

# Building Scalable and Inclusive Foundations for HPC: Lessons from UC Merced's Introductory HPC Training Program

Yue Yu

University of California, Merced  
yyu49@ucmerced.edu

## ABSTRACT

High-performance computing (HPC) is becoming essential across a broad range of disciplines, including those historically underrepresented in computational research, such as sociology, psychology, and the arts. To reduce barriers to entry, the University of California, Merced (UC Merced) developed a 90-minute introductory HPC workshop designed for participants with no prior technical background. The workshop includes a theoretical overview of campus clusters, fundamental Linux commands, and core HPC concepts, followed by a hands-on session where participants connect through SSH and browser-based tools, load software modules, and submit jobs to institutional HPC resources using Slurm. Delivered in a hybrid format with both synchronous and asynchronous learning materials, the program has been offered more than 20 sessions since 2021. It has primarily served students (75.7%), faculty (16.2%), and staff. Post-workshop surveys indicate that 83% of participants are more likely to incorporate HPC into their research after attending, contributing to a doubling of active HPC users on campus since the program's launch. This scalable and inclusive model provides an effective framework for expanding HPC adoption and fostering computational engagement across disciplines.

## KEYWORDS

HPC Training, Workshop Design, Hybrid Learning, Student Engagement

## 1 INTRODUCTION

High-performance computing (HPC) has become a critical component across a wide range of disciplines. It allows researchers to simulate complex systems where traditional experiments are infeasible, analyze massive datasets that exceed the capabilities of personal computers, and develop advanced computational tools to improve predictions and decision-making. In recent years, there has been rapid growing interest in extending HPC to fields historically underrepresented in computational research, such as social sciences, humanities, and psychology. However, adoption in these fields remains limited for several reasons. First, there is a lack of foundational training materials designed for researchers without technical backgrounds, which can make HPC seem intimidating or out of reach. Second, awareness of available HPC resources is often

low, especially in disciplines that historically have not used computational approaches. Third, there are cultural and methodological barriers that researchers may not see computational methods as aligned with their field, or they have the perception that HPC is overly complex or reserved for specialized fields.

To address these challenges, the University of California, Merced (UC Merced) developed a 90-minute introductory HPC workshop aimed at lowering barriers and making computational resources more approachable for researchers from diverse backgrounds. Based on the principles of inclusive pedagogy and experiential learning, the workshop blends conceptual instruction with hands-on practice. A key enhancement is a virtual tour of the campus data center, which allows participants to visualize the physical infrastructure behind the HPC system, fostering familiarity and reducing the intimidation often associated with large-scale computing facilities. This approach ensures that attendees not only gain a foundational understanding of HPC concepts but also build confidence and practical skills that they can immediately apply to their own research.

This paper presents the workshop as a case study in effective HPC onboarding, detailing its structure, hybrid delivery format, and support strategies. We evaluate its impact on broadening HPC participation among historically underrepresented disciplines and regions. By placing this initiative within the wider landscape of HPC education, we aim to provide a replicable model that other similar institutions can adopt to democratize access to advanced computational resources and foster a more diverse research community.

This paper presents the workshop as a case study in effective HPC onboarding, detailing its structure, hybrid delivery format, and support strategies. We evaluate its impact on broadening HPC participation among historically underrepresented disciplines and regions. By placing this initiative within the wider landscape of HPC education, we aim to provide a replicable model that other similar institutions can adopt to democratize access to advanced computational resources and foster a more diverse research community.

## 2 WORKSHOP DESIGN

The UC Merced introductory HPC training program was designed to lower barriers for new users, particularly those from non-technical or historically underrepresented disciplines. The program is grounded in educational best practices for teaching complex technical concepts to diverse audiences, emphasizing active participation, inclusivity, and confidence-building.

### 2.1 Educational Principles

The workshop includes several core educational principles to ensure both effectiveness and accessibility.

- **Active Learning:** Participants engage directly with HPC concepts through hands-on practices where they will be able to use UC Merced HPC to complete the tasks.
- **Peer Support:** Trained student assistants provide real-time guidance, fostering a collaborative and supportive learning from their peers.
- **Scaffolded Learning:** Concepts are introduced progressively, allowing participants to build knowledge step-by-step without becoming overwhelmed.

- **Hybrid Teaching Model:** Combines in-person sessions with asynchronous learning materials, ensuring participants have ongoing access to resources. This approach reduces barriers to entry and allows learners to review content at their own pace, supporting deeper understanding and retention of new concepts.

## 2.2 Workshop Structure

The workshop begins with a theoretical introduction to HPC, starting with an overview of what HPC is and how it differs from other scales of computing, such as personal or lab-scale desktops. Participants are introduced to fundamental terminology, including definitions of cores, CPUs, and nodes, to help them become familiar with the language and concepts that will be used throughout the session. Once these key terms are understood, the discussion moves to hierarchical computing structures, illustrating how these components fit together in modern HPC systems.

Next, participants are introduced to UC Merced's on-premises HPC clusters, including their total number of nodes and cores, to demonstrate the scale of computational resources available on campus. By comparing this to personal computing capabilities, attendees gain motivation to leverage HPC for accelerating their research. The session then covers the general architecture of an HPC system, explaining key components such as login nodes, compute nodes, SSH access, and high-speed interconnects like InfiniBand. This provides participants with a clear picture of how different parts of the cluster work together to execute programs.

To connect theory with practice, the workshop includes a virtual tour of the campus data center, giving participants a view of the physical layout of servers and supporting infrastructure. This visual component helps demystify HPC by illustrating how abstract concepts translate into real-world hardware, fostering confidence that these systems are both approachable and accessible. This concludes the theoretical portion of the workshop, highlighted in purple in Figure 1.

After the theoretical section, the workshop moves into a hands-on component (shown in green in Figure 1), where participants actively interact with the HPC system. To reduce cognitive load, each participant receives a Linux command cheat sheet containing only the commands relevant to the session. This focused approach prevents users from being overwhelmed by extraneous details and helps them concentrate on applying commands in context. Throughout the exercises, participants can refer to the cheat sheet as needed, allowing them to stay engaged without worrying about memorization or losing track during the session.

Participants use guest accounts created specifically for the workshop. These accounts are configured with restricted permissions and access only to designated public queues, ensuring that training activities do not disrupt overall cluster operations. Each guest account includes a preconfigured home directory with folders and files prepared for the practice session, streamlining the experience and removing setup barriers. Temporary passwords are provided so participants can log in via SSH, offering an authentic introduction to remote HPC access.

In addition to SSH, participants are introduced to Open OnDemand [4], a browser-based interface for interacting with the cluster.

This tool allows them to launch a terminal directly from their web browser using their guest credentials, further lowering technical barriers by removing the need to install or configure terminal software locally. This dual-access approach supports participants with different technical skills and resources, ensuring an inclusive experience.

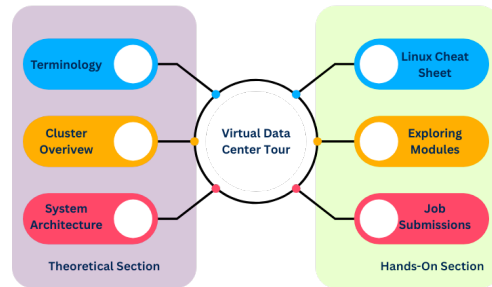


Figure 1: The general workflow of the workshop.

Once connected, participants start by exploring the basic cluster environment. They are introduced to commands such as 'module avail' and 'module load', which allow them to view and access software modules already available on the cluster. This step encourages participants to develop the habit of checking for pre-installed resources before building or installing their own, promoting both efficiency and reproducibility in their workflows.

After exploring the cluster environment, the session transitions to running programs and submitting Slurm [9] jobs onto institutional HPC cluster, introducing participants to the concept of job scheduling and resource management. They begin with serial job exercises using simple, pre-configured scripts. The process starts by examining the contents of a basic Slurm job script with the 'cat' command. This method allows participants to view the script directly in the terminal without the added complexity of using a text editor. Special attention is given to lines starting with '#SBATCH', which define job parameters such as runtime, memory allocation, and partition selection. This helps participants understand how Slurm works and how to properly request computational resources. After the script review is complete, participants submit their first job using the 'sbatch' command and monitor its progress with 'squeue'. Observing job behavior and output file generation in real time reinforces the connection between Slurm commands and the underlying cluster operations. This hands-on practice builds confidence as participants see their jobs successfully run on the system.

Building on this foundation, the session advances to parallel job exercises, introducing more complex workflows. Here, participants work with scripts that submit multiple serial jobs within a single Slurm script. They are introduced to loop syntax in Bash, which allows them to iterate over tasks and automate multiple job submissions seamlessly. These exercises include node count, core allocation, partition settings, runtime limits, and environment variables. By working through these examples, participants begin to understand how to scale their computations and optimize resource use.

This structured progression begins with basic Linux navigation, continues with single-job submissions, and culminates in automated multi-job workflows, offering participants a practical and accessible introduction to HPC job management. By the end of the session, they had gained both the knowledge and the confidence to independently perform real-world research computations on the cluster.

### 2.3 Delivery Format

The workshop is designed as a 90-minute session, divided evenly between two parts: the first 45 minutes focus on the theoretical introduction, followed by 45 minutes of hands-on practice. This duration was chosen to align with both the workshop content and the average human attention span reported in educational research [1, 7]. Keeping the session under two hours helps prevent participant fatigue and avoids overwhelming new users while still providing sufficient time for meaningful learning.

To maximize accessibility and support long-term learning, the workshop is offered in both synchronous and asynchronous formats. Live sessions give participants the opportunity to interact with instructors and student assistants, ask questions, and receive real-time guidance. Afterward, all materials are made available through an open-access HPC documentation website [8].

The documentation site has been in continuous development since 2022. It was originally built with Docsify [3] but later migrated to Docusaurus [5], a modern framework developed by Meta. Docusaurus was chosen for its native support of React components, offering greater customization, flexibility, and scalability for creating interactive documentation. This transition also follows best practices recommended by other universities [2], ensuring the platform remains sustainable and user-friendly for a broad research audience.

The updated site now includes a dedicated training materials section (Figure 2) and a blog page for community updates, announcements, and job opportunities, helping participants stay engaged beyond the workshop. It also provides essential reference materials such as cluster configurations, command explanations, and troubleshooting guides, allowing users to revisit content at their own pace. Training slides, example code, and step-by-step instructions are included to guide learners through transferring exercise files, running practice jobs, and reinforcing key concepts. This structure enables participants to continue developing their HPC skills independently long after the session has concluded.

### 2.4 Student Assistant Involvement

Currently, an undergraduate student serves as a Student Technology Consultant (STC) in the Office of Information Technology (OIT) at UC Merced. This student has contributed significantly to the migration and redevelopment of the HPC documentation website and material development, gaining valuable experience in web infrastructure and user support. Beyond documentation, they have been trained to lead hands-on practice sessions during workshops, participate in weekly HPC office hours, and assist with troubleshooting user issues through the ServiceNow ticketing system.

Through these experiences, the student has advanced from a novice HPC user to an independent and highly skilled practitioner.

This growth is demonstrated by several notable accomplishments, including leading UC Merced's first student team in the IndyCC competitions at SC24 and SC25, as well as completing two summer internships with HPC and computational science groups at Los Alamos National Laboratory, further deepening their expertise.

Inspired by this success, we plan to expand student involvement by recruiting and training additional undergraduate assistants. To foster a sustainable, peer-supported learning environment, we recently launched a Slack platform dedicated to the campus HPC community. This space is student-led and designed to encourage open discussion, promote training opportunities, and facilitate resource sharing. Our goal is to build a collaborative community where students not only receive support but also take on leadership roles in advancing HPC engagement and knowledge-sharing across campus.

## 3 IMPLEMENTATION AT UC MERCED

### 3.1 Program History

The Introductory HPC Workshop was first launched in late 2021 as part of UC Merced's broader strategy to expand access to research computing resources and support the university's transition to R1 status. As UC Merced grows its research portfolio, empowering faculty, students, and staff with advanced computational skills is essential to fostering a more active and data-driven research community.

Since its launch, the program has steadily expanded, offering more than 20 sessions to participants across diverse disciplines. By delivering accessible, foundational HPC training, the workshop has enabled researchers to integrate advanced computational methods into their work, directly contributing to the growth of UC Merced's research portfolio. These initiatives have significantly enhanced the university's research capacity and were instrumental in supporting its achievement of R1 status in February 2025. Looking ahead, the program will continue to promote computational literacy and foster collaboration, ensuring that UC Merced not only maintains its R1 designation but also strengthens its role as a hub for cutting-edge research in California's Central Valley.

### 3.2 Workshop Schedule and Requests

We offer the HPC workshop on two fixed dates each year. The first session takes place during UC Merced Research Week, typically held in March, when the campus highlights active research initiatives. As part of this event, we host an in-person data center tour alongside the workshop to showcase the HPC facilities and demonstrate their role in supporting cutting-edge research. The second fixed session is held during Graduate Orientation Week (GROW), usually in August, when the campus welcomes new graduate students. Hosting the workshop at this time introduces incoming researchers to research computing early in their programs, encouraging them to integrate HPC resources into their work from the start.

In addition to these fixed sessions, we also conduct on-demand workshops. Requests can be submitted through the ServiceNow ticketing platform using the request form linked on the HPC documentation website. To maintain the quality of the hands-on learning experience, we require a minimum of three participants and limit each session to 25 attendees.

The screenshot shows a web browser displaying the 'Introduction to HPC Workshop' page. The page has a blue header with the UC Merced and OIT logos. The main content area has a blue background with white text that reads: 'Introduction to High-Performance Computing (HPC)'. Below this, it lists 'Yue Yu Sr. Research Computing Facilitator' and 'Alex Villa Student Tech Consultant'. Underneath, it says 'Cyberinfrastructure and Research Technologies (CIRT) Office of Information Technology (OIT)'. At the bottom of the main content area, there is a terminal window showing a file transfer command: 'cp -r /home/yyu49/hpc\_training/script/serial .' and 'cp -r /home/yyu49/hpc\_training/script/parallel .'.

Figure 2: The HPC documentation website.

From these on-demand sessions, we have identified two common request patterns. The most frequent requests come from instructors who wish to integrate the HPC workshop into their courses at the beginning of each semester, particularly at the graduate level. This timing helps students become familiar with HPC early on so they can apply it to data analysis throughout the term. The second most frequent requests come from lab managers, who invite us to deliver the workshop during regular lab meetings. In these cases, the lab meeting is replaced with an HPC training session, providing a convenient and highly relevant learning experience for the entire research group.

### 3.3 Audience Profile

Across the 25 sessions delivered to date, the workshop has averaged 15 participants per session, with attendees consisting of 75.7% students, 16.2% faculty, and the remainder staff (see Figure 3). UC Merced, the most diverse campus in the University of California system, serves a student population in which the majority come from low-income households and are first-generation college students. The university is also designated as a Hispanic-Serving Institution (HSI), with Hispanic students representing more than 55% of the student body. This unique demographic underscores the importance of providing equitable access to advanced research computing resources, creating opportunities for groups historically underrepresented in computational research to engage with cutting-edge technologies.

Notably, 70% of workshop participants reported being completely new to HPC, emphasizing the program's role in onboarding first-time users. Post-workshop surveys further revealed that 83% of attendees felt more likely to incorporate HPC into their research, demonstrating the workshop's effectiveness in fostering computational engagement across a diverse campus community.

Participants particularly appreciated the hands-on format, which offered practical experience with real HPC systems. As one participant noted, "The course was amazing! I learned so much, not just about using the HPC, but also about working in the terminal and understanding the physical structure of the system. For a 1.5-hour presentation, I was very pleased with the content."

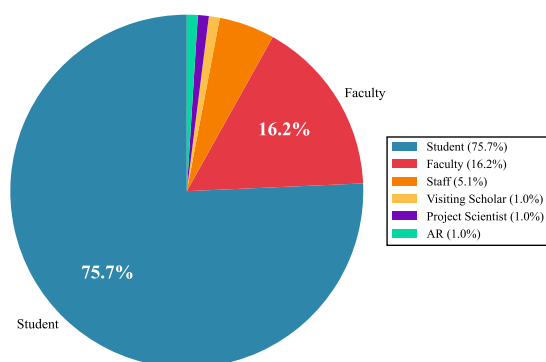
Since the program's launch, the number of active HPC users at UC Merced has doubled, reflecting its strong impact on campus-wide adoption. Collectively, these outcomes highlight the workshop's effectiveness as a scalable and inclusive model for HPC onboarding, advancing computational literacy and expanding research capacity across the university.

## 4 FUTURE DIRECTIONS

The long-term goal of this workshop program is to expand its reach beyond UC Merced, providing accessible and inclusive HPC training to a broader audience across the Central Valley of California.

### 4.1 Regional Scaling and Collaboration

We aim to scale the program to other institutions, including community colleges, California State University (CSU) campuses, and



**Figure 3: Affiliation of workshop attendees at UC Merced.**

additional University of California campuses. The geographic proximity of UC Merced, CSU Sacramento, CSU Stanislaus, and CSU Fresno creates a unique opportunity for collaboration and shared resources. These institutions are not only physically close but also share a common mission as HSI, which strengthens their collective commitment to advancing equitable access to research computing and fostering diversity in computational research.

- CSU Sacramento – Sacramento, approximately 120 miles north of UC Merced
- CSU Stanislaus – Turlock, approximately 40 miles northwest of UC Merced
- CSU Fresno – Fresno, approximately 85 miles south of UC Merced

By fostering partnerships across these campuses, we aim to establish a regional HPC training hub that supports the goals of the NSF award (#2346744) [6]. This hub will provide accessible solutions to empower historically underserved and geographically isolated regions with limited access to computational resources and collaboration opportunities. Through these efforts, we seek to overcome barriers, help researchers effectively leverage advanced computing tools, and foster inclusivity while expanding research capabilities. The hub will also serve underrepresented communities, promote interdisciplinary collaborations, and create pathways for students and researchers to engage with cutting-edge computational technologies and resources.

#### 4.2 Developing a Modular HPC Curriculum

To meet the evolving needs of participants, we plan to expand the workshop curriculum into a modular, multi-level format, covering introductory, intermediate, and advanced HPC topics. The current workshop primarily serves as an entry point, introducing basic HPC concepts and workflows to first-time users. However, as the number

of trained users grows, there has been an increasing demand for more advanced and domain-specific content.

Ongoing efforts include the development of intermediate training materials, focusing on:

- Best practices for parallel programming and workflow scaling.
- Job optimization strategies to improve efficiency and resource utilization.
- Hands-on projects tailored to specific research domains, such as bioinformatics, computational chemistry, and machine learning.

This modular approach enables participants to advance through progressively higher levels of HPC training, fostering essential skill development and building a pipeline of proficient users. By gaining the knowledge and experience to effectively utilize HPC resources, participants can enhance their research productivity and become more competitive in their future academic and professional careers.

## 5 CONCLUSION

UC Merced's introductory HPC workshop has proven effective in lowering barriers to computational research, with over 20 sessions delivered since 2021. The program has successfully engaged a diverse audience, 70% of whom were first-time HPC users, and has doubled the number of active HPC users on campus. Post-workshop surveys show that 83% of participants are more likely to integrate HPC into their research after attending.

Looking ahead, we plan to expand regionally, building a Central Valley HPC training hub in partnership with nearby institutions, and to develop a modular curriculum that supports sustained skill development from introductory to advanced levels. This initiative demonstrates how inclusive, scalable training can grow HPC adoption and empower researchers to leverage advanced computing for impactful, interdisciplinary research.

## REFERENCES

- [1] Scott Freeman, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnedi Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proceedings of the National Academy of Sciences* 111, 23 (2014), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- [2] Ben Godfrey. 2024. Practical Advice for the Creation of Effective HPC User Documentation. In *Proceedings of PEARC '24: Practice and Experience in Advanced Research Computing*. Association for Computing Machinery, 83:1–83:4. <https://doi.org/10.1145/3626203.3670621>
- [3] QingWei Hu. 2023. Docsify: A Magical Documentation Site Generator. <https://docsify.js.org/>. Accessed: 2025-09-05.
- [4] Dave Hudak, Doug Johnson, Alan Chalker, Jeremy Nicklas, Eric Franz, Trey Dockendorf, and Brian L. McMichael. 2018. Open OnDemand: A web-based client portal for HPC centers. *Journal of Open Source Software* 3, 25 (2018), 622. <https://doi.org/10.21105/joss.00622>
- [5] Meta Open Source. 2025. Docusaurus: Open Source Documentation Framework. <https://docusaurus.io/>. Accessed: 2025-09-05.
- [6] National Science Foundation. 2024. NSF Award #2346744: CENVAL-ARC: Central Valley Accessible Research and Computational Hub. [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=2346744](https://www.nsf.gov/awardsearch/showAward?AWD_ID=2346744) Accessed: 2025-09-06.
- [7] Michael Prince. 2004. Does Active Learning Work? A Review of the Research. *Journal of Engineering Education* 93, 3 (2004), 223–231. <https://doi.org/10.1002/j.2168-9830.2004.tb00809.x>
- [8] UC Merced Cyberinfrastructure and Research Technologies. 2025. UC Merced HPC & JupyterHub Documentation. [https://ucm-it.github.io/hpc\\_docs/](https://ucm-it.github.io/hpc_docs/).
- [9] Andy B. Yoo, Morris A. Jette, and Mark Grondona. 2003. SLURM: Simple Linux Utility for Resource Management. *Job Scheduling Strategies for Parallel Processing* (2003), 44–60. [https://doi.org/10.1007/10968987\\_3](https://doi.org/10.1007/10968987_3)