equipment for this project has already been purchased, our team is currently working through a variety of options for the on-site work. We are purchasing equipment for other projects, and are attempting to coincide the in-person experience with the arrival of that equipment. In addition, we are revamping our datacenter's heating and cooling system and will work with our data center team to maximize opportunities for cohort involvement in this process. Second, we are planning to open this session up to a much broader community and have this cohort be significantly larger with close to 30 participants. We want to take what we have learned from this regional experience and open it up to the science and engineering community as a whole. We have received multiple requests from participants outside of RMACC to attend, and for this last iteration, our goal is to accommodate as many individuals as possible. While there will be opportunities for pre-workshop learning engagements, we plan to hold the bulk of the effort on-premise and expand opportunities for mentoring. As in our previous cohort experiences, participants' travel is covered to allow for as much participation as possible.

There are also plans to add a badge on LinkedIn for students who participated in this cohort, by request of students who have already passed through the program. The students desire to have potential employers be aware of their knowledge gained as part of this cohort.

4 Conclusion

This project successfully expanded computing resources for RMACC members while fostering a new generation of HPC professionals. Through hands-on learning and mentorship, students developed critical technical skills and professional networks, ensuring continued growth in the HPC workforce. Other institutions are strongly encouraged to implement similar programs as part of any future infrastructure deployments. At the conclusion of this project, teaching materials will be provided to the community to aid in developing these programs. Future efforts will build on these achievements to further support RMACC's research community and expand workforce development initiatives.

Acknowledgments

This work would not be possible without the members of the CURC team, the CU Boulder datacenter staff, and the CU Boulder networking staff. The authors in particular want to acknowledge the work of Lori Flora in supporting the logistics of these efforts. This work was funded by the National Science Foundation Office of Advanced Cyberinfrastructure, #2322260.

References

[1] Jonathon Anderson, Patrick J. Burns, Daniel Milroy, Peter Ruprecht, Thomas Hauser, and Howard Jay Siegel. 2017. Deploying RMACC Summit: An HPC Resource for the Rocky Mountain Region. In *Practice and Experience in Advanced Research Computing 2017: Sustainability, Success and Impact* (New Orleans, LA, USA) *(PEARC '17)*. Association for Computing Machinery, New York, NY, USA, Article 8, 7 pages. doi:10.1145/3093338.3093379

[2] Jenett Tillotson. 2023. HPC SYSPROS Member Survey 2023. doi:10.5281/zenodo.8349359