

Next Generation HPC Workforce Development

The Computer System, Cluster, and Networking Summer Institute

Carolyn Connor, Amanda Bonnie, Gary Grider

High Performance Computing Division
Los Alamos National Laboratory
Los Alamos, NM, USA

Andree Jacobson

New Mexico Consortium
Los Alamos, NM, USA
andree@newmexicoconsortium.org

Abstract—Sustainable and effective computing infrastructure depends critically on the skills and expertise of domain scientists and committed and well trained advanced computing professionals. Unlike computing hardware, with a typical lifetime of a few years, the human infrastructure of technical skills and expertise in operating, maintaining, and evolving advanced computing systems and technology has a lifetime of decades [1]. Given that the effective operation and use of High Performance Computing systems requires specialized and often advanced training, that there is a recognized High Performance Computing skillset gap, and that there is intense global competition for computing talent, there is a long-standing and critical need for innovative approaches to help bridge the gap and create a well-prepared, next generation High Performance Computing workforce. This paper places this need in the context of the HPC work and workforce need at Los Alamos National Laboratory (LANL) and presents one such innovative program conceived to address the need, bridge the gap, and grow an High Performance Computing workforce pipeline at LANL. The Computer System, Cluster, and Networking Summer Institute (CSCNSI) completed its tenth year in 2016. The paper presents an overview of the CSCNSI and a summary of impact and success, as well as key factors that have enabled that success.

Keywords—High Performance Computing; HPC; cluster computing; supercomputing; undergraduate summer program; education; workforce development; system administration; networking; cluster administration; computer science; Trinity; curriculum; Supercomputing Challenge; Student Cluster Competition; CSCNSI; staff revitalization; Cluster Boot Camp

I. INTRODUCTION

The High Performance Computing Division at Los Alamos National Laboratory (LANL) manages world-class supercomputing centers in support of the Laboratory's national security science mission. This work includes specifying, operating, and assisting in the use of open and secure high performance computing (HPC), storage, and emerging data-intensive information science production systems for multiple programs. More broadly, LANL, and national laboratories in general, have a strong need for scientists and staff who understand and have the ability to operate, use, and help advance HPC systems and technology.

As an example, the Trinity supercomputer, the most recent supercomputer to arrive at LANL, was designed to exceed 40

Petaflops/s and is a next step toward to goal of achieving exascale computing (a million, trillion operations per second) [1][2]. Trinity is the first of the National Nuclear Security Administration's (NNSA's) Advanced Simulation and Computing (ASC) Program's advanced technology systems. Once fully installed, Trinity will be the first platform large and fast enough to begin to accommodate finely resolved 3D calculations for full-scale, end-to-end weapons calculations. The Trinity system will reside in LANL's Strategic Computing Center in the Nicholas C. Metropolis Center for Modeling and Simulation. The Strategic Computing Center is a 300,000-square-foot building. The vast floor of the supercomputing room is 43,500 square feet, almost an acre in size [3].

Trinity will have at least eight times greater application performance than Cielo, the current NNSA supercomputer sited at LANL. Trinity will introduce the "Burst Buffer" (a storage layer between the compute and storage systems to relieve the bandwidth ingest burden on the parallel file system) [4] and "Advanced Power Management" (adding power adjustment capability to the BIOS) as part of the platform [5]. These technologies will be provided as part of a fully integrated system consisting of compute nodes (>19,000), memory (>2PB), high speed interconnect, and parallel file system (>80 PB of usable capacity) as part of the planning for exascale systems [6].

In order to prepare for, operate, and fully utilize Trinity and its successors, the national laboratories require a workforce with the skills to accomplish the following.

- Manage and operate a highly complex computing infrastructure that includes miles of cable, hundreds of thousands of connections, and tens to hundreds of thousands of storage devices;
- Expertly translate computing industry trends into a robust HPC environment;
- Design and operate the physical infrastructure supporting LANL's computing needs;
- Operate and enable effective use of the full HPC environment;
- Leverage high performance computing systems, processes, and expertise, beyond traditional HPC, to advance innovative Laboratory scientific efforts;

This work was performed at LANL, supported by the U.S. DOE contract DEAC52-06NA25396, and at the New Mexico Consortium, with additional support from the National Science Foundation under awards CNS-1042537 and CNS-1042543 (PRObE).

- Engage in research, development, and state-of-the-art software engineering, supporting development, design, and effective use of increasingly complex large-scale data and computational environments.

While HPC has become a long-standing and essential tool to advance scientific discovery and achieve national goals [7][8][1] and is itself an area of active and advancing research [2][9], there exists a gap between the technical skillsets needed, particularly at the entry level, and the exposure and skillsets provided in current undergraduate (and even graduate level) academic programs. Recognizing this gap and the resulting unrelenting organizational strain due to an inadequately prepared entry-level pipeline, LANL set out just over a decade ago to actively engage to bridge this gap for a specific pipeline need, establishing the Computer System, Cluster, and Networking Summer Institute (CSCNSI) in 2007.

With a proven program with nearly five highly successful years under its belt and an expanded goal to broaden CSCNSI outreach and impact, LANL collaborated with the newly established New Mexico Consortium (NMC) [10] on a joint proposal to the National Science Foundation (NSF) with the result that the CSCNSI received five years of NSF support (2011-2015) as an educational component of PRObE, the Parallel Reconfigurable Observational Environment [11]. PRObE is an NSF-funded low-level, large-scale computer systems research facility located in Los Alamos, NM and housed by the NMC at the Los Alamos Research Park. The primary mission of PRObE is to provide the national systems research community resources in the form of several large-scale testbeds (of 1000+ computing nodes) built from repurposed supercomputers from LANL and dedicated to computer systems research. As such, PRObE is the only facility of its kind in the world.

This paper describes the CSCNSI, a unique undergraduate cluster computing technical enrichment program. In Section 2, we describe two successful outreach programs that share some important attributes with each other and with the CSCNSI while addressing different but related aspects of the HPC education gap. In Section 3 we describe the CSCNSI program itself, including motivation, goals, overview, and program components. Section 4 discusses program results and impact. Section 5 offers a discussion of notable keys to success. Section 6 touches briefly on the dynamic nature of the program. Finally, Section 7 provides concluding remarks, and Section 8 acknowledgments.

II. THREE INNOVATIVE OUTREACH PROGRAMS MOTIVATED BY PERSISTENT PIPELINE GAPS

There is an ongoing need to attract students into STEM fields generally. The need to enhance science education to increase scientific knowledge at all educational levels and fill a national need for well-trained scientists and engineers has been long recognized, dating back to the launch of the space program in the 1950s (and its expansion in the 1960s) and before. This national need has been broadly documented [1][25]. Many have worked tirelessly over the many decades to address it.

One very successful example, the New Mexico Supercomputing Challenge initiated in 1990, is a grass roots operation conceived by LANL scientists and executed through community collaboration in recognition of the need for building a stronger computational workforce. Over its 26 years of unbroken successful operation, the Supercomputing Challenge has educated ~10,000 middle and high school students in computer science, modeling, and simulation. The demonstrable success of the Supercomputing Challenge is due in part to its emphasis on a hands-on approach to imparting skills in these areas. Each student team selects a real-life problem of their choosing; then, over the course of a year, the team walks step by step through the logical analysis required to create a mathematical model that is then translated into computer code and analyzed. Students come out of this program with not only an enhanced set of Computer Science (CS) skills, but also an enthusiasm for scientific inquiry and critical thinking. The program has also had demonstrable success in entraining girls (33%) and minorities (45%), possibly because the program is team-focused and allows students to solve problems that matter to them [12][13][14].

Sensing a gap between a strong need at national labs, academic institutions, and industry to have scientists and staff who understand HPC and its complex intra-topic interconnections and a lack of sufficient educational resources in domain science and computer science undergraduate programs, as well as a deficiency in conveying the interdisciplinary and collaborative nature of the HPC environment within the academic setting, the Student Cluster Competition was developed and debuted at the 2007 Supercomputing Conference in Reno, NV. While not a LANL initiative, its conception was informed by the CSCNSI through early conversations between founder Brent Gorda and CSCNSI co-founder, Gary Grider, and LANL staff continue to actively support and participate in various capacities. The goal of the annual Student Cluster Competition is to expose undergraduate and high school students to HPC by integrating HPC education into undergraduate curriculum and across disciplines. Central aspects of this program are its hand-on nature and its structure as a competition of teams [15].

These two successful programs and the CSCNSI have differences. The Supercomputing Challenge is engaging elementary, middle school, and high school students in how real world problems are modeled mathematically, translated into computer code, and analyzed with a larger goal of attracting students to continue their education and pursue STEM fields. The Student Cluster Competition is engaging undergraduate students directly in an intense 48-hour challenge to broaden their HPC exposure, while it simultaneously seeks to affect undergraduate curricula. Further, size of student engagement also varies. The CSCNSI is a relatively small and non-competitive (in terms of competition between student teams) program that seeks to leverage (not necessarily affect) typical undergraduate computer science and computer engineering programs with the goal of systematically growing a targeted HPC skillset to provide a strong student intern and early-career staff pipeline. All three provide useful and important approaches to bridging various aspects and levels of the HPC skillset gap.

Far more important than the differences are the key similarities and shared attributes, including longevity (i.e., a decade or more of demonstrated success), an emphasis on hands-on, practical skill development, a focus on instilling a sense of wonder and accomplishment that comes from engaging with real-world problems, and a foundational reliance upon teamwork, both in terms of program planning and execution, and perhaps most importantly, directly in educational delivery structure (i.e., student teams). The significance of these attributes will become more apparent in the next section, which describes the CSCNSI in more detail.

III. COMPUTER SYSTEM, CLUSTER, AND NETWORKING SUMMER INSTITUTE

A. Motivation

The growing importance and pervasiveness of HPC for solving increasingly complex problems coupled with the fast-paced evolution of the HPC environment and the technical complexity of the challenges in the on-going march to the next generation of supercomputers has widened the entry-level skill gap for HPC organizations. LANL found itself largely unable to find student interns and early-career applicants with sufficient background and practical skills to fill its workforce needs within its HPC environment. In response to this gap, LANL created the CSCNSI to seed a strong entry-level applicant pool from which it would be able to staff and then further develop its HPC workforce as those staff progress through their individual career paths.

B. Institute Goals

To accomplish the overarching pipeline and workforce development objectives, three developmental focus areas were identified and a methodological approach was developed in which students work in small project teams to execute real-world projects on computer clusters that they assemble and configure to achieve the goal triad listed below.

- Practical skill development. The CSCNSI is designed to provide students with hands-on experience in setting up, configuring, administering, testing, monitoring, and scheduling of computer systems, supercomputer clusters, and computer networks.
- Technical broadening. The CSCNSI exposes students to a variety of key HPC-related topics, technologies, and problems of interest through execution of a research project, as well as participation in tutorials, seminars, and facility tours, including a full facilities and machine room tour of the Metropolis Supercomputing Center at LANL.
- Professional development. The CSCNSI provides students with a targeted set of professional development opportunities, including teaming and written and oral communication skill development through a series of workshops, training sessions and assignments, including resume writing, and technical project execution, poster development, and presentation development and coaching.

A second set of workforce development goals is achieved in the form of staff revitalization through mentoring, lecturing, and research collaboration opportunities. Even here, there is a

teamwork emphasis that brings staff members together to form a set of subject matter expert teams to mentor the student projects. These mentor teams are frequently cross-organizational, traversing team and group boundaries. Thus, a realized benefit has been the building of bridges and creation of collaboration opportunities that persist beyond the CSCNSI.

C. Program Overview

The primary objective of the CSCNSI is to provide a thorough introduction to the techniques and practices of cluster computing with a specific focus on system administration for large computing clusters. This objective is accomplished by leveraging and building upon the background and curriculum currently provided in a typical undergraduate computer science or computer engineering program, which has historically been more theoretically than practically based. Application to the CSCNSI is a competitive process in which student applications are solicited nationally. Although there is no hard restriction, the program typically targets students at Junior-level and above to provide a pool of applicants with sufficient technical background to maximize their ability to absorb the material and grow from the experience. Between nine and twelve students are selected each year. Students generally have had no cluster computing exposure prior to beginning the program. Many express that HPC education is either not commonplace or not available at their university (especially as part of Computer Science or Computer Engineering programs). The vast majority has never seen, much less been on the floor of a computer machine room.

The program includes lecture, laboratory, and professional development components. Students explore current challenges in HPC through the inclusion of an extensive seminar series consisting of seminars given by practitioners and researchers actively engaged in HPC-related topical areas. At its core, the CSCNSI is an intensive, project-driven, technical summer program during which each student team actively builds a computer cluster and then executes an HPC-related research project under the advisement of an assigned mentor or mentor team. The CSCNSI is executed as a paid ten-week undergraduate research internship.

D. Program Execution

The CSCNSI is structured as a set of complementary components that together realize the goal triad described in Section 3.2. After on-boarding and orientation activities, students are assigned to small teams that spend approximately two focused weeks building a small, but fully functional mini-supercomputer (one computer cluster per team). The second phase of the program consists of attending a set of HPC seminars, executing research projects on the clusters the students built, and documenting and reporting research project results in the form of a technical poster and presentation.

1) *Cluster Boot Camp:* The first phase of the program is Cluster Boot Camp in which students work in teams of three under the tutelage of an instructor and teaching assistant (both typically CSCNSI graduates) to build small supercomputers (one cluster per team), including everything from installing the hardware into the computer rack and labeling and connecting the cables and computer nodes to installing and

running an operating system and software stack similar to what national supercomputing sites use to operate their very large computing clusters. Each team is provided ten to twelve computer nodes, a computer rack, all required cabling, equipment, and machine room space to build a fully operational computer cluster over approximately a two-week period. Cluster Boot Camp lecture/laboratory topics include the following.

- Setting up and wiring a cluster,
- Linux review/ Operating system configuration,
- Linux services/ Head node configuration,
- Network booting,
- General networking,
- Warewulf cluster management/ Diskless booting,
- Shell scripting,
- Monitoring and benchmarking/ Ganglia setup,
- Parallel programming/ MPI setup,
- Lightweight Directory Access Protocol (LDAP), Kerberos, and
- Configuration management (e.g., Puppet).

However, it is important to note that Cluster Boot Camp as described above and currently implemented was not the original design, perhaps because it is not an immediately obvious starting point. The process and foundational concepts and choices that brought us to the boot camp model is described in more detail in a separate publication [16].

2) *HPC Seminar Series*: The intent of the HPC seminar series is to provide the students with a broad exposure to topics across the spectrum of HPC. Practitioners actively engaged in building, supporting, designing, and using various aspects of HPC systems and resources at Los Alamos provide the HPC seminars. The seminar set has some variation from year to year to capture timely topics in addition to including a core set of foundational topics. As an example, the HPC Seminar Series topics were as follows for the 2016 session.

- HPC Technology Overview
- The What and Why of HPC Division
- Mission Science at Los Alamos National Laboratory
- LANL Cluster Overview
- Resiliency and Reliability
- HPC Facilities Overview
- Cluster Configuration Management
- Glance and SLURM: User-Defined Image Management on HPC Clusters
- HPC Monitoring
- HPC File Systems
- HPC User Support and Consulting
- Design and Implementation of a Scalable Monitoring System for Trinity
- HPC Networking
- Neural Networks for Cross-Compiler Malware Detection
- Containers, WC, and Charliecloud
- UNIX Host Security
- Unclassified Weapons Program Overview
- Relational Statistical Study of Social Networks
- Storage (Campaign + HPSS)
- Design and implementation of a Scalable HPC Monitoring System

- DRM (Job Scheduling and Resource Management)
- Programming and Runtime Environment
- HPC Storage Systems: Serving Data to the Lunatic Fringe
- HPC Security
- LANL HPC History Project
- MarFS: A scalable POSIX File System/Data Lake on Cloud Objects
- HPC Student Mini-showcase

3) *Team Research Projects*: As previously noted, the team research projects are a core component of the CSCNSI. With Cluster Boot Camp complete, each team works with an assigned mentor or mentor team consisting of HPC subject matter experts to execute a real-world research project on their newly assembled and configured clusters. Prior to student arrival, the mentor teams have designed and proposed research projects. Sometimes, project ideas for the following year spring immediately from results obtained in the current year.

The team projects are based upon highly relevant research questions with direct applicability in an HPC production environment. Projects are often very timely and a result of needed efforts that may have been put on a backburner due to more urgent production issues or lack of staff time to complete. Projects are rarely, if ever, repeated – unless there are unresolved issues or topics that need deeper investigation, providing the follow-on project has sufficient substance to keep students busy for the duration of the summer. Most are true research projects by their nature and this is clearly communicated to the students as they start each summer. Some students choose to continue work on their projects once they return to their school or as an internship working with their mentors to explore the subject at a deeper level.

It is not uncommon for CSCNSI projects to have immediate and/or broader impact as will be noted in Section 4 below. Example research projects over the past five years have included the following. A complete list of projects with links to project posters is available [27].

- Infiniband Performance Characterization through Machine Learning (2016),
- Comparison of High Performance Network Options: EDR InfiniBand versus 100Gb RDMA Capable Ethernet (2016),
- Developing an HPC Monitoring Pipeline with Rabbit MQ (2016),
- Performance Studies of Parallel Erasure Coding on Clustered Micro Storage Servers (2016),
- Improving and Tracing Lustre Metadata (2015),
- Ceph: An Open Source Object Store (2015),
- Docker File System Isolation (2015),
- Docker: Testing the Waters (2015),
- Scalability of InfiniBand-Connected LNET Routers (2014),
- The Effects of SSD Caching on the IO Performance of Unified Storage Systems (2014),
- Backups Using Storage Clusters (2014),
- Monitoring I/O on Data-Intensive Clusters (2014),
- Cloud Management with Open Stack (2013),
- A comparison of Library Tracking Methods in High Performance Computing (2013),

- Functional Assessment of Erasure Coded Storage Archive (2013),
- Network Service Security Through Software Defined Networking (2013),
- Building a Parallel Cloud Storage System Using OpenStack's Swift Object Store (2012),
- Scalable Node Monitoring (2012),
- iSSH vs. Auditd: Intrusion Detection in High Performance Computing (2012),
- Gluster FS: One Storage Server to Rule Them All (2012).

4) Student Research Mini-Showcase: The CSCNSI culminates in a student research mini-showcase. The mini-showcase is a technical symposium and poster session for student interns at all levels (e.g., undergraduate, post baccalaureate, graduate, post-doctoral) engaged in computing work and research at LANL. The mini-showcase provides a unique opportunity for the students to present their research and results, engage in professional networking with other students and staff, broaden their technical exposure and expertise, and celebrate technical achievement.

The CSCNSI student deliverables are a technical poster for the poster session and a 20-minute technical presentation on their research for students and staff. Many students at the undergraduate or early graduate level have not yet engaged in a research project, prepared a technical poster, or prepared a technical talk. In preparation for this event, and as part of CSCNSI's professional development goal, the students receive training sessions and coaching on how to prepare a poster, how to put together technical presentation material, as well as coaching in presentation delivery.

IV. CSCNSI IMPACT

2016 marked the completion of the 10th Annual CSCNSI. Over the past decade, the core program has hosted 118 individual students participants from 37 distinct universities nationwide, including five New Mexico universities. Program graduates leave with a unique set of skills that are highly sought after at LANL and other locations that operate HPC facilities and/or engage in a wide variety of computing activities. We can point to multiple indicators of success in achieving the stated goals and objectives.

Strong student intern pipeline: Over the past decade, fifty to sixty percent (50%-60%) of CSCNSI participants have returned to LANL for follow-on internships and post-baccalaureate appointments. Other graduates have returned to intern at NMC, work for other government agencies (e.g., Sandia National Lab (SNL), Lawrence Livermore National Lab (LLNL), NASA Langley, US Army Research, Development and Engineering Command (RDECOM), and NCSA), have entered graduate school, or have moved on to computing positions in private industry.

Strong early-career pipeline: Approximately forty percent (40%) of those students who have returned to LANL for follow-on internships have been converted to permanent staff positions (i.e., approximately 20% of total program participants have been converted to permanent staff).

Broad institutional impact: Over the years, CSCNSI graduates have gone on to contribute at LANL in at least 25 different groups spanning 15 unique technical divisions.

Broader impact: In addition to the student research mini-showcase, student teams are encouraged to submit their research posters to external technical venues. Over the years, CSCNSI teams have had their posters accepted at annual computing conferences [22][23][24][25]. A CSCNSI 2011 student team submitted their results to the Graph 500 Committee at the 2011 Supercomputing Conference (SC11), which officially certified the team's test results, placing them at #27 and #33 on the 2011 Graph 500 list [17]. In addition, team research projects have attracted industry interest (e.g., the work of a 2010 CSCNSI team motivated a Mellanox white paper [18] and the work of a 2016 team has catalyzed substantive technical compatibility follow-on discussions between two networking switch vendors).

Staff engagement (project mentors): Mentors find the student engagement and technical exchange to be revitalizing and beneficial. Over the years, dozens of LANL staff members have engaged as CSCNSI project mentors, with more than 20 returning to mentor multiple years. Staff are motivated to participate because they have the opportunity to influence, design, and mentor timely research projects that are directly relevant to their professional focus and tasking, often choosing questions that they want answers to, but have not had time to dig into themselves.

Staff engagement (guest lectures): In addition, each year the CSCNSI hosts on the order of 25 predominantly LANL staff guest lecturers to execute the HPC Seminar Series and associated lectures who share their expertise on a variety of computing, science, and professional development topics. The intent of the HPC Seminar Series is to provide the students with a broad exposure to topics across the spectrum of HPC. The presenters are practitioners actively engaged in building, supporting, designing, and using various aspects of HPC systems and resources at LANL. The seminar set has some variation from year to year to capture timely topics in addition to including a core set of foundational topics.

Positive Exit Survey Results: Exit survey results consistently indicate a positive and highly productive participant experience. Graduates self-report a significant increase in breadth, depth, and technical skill base.

CSCNSI graduates return: Over the years, a significant number (10-15%) of CSCNSI graduates have returned in subsequent years to serve in a variety of roles, including CSCNSI instructor, teaching assistant, guest lecturer and/or technical project mentor. We interpret this as an indicator of a healthy, relevant, and engaging program.

Program duplication: In the summer of 2016, LLNL piloted a program closely patterned on the CSCNSI at its site. Argonne National Laboratory has also expressed interest in beginning a similar program. Both national laboratories have expressed that they are experiencing similar and significant HPC

pipeline challenges and are looking for methods to address them (Meetings and conversations 2015, 2016).

University Collaboration and Outreach: In 2016, LANL and NMC jointly explored various modes of university outreach and collaboration. We hosted a graduate student team consisting of CS and Physics Ph.D. students from New Mexico State University to engage in Mini-Boot Camp training (as described below in Section 6). The outreach experiment was quite successful, the students reported that they learned a lot and their advisors intend to use their newly gained skills to help administer research clusters in the CS and Physics departments. We also worked with the New Mexico Institute of Technology to develop a mechanism for CSCNSI students to receive college credit for completion of the program. The successful course pilot was offered as a third-year, three credit hour, CS course during the 2016 summer session. Half of the CSCNSI students enrolled, successfully completed the course, and received college credit for their work in the 2016 CSCNSI.

V. KEYS TO SUCCESS

The CSCNSI is a carefully designed program where every aspect and component has been created and honed to achieve the overall objectives of bridging a widely recognized and organizationally felt HPC skillset gap and of growing a well-prepared and excited HPC entry-level workforce. The institute's long-standing and consistent success is due to its adherence to a core set of fundamental philosophies. These include a commitment to continuous improvement and a commitment to an essential set of teaching philosophies and methodologies that result in targeted, hands-on practical skill development, technical broadening, and professional development, which in turn uniquely prepare students to enter the HPC workforce.

From application process through execution to evaluation, the program is carefully organized to maximally engage busy, high-performing HPC subject matter experts and to do so with minimal burden and with an eye to providing value in return for their mentorship.

- The program removes the administrative burden of mentoring students by design. The CSCNSI framework absorbs the administrative tasks typically associated with student mentoring, including on-boarding, office space, required training, paperwork, scheduling, lecture room reservation, machine room equipment, machine room maintenance, desktop computers, poster printing, etc., removing the burden from the staff and thus allowing them to fully direct their efforts toward technical mentoring of the research projects.
- The program provides a full time instructor who leads the base instruction and is always available to the students during project time to keep projects moving and on track.
- The CSCNSI works closely with mentor teams to help them design, plan, and execute highly relevant and timely research projects that are directly related to or an extension

of their professional interests and duties, making mentoring a win/win proposition for the mentors and their home organizations. In turn, the students are interacting with a wide range of HPC technical leaders and actively engaged expert practitioners (through their projects and through their attendance and interaction at the HPC seminars), working on cutting edge equipment and engaging in relevant and timely research questions.

- Staff receive an extended interview with students who are attaining the exact HPC skill base and technical exposure that they are looking for, meaning that their recruiting burden is lessened and organizations receive well-prepared follow-on interns (CSCNSI graduates) that are ready to hit the ground running.
- Mentor teams consist of two to three staff members, thus lessening the individual mentoring burden and time commitment and allowing the mentoring interactions to fit more easily into busy work schedules, while ensuring that mentors are highly available to students and teams remain highly successful.
- The Cluster Boot Camp is designed to be comprehensive and challenging while ensuring that all students have the resources and the support to be successful.

Another key to success is the CSCNSI's focus on teamwork, both in the mentor teams and as a student engagement and teaching methodology. One of the most important components of the CSCNSI, both as predicted, and as reported by the participating students in surveys during and after program completion, is the opportunity to work in a team, as well as the close interaction of students across teams. The planning and placement of the teams start months before the program commences. The CSCNSI is advertised nationally through job fairs, forums, conferences, recruiting trips, and by, most appreciated, word of mouth (we have many applicants from colleges of previous program graduates). Applications are accepted during approximately a six-week time-window. A committee of technical staff from LANL and the NMC reviews the student resumes, and the top students are selected and invited to the CSCNSI. A typical summer invites nine to twelve students for attendance. Once offers are accepted, student teams are formed, considering both student skill sets, project needs, and typically an attempt to place unique schools (if possible) in each team. This increases the diversity of the teams and makes sure that teams are as well prepared as possible for the research project portion of the program. A good example for placement consideration is a project that requires heavy coding; students with a stronger background or preference for coding would likely be directed to that team.

The cluster building exercise, our signature Cluster Boot Camp, proves to be much more than just a crash course in Linux and systems administration. It is also an incredibly useful tool in team building, as each team is taken through a myriad of obstacles, and a very intense start of the summer. This helps to set the team up for productive and positive interactions during the research phase. In the early CSCNSI offerings, each team was introduced to their project and

mentors before Cluster Boot Camp, but they were not allowed to do project work prior to the completion of Cluster Boot Camp. This allowed mentors to be introduced to students earlier, and optionally to be involved in Cluster Boot Camp as extra support to the instructor team during exercises. Beginning with the 2015 and 2016 summers, those introductions have been held until after the successful completion of Cluster Boot Camp to allow students to focus on cluster building without distraction. In both scenarios the students have striven to build clusters correctly in order to successfully utilize them for their known/unknown project. They are motivated to work to build as stable and reliable cluster as possible to facilitate their project work and avoid lost project time due to cluster rebuilds and fixes. As Cluster Boot Camp concludes, none of the teams want to be the one still fixing their systems, but rather, they want to be launching their research projects.

In the past few years, we have introduced non-technical, team building courses and training to further encourage team building and an understanding of different working styles. In particular the 2016 exercise evaluated each student's working style (through survey) and interactively set out to develop ways to interact, given the differences between team members, and to better understand and value and benefit from similarities and differences and varied problem solving approaches.

The CSCNSI instructor teaching philosophy is essentially "baptism by fire." There is a large amount of information to digest in a short amount of time, a need to divide and conquer, and a great need to share knowledge between team members. Questions are not answered until students have at least tried to research the problem first. The idea is to give students hints at how to do things but not show them the solution. There are often many ways to get things working correctly. The students become familiar with a situation in which things do not immediately work. They experience both success and failure in a setting that is a microcosm of a real world production-computing environment.

VI. PROGRAM EVOLUTION

The CSCNSI was conceived to be a continuously improving, living program. Technology is ever changing. Every year the CSCNSI evolves to stay on the cutting edge, and 2017 will be no exception. Approximately every four to five years, the CSCNSI affects a complete hardware refresh, which will happen in 2017 when the hardware purchased in 2013 will be replaced.

Likewise, while the overall topics covered in the fast-paced boot camp and guest lectures are essentially the same now as they were at inception in 2007, it is important to point out that the content is in no way, shape, or form static. As systems software (like operating systems) evolve, so must the course material. Most of the projects that students execute, deal with advanced topics at the bleeding edge of computation, and it is never an option to use older software. Therefore, before each summer, the lecture slides are updated to reflect the latest software revisions available in standard OS distributions, as well as updated communications technology. This is one of the major strengths of the program. The CSCNSI does not have outdated textbooks, but a living, hands-on curriculum.

In 2015, an addition to the core CSCNSI program was proposed. A Cluster Mini-Boot Camp was planned and executed as a skill-broadening intensive for new or early career HPC staff without previous computer cluster experience or background. The training format is a condensed, six-day version of the regular CSCNSI Cluster Boot Camp in which participants work under the guidance of an instructor through a series of lecture, demonstration, and hands-on implementation cycles to build and configure a working computer cluster. The 2015 pilot was well executed and well received, and two additional Mini-Boot Camps were executed in 2016, one immediately preceding and one immediately following the regular ten-week CSCNSI. The second included LANL Staff, as well as a Ph.D. graduate student team from New Mexico State University, demonstrating the broad applicability of the program. As a result of these successes, increasing technical sophistication in the student projects, and a desire to affect the pipeline more broadly, the CSCNSI is likely to be including more graduate student participants in the future.

Finally, although the CSCNSI female engagement rate (at 17%) is on par with (or slightly higher than) national statistics for female enrollment in computer science and engineering [19], diversity is an area in which we constantly seek to improve through local and national recruiting, as well as through targeted engagement with conferences such as the Grace Hopper Celebration of Women in Computing conference [20] and the ACM Richard Tapia Diversity in Computing conference [21]. In addition, this year we will be more actively engaging with diversity affinity email lists to achieve broader awareness of the CSCNSI student opportunity.

VII. CONCLUSIONS

The HPC skillsets required at national laboratories are many and must address multi-faceted needs, including but not limited to computer architecture, operating systems, algorithms, application development, data analytics, machine learning, computational modeling, system architecture, networking, storage, monitoring, system monitoring, programming languages, workflow, system administration, facilities and operations. It is unlikely that any single program or initiative or university curriculum will be able to fill these needs in isolation or that students will gain these skills on their own. Hence the necessity for continued effort, collaboration, programs, and innovative initiatives continues to grow.

Our metrics indicate that the CSCNSI has been successful in terms of its objectives to create a strong student intern and early-career pipeline that provides both HPC-specific technical skills and professional skills (e.g., oral and written communication and teaming skills) well suited for the HPC environment. Both the organization and the student participants benefit from an extended interview to examine opportunity and fit. Students are able to network to find opportunities of specific professional interests. Over the past decade, CSCNSI graduates have returned for follow-on internships at a high rate, integrated well into their chosen opportunities, and have been converted to career staff at a high rate.

This paper described the challenges and motivations that led LANL to design a program that existed nowhere else, remains unique, and is being patterned. The CSCNSI has been

successful far beyond expectations and continues to create opportunity and generate success for its students, its graduates, the staff mentors, and LANL's HPC Division. Students comment and say, "I've learned more in these nine weeks of summer than I did in my first two years of college." CSCNSI has been, in terms of recruiting and retention, one of the most successful student programs at LANL, an organization that prides itself on its student programs and the large fraction of current staff members overall who started as LANL students. CSCNSI is proud to be part of that legacy. We see no signs of slowing as the need for qualified computer systems professionals continues to increase as high-end computing becomes more ubiquitous within industry and government.

ACKNOWLEDGMENT

The authors would like to thank all of the LANL staff mentors who are the engine behind the continued success of the CSCNSI, past instructors, including Andree Jacobson, Dane Gardner, Matthew Broomfield, and Jarrett Crews, the CSCNSI Teaching Assistants, as well as all of the students who have participated in the program and provided valuable feedback to help us evolve and improve the program. The co-authors are grateful to Gary Grider who articulated the critical entry-level HPC pipeline need and then reached out to partner with other creative and adventurous souls to co-found this Summer Institute to address it.

REFERENCES

- [1] Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020, National Academy of Sciences Final Report, National Academies Press. Washington, D.C., 2016. Doi: 10.17226/21866. Downloaded PDF, <http://www.nap.edu/catalog/21886/future-directions-for-nsf-advanced-computing-infrastructure-to-support-us-science-and-engineering-in-2017-2020>
- [2] Top Ten Exascale Research Challenges, DOE ASCAC Subcommittee Report. February 10, 2014. Downloaded PDF <http://science.energy.gov/~media/ascr/ascac/pdf/meetings/20140210/Top10reportFEB14.pdf>
- [3] "The Tip of the Iceberg: Say, what is "under the floor" of a supercomputer?" National Security Science, April 2013. Available online: <http://www.lanl.gov/discover/publications/national-security-science/2013-april/assets/docs/under-supercomputer.pdf>
- [4] N. Liu, J. Cope, P. Carns, C. Carothers, R. Ross, G. Grider, A. Crume, and C. Maltzahn, "On the role of burst buffers in leadership-class storage systems," IEEE 28th Symposium on Mass Storage Systems and Technologies, April 2012, pp.1-11.
- [5] J. H. Laros III, K. Pedretti, R. E. Grant, S. Olivier, M. Levenhagen, D. DeBonis, S. Pakin, S. Martin M. Kappel, P. Falde, "ACES and Cray collaborate on advanced power management for Trinity," 2016 Proceedings of the Cray User Group, London, May 2016. https://cug.org/proceedings/cug2016_proceedings/includes/files/pap107.pdf
- [6] National Nuclear Security Administration Ensures Safety and Security of U.S. Nuclear Stockpile with Cray XC40 "Trinity" Supercomputing Solution. Trinity Cray Press Release, 2015. <http://science.energy.gov/~media/ascr/ascac/pdf/meetings/20140210/Top10reportFEB14.pdf>
- [7] America's Next Generation Supercomputer: The Exascale Challenge: Hearing before the Subcommittee on Energy, Committee on Science, Space, and Technology," House of Representatives, One Hundred Thirteenth Congress, First Session, Wednesday, May 22, 2013. Online, Accessed August 22, 2016. <http://purl.fdlp.gov/GPO/gpo40360>
- [8] S. Graham, M. Smir, and C. Patterson, "Getting up to speed: the future of supercomputing," Technical report, National Research Council of the National Academies, 2005.
- [9] "Exascale timeline pushed to 2023: what's missing in super computing? Web page editorial. The Next Platform, April 27, 2016. Accessed August 22, 2016. <http://www.nextplatform.com/2016/04/27/exascale-timeline-pushed-2023-whats-missing-supercomputing/>
- [10] New Mexico Consortium. Online: Accessed August 25, 2016. <http://newmexicoconsortium.org/>
- [11] PROBE: A Thousand-Node Experimental Cluster for Computer Systems Research. Garth Gibson, Gary Grider, Andree Jacobson, and Wyatt Lloyd. USENIX; login;, Vol 38, No 3, June 2013. <https://www.usenix.org/publications/login/june-2013-volume-38-number-3/probe-thousand-node-experimental-cluster-computer>
- [12] Cohen, M., Foster, M., Kratzer, D., Malone, P., and Solem, M., A High School Supercomputing Challenge. *Proceedings of the 1991 ACM/IEEE conference on Supercomputing*, Pages 68-75.
- [13] Foster, M. and Kratzer, D. New Mexico High School Supercomputing Challenge, 1990-1995: Five Years of Making a Difference to Students, Teachers, Schools, and Communities. Technical Report. Los Alamos National Laboratory. LA-13091-PR. 1996.
- [14] Supercomputing Challenge: Taking the Challenge since 1990. <http://www.supercomputingchallenge.org/>
- [15] Harrell, S. L., Nam, H. A., Vergara, V. G., Keville, K, and Kamalic, D., Student Cluster Competition: A Multi-disciplinary Undergraduate HPC Educational Tool, EDUHPC2015, Nov. 2015, Austin TX.
- [16] C. Connor, A. Jacobson, A. Bonnie, G. Grider, "An innovative approach to bridge a skill gap and grow a workforce pipeline: The Computer System, Cluster, and Networking Summer Institute," USENIX Journal of Education in System Administration (JESA). LA-UR-16-26535, in press.
- [17] The Graph 500 List. Official 2001 Results. http://www.graph500.org/results_nov_2011
- [18] Shainer, G., Los Alamos National Labs and Student Team Validate Performance Benefits of RDMA over Converged Ethernet (RoCE). Feb 21, 2011, available online. <http://www.techonline.com/electrical-engineers/education-training/tech-papers/4228835/los-alamos-national-labs-and-student-team-validate-performance-benefits-of-rdma>
- [19] National Center for Women & Information Technology Fact Sheet. <https://www.ncwit.org/ncwit-fact-sheet>
- [20] Grace Hopper Celebration of Women in Computing. <http://gracehopper.anitaborg.org>
- [21] ACM Richard Tapia Celebration of Diversity in Computing. <http://tapiaconference.org>
- [22] Stitt, T., Bonnie, A., Fuerst, Z. 2015. *Virtualizing File Transfer Agents for Increased Throughput on Single Host*, SC'15 http://sc15.supercomputing.org/sites/all/themes/SC15images/tech_poster/tech_poster_pages/post192.html Poster; Supercomputing Conference,
- [23] Wiens, C. L., Long, J. M. C., Ornstein, J.R. 2014. I/O Monitoring in a Hadoop Cluster; Poster; Supercomputing Conference, SC'14. http://sc14.supercomputing.org/sites/all/themes/sc14/files/archive/tech_poster/tech_poster_pages/post144.html
- [24] Erickson, K., Kachelmeier, L., Van Wig, F.. 2016. *Comparison of High Performance Network Options: EDR InfiniBand versus 100Gb RDMA Capable Ethernet*; Poster; accepted Supercomputing Conference, SC'16.
- [25] Easterday, H., Ogas, J., Smith, A., 2016. *InfiniBand Performance Characterization Through Machine Learning*; Poster; accepted to the Supercomputing Conference, SC'16.
- [26] Lax, P. *Report of the Panel on Large Scale Computing in Science and Engineering*. Sponsored by the U.S. Dept. of Defense and the National Science Foundation, in cooperation with the Dept. of Energy and the National Aeronautics and Space Administration, 1982.
- [27] CSCNSI Project page. Online: Accessed October 6, 2016. <http://www.lanl.gov/projects/national-security-education-center/information-science-technology/summer-schools/cscnsi/student-projects.php>