The HPC Best Practices Webinar Series

Osni A. Marques*
Lawrence Berkeley National Laboratory
oamarques@lbl.gov

David E. Bernholdt†
Oak Ridge National Laboratory
bernholdtde@ornl.gov

Elaine M. Raybourn‡
Sandia National Laboratories
emraybo@sandia.gov

Ashley D. Barker†
Oak Ridge National Laboratory
ashley@ornl.gov

Rebecca J. Hartman–Baker*
Lawrence Berkeley National Laboratory
rjhartmanbaker@lbl.gov

ABSTRACT

In this contribution, we discuss our 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. HPC-BP is an outreach component of the IDEAS Productivity Project [4] and has been designed to support the IDEAS mission to work with scientific software development teams to enhance their productivity and the sustainability of their codes. The series, which was launched in 2016, has just presented its 22nd webinar. We summarize and distill our experiences with these webinars, including what we consider to be "best practices" in the execution of both individual webinars and a long-running series like HPC-BP. We also discuss future opportunities and challenges in continuing the series.

1 INTRODUCTION

The Best Practices for HPC Software Developers (HPC-BP) webinar series is a major component of the outreach efforts of the IDEAS Productivity Project[1]. Since its inception, the project has been addressing the confluence of trends in hardware and increasing demands for predictive multiscale, multiphysics simulations. The project has also been responding to trends for continuous refactoring with efficient agile software engineering methodologies and improved software design. The webinar series began when the IDEAS Project was particularly focused on working with the terrestrial ecosystem modeling community within the U.S. Department of Energy (DOE), and the series has been adapted as the focus (and sponsorship) of the IDEAS Project shifted to DOE's Exascale Computing Project (ECP) [5], wherein the focus is on helping teams in both scientific applications and software technologies to be more productive in their software development efforts and to produce more sustainable results.

To achieve its goals, IDEAS-ECP (as we refer to this phase of the IDEAS Productivity Project) has implemented an agenda for training and outreach, including HPC-BP. The target audience for the series largely overlaps with the community of users of the computer facilities under DOE's Office of Science: the Argonne Leadership Class Facility (ALCF) [2], the Oak Ridge Leadership Class Facility (OLCF) [3], and the National Energy Research Scientific Computing Center (NERSC) [1]. However, the webinars can be (and typically are) attended by a much broader community: announcements are done through various email lists, participation is free, and only a simple registration is required for each event. The series may have similarities with other efforts (e.g., [6, 7]), but it is also distinct in many ways. For example, when the webinar focuses on tools, it is often necessary to address differences in the corresponding installations at the computing facilities.

In this contribution, we discuss the process, i.e., the *HowTo*, that we have adopted for HPC-BP, and we describe some webinars that we have organized and delivered in the series. We provide an overview of the process we follow for the selection of topics, how the webinars are executed, and the future we foresee for the series.

2 SELECTION OF TOPICS

The webinars in the HPC-BP series occur approximately once a month and last about one hour each. Earlier in the series, the topics were set well in advance. Currently, we maintain a dynamic pool of potential topics (and speakers), from which we pull the ones that we consider to be timely. This decision is based on interactions with ECP application teams; interactions with staff at ALCF, OLCF and NERSC; perceived trends in hardware evolution; and discussions

*This manuscript has been authored by an author at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges, that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

†This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

‡Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

[1]IDEAS stands for *Interoperable Design of Extreme-scale Application Software*.

https://doi.org/10.22369/issn.2153-4136/10/1/19

among members of IDEAS-ECP. In addition, on several occasions the series has featured presentations by "volunteers," i.e., members of the HPC community at large who learned about HPC-BP and contacted us about the suitability of their topics for the series.

At the time of this writing, we have delivered 22 webinars in the HPC-BP series. Seven webinars in 2016 were in conjunction with an earlier phase of the IDEAS project. We relaunched the webinar series in June 2017 as part of the IDEAS-ECP project, and have so far delivered 15 webinars on an approximately monthly cadence. The following is a sample of topics that have been covered: intermediate Git (presenting Git in terms of a data structure and a set of algorithms to manipulate that data structure); the Roofline Model (introducing the concepts of data locality and arithmetic intensity, and the impact of their interplay in computational performance); the management of defects in HPC software development (discussing the relevance of software verification as a method for removing defects at and earlier development phase, and its impact on productivity); scientific software development with Eclipse (focusing on the latest features that are mostly useful for scientific applications); on-demand learning for better software development (demonstrating how a personalized transmedia learning framework based on on-line resources can be applied towards learning more productively); the importance of proper citation mechanisms for software (so software developers and maintainers can get academic credit for their work); and software sustainability seen from different angles (cultural issues, software sharing, adoption of best practices, etc.)

Usually, we select a topic at least two months in advance, marking the beginning of our "formal" interactions with the presenter(s). Speakers are asked to prepare a 45-minute presentation and include breaks to answer potential questions from the participants. About six weeks before the webinar we ask the presenter for a title, abstract and a short bio. About a month before the webinar we prepare a registration page and start an announcement campaign. We also schedule a dry-run, which is typically conducted the week prior to the webinar. The dry-run is usually attended by a few members of the IDEAS-ECP project, and has proven to be essential for fine-tuning the presentation. The day before the webinar we send a link to the event to the registrants, together with a copy of the slides to be presented.

3 MISE EN SCÈNE

Based on our experience, we have developed a model for a number of distinct roles and steps in the process of developing and delivering a specific webinar event. Figure 1 provides an illustration. The letters to the right of the figure indicate the steps, the corresponding "actors" and their (leading or supporting) roles:

(a) Actor: meeting manager. Roles: displays introductory slide (typically prepared by the moderator), which contains information about how participants can ask questions and provide feedback; monitors connections and checks whether participants are having problems with video/audio; starts recording.
(b) Actor: moderator. Role: introduces speaker(s) and other people that will help answer participants' questions.

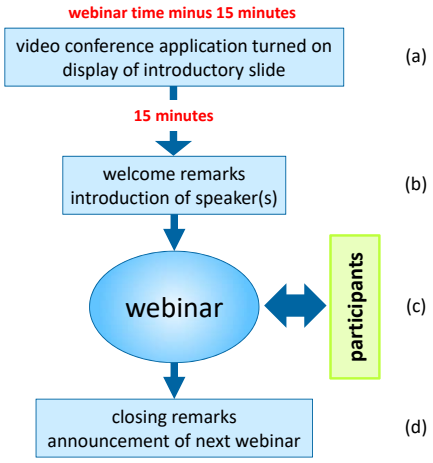

Figure 1: Steps and actors in HPC-BP webinar series.

(c) Actor: speaker. The meeting manager and the moderator check questions submitted through chat and Google doc and relay them to the speaker at specific breaks.
(d) Actors: moderator and meeting manager. Roles: ask audience whether there are additional questions; display slide announcing the next webinar; end the recording and video service.

The number of registrants for the webinars has ranged from 69 to 207, and our experience is that roughly 50% of those who register actually attend the event. Table 1 gives the total number of registrations, registrations affiliated with ECP, and number of participants in the webinars of the HPC-BP series. We observe that the number of registrations varies significantly across the 15 webinars. We believe this can be attributed to a variety of sources. Although we haven't attempted to analyze the variation, we believe that the webinar topic is likely to be the most significant driver for registrations. Speakers who are well-known in the community may also attract more registrations. The detailed data also suggests the possibility of seasonal variation, but we do not yet have sufficient experience to consider this conclusive. There are also variations in how far in advance we are able to get our announcements out, which may impact registration rates. In terms of timing, we have a specific time slot (1:00–2:00pm U.S. Eastern Time, on a Wednesday) that we have successfully held to for the vast majority of our webinars. However the week in the month will vary, so the schedule is not completely regular.

We note that the level of interest in the series, as measured by the number of people registering, and the number of participants, often significantly exceeds the typical experience that the ASCR facilities have with HPC-oriented training. We also note that participation of the audience has motivated us to include topics in our pool, or to consider offering webinar topics in different formats such as workshops or tutorials (e.g., for Eclipse and CMake).

Questions from the participants are highly encouraged, and are accepted through the chat capability in the webinar tool and also a shared Google Doc. While participants are muted by default, in

some cases we allowed participants to directly ask specific questions (by unmuting) to the presenter(s) at the end of the webinar.

Recordings of the webinars, along with the corresponding slides, are posted about a week after the webinar, together with a revised and curated Q&A (i.e., answers to the questions from the participants). An announcement of the "archival" copy of the webinar is distributed to those who registered for and/or attended the webinar, in case they wish to refer back to it or share with colleagues. All the relevant information is available under the "Events" tab in the IDEAS web site[4].

Table 1: Registrations and attendees in the 15 webinars of the HPC-BP Series (June 2017 - September 2018).

	Total Registrations	ECP Affiliated Registrations	Attendees
Minimum	69	22	22
Average	134	42	75
Maximum	210	71	127
Std. Deviation	50	15	35

4 OVERALL CONSIDERATIONS AND NEXT STEPS

One of the unique features of HPC-BP is that its agenda is implemented in concert with three major computer facilities (i.e. ALCF, OLCF and NERSC). This feature also poses challenges: for example, one of our webinars focused on Python in HPC. Maximizing performance from Python applications can be demanding on supercomputing architectures, even more so for different installations. The approach we used for the webinar was to have it presented by three staff, one from each facility. This exercise also presented an opportunity for an exchange of knowledge among the facilities.

One challenge is to capture meaningful data for success metrics. Although each workshop is accompanied by a survey, the number of respondents is typically low. Tracking the number of registrants and participants seems more appropriate and straightforward. (Sometimes, one registration corresponds to multiple people viewing the event from a single location.) We take the number of questions in the Q&A into consideration for potential follow-ups on the topic, e.g., further training through tutorials.

As mentioned before, IDEAS-ECP is particularly (though not exclusively) focused on topics relating to software productivity and sustainability, and also works in collaboration with the Training component of the ECP project. The training team is in the process of compiling a list of tools, programming languages, programming models and runtimes, etc., which will be used to construct a survey for ECP developers of applications and software technology to gauge their needs for training, plus the optimal format(s) (webinar, tutorial, etc.) and level(s) (novice, etc.). We anticipate that this survey will enable us to identify opportunities to further enrich the pool of topics related to software productivity and sustainability and further expand our outreach activities.

Looking into the future, once the early exascale computers are in production and the ECP has completed its mission, the IDEAS Productivity Project will need to find new ways to support and sustain the HPC-BP webinar series.

ACKNOWLEDGMENTS

This research 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.

This work was carried out in part at Lawrence Berkeley National Laboratory, managed by the University of California for the U.S. Department of Energy under contract number DE-AC02-05CH11231.

This work was carried out in part at Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract number DE-AC05-00OR22725.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] National Energy Research Scientific Computing Center. 2018. http://www.nersc.gov. (2018).
[2] Argonne Leadership Computing Facility. 2018. https://www.alcf.anl.gov. (2018).
[3] Oak Ridge Leadership Computing Facility. 2018. https://www.olcf.ornl.gov. (2018).
[4] IDEAS. 2018. https://ideas-productivity.org. (2018).
[5] The Exascale Computing Project. 2018. https://www.exascaleproject.org. (2018).
[6] Blue Waters Webinars Series. 2018. https://bluewaters.ncsa.illinois.edu/webinars. (2018).
[7] ACM Learning Webinars. 2018. https://learning.acm.org/webinars. (2018).