Introduction to Volume 10 Issue 2: Special Issue on HPC Training and Education

Nitin Sukhija
Slippery Rock University of Pennsylvania
Slippery Rock, PA

FORWARD

High performance computing is becoming central for empowering scientific progress in the most fundamental research in various science and engineering disciplines as well as broader societal domains. It is remarkable to observe that the recent rapid advancement in the today's and future computing and software environments provide both challenges and opportunities for cyberinfrastructure facilitators, trainers and educators to develop, deliver, support, and prepare a diverse community of students and professionals for careers that utilize high performance computing to advance discovery. This special issue focuses on original research papers submitted to the First Workshop on Strategies for Enhancing HPC Education and Training (SEHET18), which was held in conjunction with PEARC18 in Pittsburgh, Pennsylvania, U.S.A., July 25, 2018 and the Fifth SC Workshop on Best Practices for HPC Training and Education (BPHTE18), which was held in conjunction with SC18 conference in Dallas, Texas, U.S.A., November 12, 2018.

This special issue begins with an article by Ponce et al. that presents tools, techniques and a methodology for developing curriculum courses aimed at graduate students in emerging computational fields, including biology and medical science. The teaching methodology used focuses on computational data analysis and statistical analysis, while at the same time teaching students' best practices in coding and software development. They found significantly good evaluation results especially in the Institutional composite questions. Those include items such as an intellectually stimulating course, provides a deeper understanding of the subject matter, learning atmosphere, and overall quality of the learning experience

The article by Stieber describes the curricula of a 10-week advanced computational inorganic chemistry course based on the course-based undergraduate research experiences (CUREs). The students used the computational resources enabled by the National Science Foundation's Extreme Science and Engineering Discovery Environment (NSF XSEDE) to conduct independent research projects following in-class lectures and tutorials. They conclude by highlighting the impact of the program on the students and its relevance to workforce training.

The article by Coulter et al. presents the XSEDE National Integration Toolkit that provides the software used on most XSEDE systems in an effort to enable easy knowledge and best-practice transfer from XSEDE service providers to campus cyberinfrastructure (CI) professionals. They describe the Use Case based approach for the creation of the XNIT Toolkit to provide campus cyberinfrastructure administrators a way to include software that is commonly found on XSEDE resources, without having to do a complete re-installation of operating system that the XCBC cluster distribution would require. They conclude by discussing more flexible, extensible solutions for scientific software delivery for the XNIT toolkit to continue providing a useful solution to easing the pain of administrating a general HPC-style resource at the campus level.

The article by Fietkiewicz et al. presents a comparative study of student's experiences in a high performance computing course. They utilized a recursive matrix multiply as an exercise for designing a parallel program and to have students use an application that contrasts with a previous experience using the iterative, loop-based algorithm. They conclude by reporting on student outcomes with respect to their prior multithreaded programming experience and on the common errors that included recursively creating excessive threads, failing to parallelize all possible mathematical operations, and poor use of OpenMP directives.

The article by Ponce et al.. surveys the current academic and non-academic programs across the globe, and presents quantitative evidence demonstrating the need for programs in higher education in High-Performance Computing and Data Science. They focus on Canadian programs and specifically on the education program of the SciNet HPC Consortium, using its detailed enrollment and course statistics for the past six to seven years. They conclude by illustrating SciNet's path in developing and transgressing the usual role of training events for users, into full credited graduate courses recognized at the university level for masters and doctorates degrees.

The article by Destefano and Sung describes the external evaluation activities in the first three years of the Blue Waters (High Performance Computing) Community Engagement program for graduate fellows and undergraduate interns. The evaluation utilized the 'Educative, Value-Engaged Approach' and conducted formative and summative evaluations to improve the programs and activities based on continuous feedback, while collecting appropriate data and information to conduct a longitudinal analysis of the impact of the programs over the life of the project. They report that the evaluation plans, activities, findings significantly affected the fellowship and internship program implementation, and program impact.

The article by Chakravorty et al. describes an informal curricular model of short, intensive and applied micro-courses offered at Texas A&M University High Performance Research Computing (TAMU HPRC) that address generalizable competencies in computing as opposed to content expertise. The model used a number of interventions that have been systematically applied on a semester-by semester basis for greater visibility for the courses, engaging students with active learning methods and retaining student interest via the HPRC seminar series. They found number of factors that amount to 300% growth in participation in HPRC short courses and provide a longitudinal report that assess the success of strategies implemented by TAMU HPRC to promote cybertraining efforts across campus.

The article by Nunez et al. presents different learning methods and tools employed by Pawsey Supercomputing Centre to address specific educational and training purposes. They used various techniques ranging from traditional on-site training, through self-guided training materials to online training and webinars and an open repository of materials, covering different aspects of HPC systems usage, parallel programming techniques as well as cloud and data resources usage. They found through numerous feedbacks that there is no universal learning solution; instead, various solutions and platforms need to be carefully selected for different groups of interest.

The article by Zon et al. describes the changes in the training and educational efforts the Canadian academic high performance computing center in the areas of scientific computing and high performance computing over the last six years. They used enrollment data, attendance and certificate numbers to report on trends on the growth, demand and breadth of the Scinet's training program. They conclude by highlighting the results of the assessment which is a steady increase in the demand for the data science training and wider participation of 'non-traditional' computing disciplines.

The second article by Chakravorty et al. describes a conceptual framework to evaluate and explore different pedagogical approaches that utilize technologies to introduce learners to complex programming scenarios suitable for the intermediate level. They provide quantitative and qualitative evaluation of three distinct models of programming education: (i) connect coding to hands-on "maker" activities using Raspberry Pi's; (ii) incremental learning of computational thinking elements through guided exercises that use Jupyter notebooks; and (iii) problem-based learning with step-wise code fragments leading to algorithmic implementation. They conclude by reporting on the student assessment outcomes that involve using the Jupyter notebooks to accelerate student learning with coding concepts.

The article by Mullen et al. describes the applicability of Massively Open Online Courses (MOOCs) for scaling High Performance Computing (HPC) training and education. They used two MOOC case studies, including the design decisions, pedagogy and delivery to outline how MOOC courses differ from face-to-face training, video-capturing of live events, webinars, and other established teaching methods with respect to pedagogical design, development issues and deployment concerns. They conclude by reporting on best practices for segmenting content into smaller concept sized chunks for addressing HPC specific technical needs and concerns.

The article by Bisbal et al. presents training guidelines for building open-source software to address common problems encountered in the scientific community members when developing their own codes and building codes written by other computational scientists. The article outlines topics that are needed to be taught to computational scientists in a logical order to train them to build open-source software. The article concludes by providing references to some of the topics that could be used to develop training materials or distributed directly to students as part of the training materials for bridging a major skills gap for the computational scientists.

The article by Seo et al. describes the design of the Cyberinfrastructure Security Education for Professionals and Students (CiSE-ProS) system. They used engaging approaches to evaluate the impact of learning environments produced by augmented reality (AR) and virtual reality (VR) technologies for teaching cybersecurity concepts. They conclude by reporting on the successes and feedback of CiSE-ProS virtual reality (VR) program using the pilot study with high school students at the Summer Computing Academy at Texas A&M University.

The article by Kunkel discusses the preliminary design of the HPC Certification Program and an independent body that curate the competencies and issue certificates for the users. The article presents two purposes of the program: defining and organizing the fine-grained skills and the establishment of the certificates and online exams that confirms that the user possess those skills. They conclude by reporting on the HPC certification forum which plays a virtual central authority to curate and maintain the skill tree and certificates.

The second article by Fietkiewicz discusses the experiential differences in student performance and perceptions in the undergraduate level high performance computing (HPC) at Case Western Reserve University. They used six main HPC techniques, which includes: batch job processing, general optimization for sequential programming, parallel programming using spawned (forked) processes, parallel programming using OpenMP and multithreading, parallel programming using OpenACC and GPUs, and parallel programming using message passing and MPI. They found that academic experience was correlated to performance, and technical experience may have no correlation at all, assuming adequate coverage in class is provided.

The article by Chen et al. describes an online education and training platform designed by NVIDIA Deep Learning Institute (DLI) that helps students, developers, and engineers solve real-world problems in a wide range of domains using deep learning and accelerated computing. The new educational platform uses a combination of current online learning andragogy along with a cloud computing platform consisting of a VM and Docker containers. The online platform enables students to use the latest AI frameworks, SDKs, and GPU-accelerated technologies on fully-configured GPU servers in the cloud so that the focus is more on learning and less on environment setup. They conclude by reporting on the project-based assessment and certification offered to the students at the end of some courses and the feedback from the DLI University Ambassador Program offered to educators to teach free DLI courses to university students, faculty, and researchers.

The article by Ngo and Denton describes the approaches in leveraging Cloud-Lab, a publicly available computing resource, as a platform to develop and host materials to support teaching topics in cluster computing. They used two approaches in using CloudLab to teach advanced concepts in cluster computing: direct deployment of virtual machines (VMs) on bare-metal nodes and indirect deployment of VMs inside a CloudLab-based cloud. They found that the flexibility, availability, and scale of CloudLab bring significant applicability to other topics in computer science, including operating system, networking, and cyber-security.

The second article by Coulter discusses the elasticity and programmability of cloud resources to be used as a tool for educators who require access to a wide range of computing environments. The article presents the Jetstream cloud environment to provide training for both new HPC administrators and users, by showing a ground-up build of a simple HPC system that allows an educator to tackle everything from basic command-line concepts and scheduler use to advanced cluster-management concepts such as elasticity and management of scientific software.

The article by Marques et al. presents the best practices and experiences organizing the Best Practices for HPC Software Developers (HPC-BP) webinar series, an effort for the dissemination of software development methodologies, tools and experiences to improve developer productivity and software sustainability. They report on a model for a number of distinct roles and steps in the process of developing and delivering a specific webinar event. They conclude by reporting opportunities to further enrich the pool of topics related to software productivity and sustainability and further expand their outreach activities.