# HPC Workforce Preparation

Scott Lathrop
Shodor Education Foundation, Inc.
NCSA/University of Illinois
1205 W. Clark St.
Urbana, IL 61801 USA
lathrop@illinois.edu

***Abstract*—Achieving the full potential of today's HPC systems, with all of their advanced technology components, requires well-educated and knowledgeable computational scientists and engineers.**

***Keywords—workforce, education, training, internships, fellowships.***

## I. Introduction

It is well documented in numerous national reports that Computational Science and Data-Enabled science and Engineering (CSDE) are integral to science, technology, engineering, mathematics, and scholarly studies within academia, government agencies, and business and industry [3, 4, 5, 6]. CDSE is dependent upon a pervasive, extensible, sustainable cyberinfrastructure that includes high performance computing, high throughput computing, data-intensive computing, scientific visualization, and networking ecosystem that this document will reference as the high performance computing (HPC).

Economic competitiveness requires a significantly larger and more diverse computationally literate workforce. The workforce includes educators and researchers spanning all disciplines; software and hardware developers of algorithms, applications, community codes, and tools; as well as infrastructure staff able to install, operate and maintain the rapidly evolving complex integrated computational ecosystem.

National-scale strategies are critically needed to prepare the extensive human capacity required to advance scientific discovery. The strategies will require affordable, sustainable, systemic changes across a range of disciplines, educational levels, institutions, and sectors of society.

Demographic and economic needs make it critical that every effort is made to recruit, prepare and sustain significantly larger numbers of women, minorities and people with disabilities among the CDSE workforce.

"*No first-world nation can maintain the health of its economy or society when such a large part of its population remains outside all scientific and technological endeavors.*" Richard Tapia, Rice University

The national STEM workforce is inadequate to meet today's needs, especially among women and minorities. The lack of graduates in science, technology, engineering and mathematics (STEM) is undermining the potential of research to capitalize on the computational ecosystem. The nation is failing to prepare an adequate workforce for a number of reasons including, but not limited to:

- CDSE and HPC information learning opportunities are limited, scattered, incomplete, and of varied quality
- Students are not provided adequate opportunities and motivation to gain in-depth knowledge to pursue timely solutions to real-world problems

## II. Challenge

A diverse population, motivated and engaged in the full spectrum of computational activities, is critical to the advancement of scientific discovery and scholarly research and education for generations to come. Today's leading edge research endeavors require inter-disciplinary teams working across geographical distances. To prepare individuals for these research teams, today's educational system must keep pace with the rapid pace of technological change. This is in reality an international-scale challenge requiring a computationally literate workforce that reflects the diversity of each country's population.

## III. Strategies

The Blue Waters project is actively tackling these challenges through strategies that are based on research findings and being conducted in collaboration with other organizations actively pursuing these challenges. The purpose of this paper is to present ideas for addressing the HPC workforce challenge and to foster discourse on these and other strategies the community can and must pursue to prepare a larger and more diverse HPC workforce.

The Blue Waters project is actively pursuing these strategic activities:

- Training webinars, workshops and summer schools,
- Web-based graduate credit courses,
- Fellowships and internships,
- Education allocations,
- Dissemination of materials, and

- Evaluations for continuous improvement.

Blue Waters is working in collaboration with the XSEDE Training, Education, and Outreach Services (TEOS) team, the University of Michigan, the ACM SIGHPC Education Chapter, and numerous national and international organizations and professional societies. The sharing of lessons learned from these efforts helps to inform all programs, resources, and services to improve.

According to African folklore, "*it takes a village to raise a child*[1]". To prepare a large and diverse HPC workforce will require a national-scale inter-disciplinary effort. We heartily welcome others to join in the effort!

## IV. TRAINING

Today's HPC research community includes faculty, scientists, engineers, postdocs and graduate students from all fields of study. These individuals benefit from training offered by Blue Waters via live in-person events, through webinars to reach people at their desktop, through videoconferencing facilities to facilitate group learning, and through self-paced tutorials to support just-in-time learning.

Blue Waters provides webinars to address the on-going training requirements of the science, technology, engineering, mathematics, and scholarly studies communities. In these webinars participants learn about tools, technologies and resources, empowering them to make the most effective use of a diverse array of HPC platforms, including current petascale and emerging extreme scale systems. The topics for the webinars are responsive to the needs of the community. Presenters are drawn from leading organizations including Blue Waters, vendors, end-users, and the broader HPC community, thereby ensuring that the sessions provide the participants with current information and insights into emerging systems.

Hands-on workshops are provided to demonstrate the use of the software tools and libraries from established HPC vendors as well as those developed though through R&D efforts to enhance extreme scaling capabilities. Through these workshops, Blue Waters staff work with individuals on their applications to incorporate emerging libraries and tools into the research software development process.

The Virtual School of Computational Science and Engineering [2] (VSCSE) supported by the University of Michigan and Blue Waters offers multi-day and week-long summer school sessions for graduate students, post-doctoral students and professionals from academia, government and industry to gain the skills they need to leverage the power of cutting-edge computational resources. Often times, the practical aspects of computational science fall between the cracks as computer science departments focus on what computer scientists need to know and domain science and engineering departments focus on the applications of computer science to those disciplines. The VSCSE was created to help students fill those knowledge gaps, preparing them to use petascale and extreme scaling computing resources. Participating in the Virtual School also helps students build networks of fellow researchers who they can turn to for support and collaboration. The VSCSE courses are delivered simultaneously at multiple locations across the country using high-definition videoconferencing technology thereby affording participants the opportunity to directly interact with the instructors.

For people who need access to just-in-time training, there are numerous peer reviewed self-paced tutorials offered via XSEDE[3], CI-Tutor [4] and the Cornell Virtual Workshop[5]. The largest number of people being trained on CDSE and HPC are learning via on-line tutorials. XSEDE reported in September of 2014 that over 50,000 registrations for these on-line tutorials had been received during the prior three years. The number of topics and the number of registrations are continuing to rise as the number of courses is increased and the community becomes aware of these high quality peer-reviewed resources.

## V. COURSES

There are numerous efforts to provide access to on-line learning such as through Coursera[6], edX[7]. The Blue Waters project is building on the CourseShare[8] model of course sharing within the Committee on Institution Cooperation to offer on-line semester-long credit bearing graduate courses using an innovative format to allow participation by students at multiple institutions across the country.

Course lectures are delivered as asynchronous online videos. The courses are offered as a traditional graduate level college credit course, including a syllabus with learning outcomes, 40 hours of instruction, reading assignments, homework and exercises, and assessment of learning progress. The program provides opportunities for students to earn graduate credit at their own institution. The courses are taught by experts in the field. The experts record their lessons and provide testing and practice (e.g. homework, mini-projects) material. The recordings are played at the remote participating institutions according to the local semester schedules. Local faculty supervise the classes and provide support to the students.

During the fall of 2014, the course *Algorithmic Techniques for Scalable Many-Core Computing* was offered to multiple institutions. The findings of the evaluation indicated that the course provided students with access to course content not previously available on their campus, and that the students benefitted from quality hands-on experiences. Further, the

---

[1] Wikipedia (http://en.wikipedia.org/wiki/It_Takes_a_Village)

[2] www.vscse.org

[3] https://www.xsede.org/web/xup/online-training

[4] https://www.citutor.org/

[5] https://www.cac.cornell.edu/vw/

[6] https://www.coursera.org

[7] https://www.edx.org

[8] https://www.cic.net/projects/shared-courses/courseshare/introduction

findings indicate that the students gained knowledge/skills through this course that will contribute to their work/research, and that the students are interested in learning more about HPC as a result of this experience. The following improvements are being made to the next set of courses based on the lessons learned from offering this course:

- more complete quizzes that reflect final exam content,
- more examples of final projects to assist students and faculty with choosing projects,
- changes to underlying course management infrastructure to provide multiple ways to interact with participating students and faculty, and
- additional coordination with collaborating faculty

Blue Waters is currently supporting two additional courses: *Designing and Building Applications for Extreme Scale Systems*, and *High Performance Visualization for Large-Scale Scientific Data Analytics*. These courses are reaching students who would otherwise not be able to enroll in these courses on their own campuses. The Blue Waters project seeks faculty interested in teaching additional HPC related courses to enhance the education of students across the country.

## VI. Student Internships and Fellowships

Educational research demonstrates that student attitudes and commitments to pursuing science and engineering are directly related to research experiences, internships and fellowships provided during their educational experience. The Blue Waters project is working to recruit, prepare and motivate undergraduate students through internships and graduate students through fellowships. Particular attention is given to pro-actively recruiting under-represented students including women, minorities, and people with disabilities.

The Blue Waters Internship Program is designed to immerse undergraduate students within projects associated with the Blue Waters and/or XSEDE efforts. The students are provided with an introductory two-week institute to ensure they are familiar with the Blue Waters and XSEDE environments, including the relevant tools and resources they will need to be familiar with to contribute to their assigned project. The students are then immersed in a year-long research project using the Blue Waters and/or XSEDE computing systems. The students are provided with on-going seminars, training sessions, and informal discussion groups to further enhance their knowledge and skills as well as motivate them to continue their formal education and careers in science and engineering. Twenty students are selected annually, and over half of this year's students this year are women and/or minority students.

The Blue Waters Graduate Fellowship Program is modeled after the NSF Graduate Fellowships, using similar requirements for eligibility, level of support, application process, and reporting requirements. The fellowship support is for one year up to $50,000 per year encompassing a stipend of $38,000 plus $12K in support of tuition and fees, as well as support for travel to Blue Waters sponsored events. The program supports candidates who are engaged in a program of study and research that is directly relevant to the utilization of Blue Waters or XSEDE resources in support of their research. Preference is given to candidates engaged in a multidisciplinary research project that combines disciplines such as computer science, applied mathematics and computational science applications. Ten graduate students were selected to participate during this year, and half of them were women. The second cadre of students will be announced at the 2015 Cray User Group meeting, and they will participate in the annual Blue Waters Symposium in May of 2015.

The evaluations demonstrate that the interns and fellows are benefitting tremendously from these in-depth immersions into HPC. The interns rated the two-week summer institute at 95% rated the event "above average" or "excellent" compared to similar events. The students indicated that the networking and social aspects were strengths of the program, and that working groups helped build community and produced valuable outcomes to guide the participants.

The fellowship program has allowed the graduate students to find intersections between their science fields and HPC community, to achieve independence in their research area, to gain access to an abundant amount of resources, and to gain better directions for their research. The fellows are making connections with students and faculty in new fields, and they are highly interested in using the computational techniques they are learning in their research. A few quotes from the fellows follow:

- *"I've worked with HPC for a long time, but it's been nice to reconnect into the community and see all the expertise that's there and that goes in to it. And have contacts for possible employment opportunities looking to the future."*
- *"I'm forced to learn. And sometimes it sucks being knee deep in something that you don't know what you're doing, but a lot of times that forces you to learn and get better at what you're doing."*
- *"I've worked on several HPC machines, and Blue Waters staff is by far the most responsive and helpful."*
- *"And not only is this an opportunity to learn the material, but the people at Blue Waters have been really helpful in pointing us in the right direction about what material to even look at, which I don't think I would have gotten out of a different program."*

## VII. Education Allocations

Blue Waters provides up to 1% of the computing resources in support of education. Education allocations have been used by over 860 participants from 28 institutions in one-day workshops, semester courses, and multi-day summer schools, as well as by the Blue Waters graduate fellows and the Blue Waters student interns. The Blue Waters system has been uniquely able to support the OpenACC workshops offered by

XSEDE, as no current XSEDE allocated production system supports OpenACC. Requests for education allocations are open to the academic community.

Impact statements from the faculty receiving these allocations include:

- *"the resources used are significantly beyond what they have access to on campus"*
- *"this experience will better prepare the students to use petascale resources in the future"*
- *"team gained invaluable knowledge of Blue Waters and High Performance Computing"*

## VIII. Dissemination of Materials

Education and training materials developed by practitioners are collected, peer-reviewed, and shared with the community to foster the development of a broader, well-educated community able to conduct computational science and engineering research using petascale technologies, resources and methods. This includes 30 undergraduate course modules developed by faculty with Blue Waters funding, along with other resources collected from XSEDE, HPC Centers, GLCPC institutions, Campus Champions, and the broader community.

Information is shared via the HPC University portal (www.hpcuniversity.org) that has been developed with contributions from Blue Waters. Included among the collections are the education modules developed by Shodor with Blue Waters support, as well as the materials developed for the Virtual School of Computational Science and Engineering Summer Schools through support from Blue Waters. The HPC University portal is actively collecting CS&E and HPC education and training content, workshop and training event information, fellowship and internship opportunities for students, student competitions, and other information relevant to learning about HPC. There were over 20,000 downloads of learning materials from the portal during the past year.

The team conducts regular focus groups to gather community input to improve the repository. The findings from the most recent focus group of college students found that the students felt that they could trust all the information on this website, because the links and resources it provided are from credible organizations. The focus group participants said the website provided many different perspectives for each topic. Most of the participants said one of the strengths of this website is that it provides students with internship and fellowship program information, which the students highly valued. Some comments include:

- *"Most websites that I can't trust gives you one perspective, but this website since it gives you so many different links to go...."*
- *"I would recommend (to CS students) this website, because I noticed that it has internships and fellowships, things to do after graduation."*

## IX. Evaluations Drive Improvements

An external evaluator conducts formative and summative evaluations to improve the programs and activities based on formative feedback, while collecting appropriate data and information to conduct a longitudinal analysis of the impact of the programs over the life of the project. The evaluation includes surveys, interviews, and focus groups with members of the petascale community.

The evaluations are leading to improvements in the programs and are helping to document the impact of the programs on the community.

## X. Summary

This paper describes a number of programs focused on preparing the HPC workforce. The efforts are informed by research and guided by on-going evaluations, focus groups, and community requirements. We welcome participation and collaborations from among the community for enhancing our joint efforts.

## References


[1] NSF Task Force on Cyberlearning and Workforce Development Open Community Meeting Agenda, http://groups.google.com/group/nsf-clwd-task-force

[2] Cyberinfrastructure for 21st century science and engineering, http://www.nsf.gov/pubs/2010/nsf10015/nsf10015.jsp

[3] Cyberinfrastructure Vision for 21st century discovery: Chapter 5: Learning and workforce development (2006-2010), http://www.nsf.gov/pubs/2007/nsf0728/nsf0728_6.pdf

[4] Fostering learning in the networked world: The cyberlearning opportunity and challenge, NSF Task Force on Cyberlearning, 2008, http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf.