Title
Training Efforts in the Exascale Computing Project

Journal
Computing in Science & Engineering, 22(5)

ISSN
1521-9615

Authors
Marques, Osni
Barker, Ashley

Publication Date
2020-09-01

DOI
10.1109/mcse.2020.3010596

Peer reviewed

Training Efforts in ECP

Osni Marques
Lawrence Berkeley National Laboratory
email: oamarques@lbl.gov

Ashley Barker
Oak Ridge National Laboratory
email: ashley@ornl.gov

***Abstract*—This article delineates the training activities carried out under the auspices of the U.S. Department of Energy's (DOE) Exascale Computing Project (ECP). While some of these activities are specific to members of ECP, others can be beneficial to the community at large. We report on training opportunities and resources that the broad Computer Science and Engineering (CS&E) community can tap into; we seek to increase awareness about these resources, which we expect to go beyond ECP's scope and life cycle.**

I. Introduction

The mission of the Exascale Computing Project (ECP) of the U.S. Department of Energy (DOE) is to put together all the pieces in the puzzle for the realization of the nation's first exascale systems. These pieces involve mission critical applications and an integrated software stack, assembled in concert with the efforts of U.S. high-performance computing (HPC) hardware companies for the identification and development of advanced computer system engineering and hardware components. This amalgam is deemed essential for the deployment of fully functional, capable exascale computing environments.

ECP involves two DOE organizations, the Office of Science (DOE-SC) and the National Nuclear Security Administration (NNSA), and it is structured on the principles of co-design and integration. Its three key focus areas are: 1) Application Development (AD): exascale-capable applications are a basic element of ECP and will be the vehicle for delivering solutions and insights to crucial but up to now intractable challenges. 2) Software Technology (ST): software technologies play an essential supporting role in application efficiency, and span from low-level system software to high-level applications development tools and libraries. 3) Hardware and Integration (HI): this area ensures the integration of ECP applications, software and hardware innovations within DOE's computing facilities.

HI also includes a Training and Productivity (T&P) effort to help address the challenges facing the ECP teams by providing training on key exascale hardware and software technologies. In addition, one of the primary goals of T&P is to help the ECP's application development and software technology teams become better software developers. As researchers and scientists, team members don't necessarily have backgrounds in computer science or specialize in writing codes. However, in order for science projects to fully exploit the potential of exascale computing, the software must be optimized and ready for the launch of these new exascale systems. Consequently, the T&P activity focuses on critical technologies to improve the productivity of software developers, and to advance the sustainability of software products, with emphasis on the dissemination of best practices for software development, testing and deployment.

The T&P effort facilitates selected application-driven training on topics such as programming models, tools, libraries and frameworks, data management and workflows, data analysis and visualization tools, system software, on-node parallelism and vectorization, application portability techniques, and software engineering design. The training is conducted through a variety of activities such as seminars, webinars, deep-dive workshops, lectures, hackathons, and tutorials. Often, these activities serve as an introduction to topics that are pursued in depth by ECP teams that specialize on those topics.

In the next sections, we elaborate on the mechanisms we use for setting up the ECP training agenda, give examples of training sessions (held and planned), and provide some statistics about our target audience. We finish by giving pointers to available resources.

II. A Training Agenda

The agenda for the T&P effort is determined through feedback we receive from an annual survey of the ECP community, from an advisory council formed by representatives from DOE computing facilities, from interactions with the ECP community (e.g., at the ECP Annual Meeting, see Box 1), from participants in the Argonne Training Program on Extreme-Scale Computing (ATPESC) [1], and also the HPC community at large (e.g., at SC conferences).

The annual survey asks participants to rate topics in the following categories: programming languages; programming models and runtimes; compilers; package managers and software distribution tools; data analytics and visualization; data management and workflows; mathematical libraries, scientific libraries and frameworks; performance analysis tools; version control and Web-based hosting services; automated testing frameworks; collaboration tools; and "something else." These categories are defined in coordination with the AD, ST and HI areas, and also the advisory council. According to the last survey, the top most desirable topics for training (aggregate of "may have impact" plus "high impact" answers) are listed in Table I. The table includes information about events that have already been offered (o), planned (p), or remain "to be determined" (u), meaning that as it is the topic is either too broad or training opportunities may be available elsewhere. In the next annual survey, we will attempt to narrow down the

TABLE I: Top areas for training according to 2019 ECP Annual Survey.

topic	event
CMake	webinar (o), tutorial (o/p)
Spack	webinar (p), tutorial (o)
Building applications within containers	webinar (o/p), tutorial (o)
Git, GitHub and Gitlab	webinar (o), tutorial (o/p)
Python	webinar (o)
CUDA	hackathon (o), tutorial (o)
HIP	webinart (o)
SYCL	webinar (o)
Advanced MPI	classes (u)
C++ and Using C++14/17 effectively	webinar (o), classes (u)
OpenMP 4.5+	hackathon (o)
AMD GPU	hackathon (o)

needs for the topics considered to be too broad but that are nonetheless desirable.

Communication is an important aspect of the T&P activity. The training section (tab) of the ECP website [2] is the primary way that training opportunities are conveyed to the ECP community. This website provides a comprehensive list of upcoming and past training events (e.g., events related to the topics listed in Table I). The website also contains archives for all the materials from previous training events, including slides and video recordings of the presentations. These videos are also available on the YouTube channel *ExascaleComputingProject*. This approach allows us to share these materials outside of the ECP. Reaching beyond the researchers involved in ECP today helps to develop the HPC workforce, while also laying the groundwork for the second generation of exascale developers and users, after ECP ends and exascale systems become more common and accessible to a broader base of users.

The website also points to specific events hosted by the three computing facilities of DOE-SC's Advanced Scientific Computing Research (ASCR) program: Argonne Leadership Computing Facility (ALCF), National Energy Computational Research Center (NERSC), and the Oak Ridge Leadership Computing Facility (OLCF). These facilities also conduct training on topics of great importance to ECP. Thus, we wish to make sure the ECP community is aware of these training opportunities.

In order to coordinate training events between the ECP and ASCR's computing facilities, the Training Advisory Group (TAG) was formed. TAG meets each month to discuss training plans at each of the facilities and inside ECP; if there is interest and benefit, the members initiate jointly sponsored events, from hackathons to tutorials. This cross-lab collaboration through regular communication allows the ECP to better identify cross-training opportunities as well as potential scheduling conflicts.

The T&P effort uses a monthly training newsletter to communicate upcoming training events. Members of the CS&E community, not necessarily affiliated with ECP, can also subscribe to the (monthly) ECP Training Newsletter, which has been recently redesigned, and stay informed about training activities that are open to more than just members of ECP.

A notable component of ECP's training effort is the Best Practices for HPC Software Developers (HPC-BP) series of webinars, which is carried out in coordination with the IDEAS-ECP Productivity Project (see Box 2). HPC-BP is opened to the general public, all is needed is a "ticket reservation," which is used for planning and statistics purposes. More information about the HPC-BP series can be found in [3], including the process used for delivering each webinar. We aim at providing a reproducible workflow, not only to produce the webinar series but also to extract best practices that that can be helpful to the community. This includes early interactions with the presenters to ascertain that the webinars contain "big-picture takeaways" and pointers to supplementary information. Subsequently, we request presenters to curate the answers to eventual questions from the audience (i.e., in a Q&A document), and add more information as needed. At the time of this writing, 33 webinars have been offered, as listed in Table II. The speakers are from a variety of institutions, including Europe. The average number of tickets we have "sold" (through Eventbrite) for each webinar is 152, which includes (in average) 47 "affiliated with EC." Actual participation in the webinars has been (in average) 50% of the number of tickets sold. Many participants in the webinars are recurrent. Concerning participants' affiliations, and as an example, the attendance of the 2020 May webinar was comprised of 30% academia, 56% research laboratories, and 14% industry. Currently, of particular interest to the series are topics that serve as motivation and preparation for the exascale machines that will be fielded at ASCR's computing facilities. The webinars about the program models Kokkos and SYCL are examples of those. While Kokkos and SYCL warrant more than a one-hour webinar, these presentations pave the way for further learning: we are currently interacting with the developers of Kokkos for the development of an online course on that topic.

III. Available Resources and Final Remarks

At the time of this writing, we have organized or been involved in more than 130 training events (since the start of ECP in September 2016), which have been attended by more than 5,000 people. (We keep track of the events in an internal Jira dashboard for the HI component of ECP.) Information about a good fraction of these events can be found under the training tab of the ECP website [2]. Specifically, the memory of the HPC-BP webinar series (slides, Q&A documents) is preserved under *events* in [5]; recordings are available at the *ExascaleComputingProject* and *IDEASProductivity* YouTube channels. Given the constraints posed by the COVID-19 pandemics, readers may be interested in the panel series *Strategies for Working Remotely*, found under *events* in [5], in which the panelists discuss experiences in working remotely, lessons learned, unforeseen benefits, etc.

As ECP moves to completion (in mid 2023), we will continue to parse the needs of the ECP community and set up a matching T&P training agenda. In this agenda, we will seek to identify best practices that can be adopted or

TABLE II: Webinars of the series *Best Practices for HPC Software Developers* in ECP.

Date	Title, Presenter(s)
2020-june	*SYCL – Introduction and Best Practices*, Thomas Applencourt (ALCF)
2020-may	*Accelerating Numerical Software Libraries with Multi-Precision Algorithms*, Hartwig Anzt (KIT) and Piotr Luszczek (UTK)
2020-apr	*Best Practices for Using Proxy Applications as Benchmarks*, David Richards (LLNL) and Joe Glenski (HPE)
2020-mar	*Testing: Strategies When Learning Programming Models and Using High-Performance Libraries*, Balint Joo (JLAB)
2020-feb	*Introduction to Kokkos*, Christian Trott (SNL)
2020-jan	*Refactoring EXAALT MD for Emerging Architectures*, A. Thompson and S. Moore (SNL), and R. Gayatri (NERSC)
2019-dec	*Building Community through xSDK Software Policies*, U. M. Yang (LLNL) and P. Luszczek (UTK)
2019-oct	*Tools and Techniques for Floating-Point Analysis*, I. Laguna (LLNL)
2019-sep	*Discovering and Addressing Social Challenges in the Evolution of Scientific Software Projects*, R. Gassmoeller (UC Davis)
2019-aug	*Software Management Plans in Research Projects*, S. A. Sufi (SSI)
2019-jul	*When 100 Flops/Watt was a Giant Leap: The Apollo Guidance Computer Hardware, Software and Application in Moon Missions*, M. Miller (LLNL)
2019-jun	*Modern C++ for High-Performance Computing*, A. Lumsdaine (PNNL & U. of Washington)
2019-may	*So You Want to be Agile? Strategies for Introducing Agility into Your Scientific Software Project*, M. Heroux (SNL)
2019-apr	*Testing Fortran Software with pFUnit*, T. Clune (NASA Goddard)
2019-mar	*Parallel I/O with HDF5 - Overview, Tuning and New Features*, Q. Koziol (NERSC)
2019-feb	*Containers in HPC*, S. Canon (NERSC)
2019-jan	*Quantitatively Assessing Performance Portability with Roofline*, J. Pennycook (Intel), C. Yang and J. Deslippe (NERSC)
2018-dec	*Introduction to Software Licensing*, D. Bernholdt (ORNL)
2018-oct	*Open Source Best Practices: From Continuous Integration to Static Linters*, D. Smith and B. Pritchard (MSSI)
2018-sep	*Modern CMake*, B. Hoffman (Kitware)
2018-aug	*Software Sustainability: Lessons Learned from Different Disciplines*, N. C. Hong (SSI)
2018-jul	*How Open Source Software Supports the Largest Computers on the Planet*, I. Lee (LLNL)
2018-jun	*Popper: Creating Reproducible Computational and Data Science Experimentation Pipelines*, I. Jimenez (UCSC)
2018-may	*On-demand Learning for Better Scientific Software: How to Use Resources and Technology to Optimize your Productivity*, E. Raybourn (SNL)
2018-apr	*Software Citation Today and Tomorrow*, D. Katz (NCSA and UIUC)
2018-mar	*Scientific Software Development with Eclipse*, G. Watson (ORNL)
2018-feb	*Jupyter and HPC: Current State and Future Roadmap*, M. Bussonnier (UC Berkeley), S. Somnath (ORNL) and S. Cholia (NERSC)
2018-jan	*Bringing Best Practices to a Long-Lived Production Code*, C. Ferenbaugh (LANL)
2017-dec	*Better Scientific Software (https://bssw.io): So your code will see the future*, M. Heroux (SNL) and L. C. McInnes (ANL)
2017-nov	*Managing Defects in HPC Software Development*, T. Evans (ORNL)
2017-sep	*Barely Sufficient Project Management: A few techniques for improving your scientific software development efforts*, M. Heroux (SNL)
2017-aug	*Using the Roofline Model and Intel Advisor*, S. Williams and T. Koskela (LBNL)
2017-jul	*Intermediate Git*, R. Bartlett (SNL)
2017-jun	*Python in HPC*, R. Thomas (NERSC), W. Scullin (ALCF) and M. Belhorn (OLCF)

adapted by the CS&E community in general. (We encourage readers to monitor the training section of [2].) Apart from the technical contents of the events we organize, we are interested in exploring innovative modes for delivering online content, which we think would be beneficial for training across a distributed community. This is also motivated by the observation (instructors' and ours) that events that involve hands-on exercises or demos, for example, tend to be more effective in a face-to-face environment.

Acknowledgments

This work was supported by the Exascale Computing Project (ECP 17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. The authors are grateful to the presenters of webinars and tutorials mentioned in this article; to people who have provided suggestions for topics and feedback on draft material; and to people who have helped with the logistics of the various T&P events. The work of many has been essential for making a vast amount of information available to the CS&E community.

References

[1] ATPESC, "Argonne Training Program on Extreme-Scale Computing," https://extremecomputingtraining.anl.gov.

[2] The Exascale Computing Project, https://www.exascaleproject.org.

[3] O. Marques, "Producing a Webinar Series," https://bssw.io/items/producing-a-webinar-series, 2019.

[4] ECPAM20, "2020 Exascale Computing Project Annual Meeting," https://www.ecpannualmeeting.com/overview.php, 2020.

[5] IDEAS, https://ideas-productivity.org.

Box 1: The ECP Annual Meeting

The ECP Annual Meeting is paramount to the success of the ECP endeavor. The ECP community includes researchers, domain scientists, mathematicians, computer and computational scientists, U.S. HPC vendors, project management experts, an external advisory group, an industry council (formed by senior technology executives from prominent industrial organizations), and ECP's sponsors and program managers. The meeting highlights technical accomplishments of the ECP teams, and provides a collaborative environment that includes featured speakers, workshops, tutorials, and numerous planning and co-design meetings. Participants are usually unable to attend all discussions and training opportunities that could be relevant to their work: the 2020 annual meeting [4] included a diverse set of 53 BoFs, breakout sessions and panels, and 35 tutorials of various lengths. During the year, we seek to offer training events that are motivated by the agenda of the ECP annual meeting. As an added value, these events are typically open to the participation of non ECP affiliates.

Box 2: IDEAS-ECP

The Interoperable Design of Extreme-scale Application Software (IDEAS) Project [5], also part of the HI area, comprises a set of activities addressing challenges in software development productivity and software sustainabilty in CS&E on high-performance computers. These activities include advanced methodologies for application productivity (e.g., agile workflows for scientific software, supported by metrics and diagnostics to gauge progress); customizable resources for improving the development of scientific application codes (e.g., through software productivity and sustainability plans); engagement with ECP teams to gradually improve software development practices; and training, outreach and community building. As part of its outreach, IDEAS-ECP has offered the *Better Scientific Software* tutorial at various venues, including SC conferences.