

HPC Carpentry—A Scalable, Peer-reviewed Training Program to Democratize HPC Access

Andrew Reid*
National Institute of Standards &
Technology
andrew.reid@nist.gov

Alan Ó Cais
University of Barcelona
alan.ocais@cecam.org

Trevor Keller
National Institute of Standards &
Technology
trevor.keller@nist.gov

Wirawan Purwanto
Old Dominion University
wpurwant@odu.edu

Annajiat Alim Rasel
Brac University
annajiat@gmail.com

ABSTRACT

The HPC Carpentry lesson program is a highly interactive, hands-on approach to getting users up to speed on HPC cluster systems. It is motivated by the increasing availability of cluster resources to a wide range of user groups, many of whom come from communities that have not traditionally used HPC systems. We adopt the Carpentries approach to pedagogy, which consists of a workshop setting where learners type along with instructors while working through the instructional steps, building up “muscle memory” of the tasks, further reinforced by challenge exercises at critical points within the lesson. This paper reviews the development of the HPC Carpentry Lesson Program as it becomes the first entrant into phase 2 of The Carpentries Lesson Program Incubator. This incubator is the pathway for HPC Carpentry to become an official lesson program of The Carpentries.

KEYWORDS

HPC, training, open source, lesson

1 INTRODUCTION

HPC Carpentry is a set of lessons whose goal is to introduce the ins and outs of running applications on HPC resources to new audiences, including investigators from fields which are not traditional users of high-performance computing, as well as novice users from fields in which HPC is commonly used.

For historical reasons, we have a variety of lessons at varying stages of maturity, and are in the process of crafting some of these lessons into workshops or tracks, which can be presented together to bring learners up to speed on modern HPC resources.

1.1 The Carpentries

We take advantage of pedagogical methods built up by The Carpentries [3]. This organization was founded 25 years ago to solve a related problem, that of getting scientists to engage better with

*Presenter, all authors contributed equally to this submission.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Copyright ©JOCSE, a supported publication of the Shodor Education Foundation Inc.

code development best practices, to help investigative teams capture knowledge generated by possibly-short-term participants, and to improve reproducibility of computational science.

The fundamental idea of Carpentries workshops is two-fold. Firstly, within the workshop setting, instructors and learners live-code the lesson steps together, with a strong emphasis on ensuring that instructors explain the process in detail, including real-time post-mortems on typos and errors. Learners exercise the correct workflow, building up “muscle memory” of the process, and also get an organic sense of what common errors look like, and how to recover from them. Secondly, the workshop material is open-source, and feedback gathered after a workshop, from learners, instructors, and organizers, and can be translated into modifications to the lessons, so that they are continuously improved. Because the lessons are shared across the Carpentries community, there are many eyes on the content, and “bugs” are quickly found and corrected for all future users.

The limited time of the workshop setting admits modest but important pedagogical goals – learners are expected to come away not as experts in the material they have just worked through, but rather with a solid grounding in what success looks like, search terms to drive later discovery, and continued access to the lesson material itself.

2 HPC CARPENTRY LESSON PROGRAM

HPC Carpentry addresses a similar need to The Carpentries, namely on-boarding novice HPC users. Many of the same constraints that motivated The Carpentries strategy are similar – there is no room in anyone’s curriculum for a formal course in HPC usage, but the material is unfamiliar, and HPC interactions are unlike other ways that computers are used. The same strengths of The Carpentries – the workshop setting and live-typing technique – help novice HPC users just as they do novice users of Git, Bash, or Python.

2.1 A Brief History

Birds of a Feather sessions at SC17 [13] and SC21 [11] demonstrate that, for some time, there has been significant interest among the HPC community in The Carpentries approach to generating and delivering training content.

The HPC Carpentry GitHub organisation was created in 2017 as part of HPC lesson development efforts by Compute Canada (now Digital Research Alliance of Canada). There were also other ongoing HPC lesson development efforts such as the HPC Parallel

novice lesson [12]. At CarpentryCon 2018, HPC Carpentry had 2 sessions [2, 9] (each with ~40 participants), where much of the discussion centered around how to merge existing efforts and form a single group of collaborators to drive the lesson development forward.

Over the past few years, in working towards Carpentries lesson-program incubation status, the current team has identified components which can be organized into thematic Carpentries-style workshops. The themes identified are a user workshop, taking learners from an introduction to the shell through to running a simple application on a cluster, and a developer workshop, taking learners through the execution of a code example for a parallel framework, such as MPI. At the same time, we do not wish to devalue or discard the niche or advanced material which we have access to, and which does not neatly fit into the workshop themes. Identifying and developing the workshop-appropriate material has been one of the significant challenges of the past few years.

Over the course of all of this development, a number of teams have forked lesson material on GitHub at various points in its development. We recently reached out to many of these teams to try to ascertain whether there were valuable additions to the material that had been added, that we might want to capture, and to try to identify opportunities for re-integration or collaboration. This effort promises to rekindle some prior relationships, and we are hopeful that this will help us serve a broader community than would otherwise be reached, but re-integration has certainly been another significant challenge.

2.2 Lessons in the Program

The current user-oriented lesson program of HPC Carpentry consists of three lessons:

- **“The Unix Shell”** [1] – This lesson is included directly from Software Carpentry. To quote from the lesson itself: “Use of the shell is fundamental to a wide range of advanced computing tasks, including high-performance computing.”
- **“Introduction to High Performance Computing”** [5] – The most highly developed of our lessons, which takes learners from basic shell use to dispatching parallel jobs on an HPC cluster, and includes careful feedback-driven descriptions of various jargon terms.
- **“HPC Workflow Management with Snakemake”** [4] – Currently under development, this lesson is meant to follow the “Introduction to HPC” in a workshop setting. It takes the user from dispatching jobs to performing a workflow-managed scaling study on a reference executable which illustrates Amdahl’s law (using Snakemake [10] as the workflow tool).

As mentioned earlier, for historical reasons, there are a number of other lessons in The Carpentries format, created and maintained by the wider community, which encapsulate valuable material but are not part of the initial workshop plan. Examples include a lesson on the Chapel program language [6] and the use of containers [7]. The Carpentries lesson incubator provides a venue where relevant new material can be identified, and possible new contributors brought into the community.

2.3 Continuous Improvement Through Continuous Assessment

The Carpentries model seeks to actively gather learner feedback to foster lesson improvement. Each delivery of the lesson program includes a pre- and post- workshop surveys for each Carpentries Workshop to help evaluate the effectiveness of the lesson and identify opportunities for further improvement.

2.4 Portability

All of our lessons are maintained in public GitHub repositories. The lesson “Introduction to High Performance Computing” currently allows for extensive customisation for use at a particular site. For the purposes of providing community access to HPC resources, we are planning to converge our lessons such that the default content refers to a “reference cluster,” built in the cloud, that meets our general requirements and can be easily reproduced (and hence scaled out). Adapting the lessons to a specific site, for workshops that benefit from this, will be still be possible.

2.5 Supporting Infrastructure

We currently have access to cloud resources, where we use Magic Castle [8] to create instances of “HPC clusters” that can serve the requirements of the lesson program. We anticipate that our relationship with The Carpentries will help us to acquire and manage additional cloud resources, either in-kind or via funding, which should facilitate future workshops.

3 FUTURE DIRECTIONS

Our near-term strategy at this point is to complete and begin teaching the material for the two workshops mentioned above, one for HPC users, and one for developers. The workshops will both start with the existing “The Unix Shell” lesson, maintained by The Carpentries, followed by an “Introduction to High Performance Computing”. Afterwards, the workshops will diverge, with the “user” workshop continuing with the “HPC Workflow Management with Snakemake” lesson, and the “developer” workshop continuing with a lesson in the operation of a parallel framework, such as MPI.

3.1 The Carpentries Lesson Program Incubator

With the ongoing support of The Carpentries, it is hoped that the HPC user-oriented workshop described above will be the first entrant into Phase 2 of The Carpentries Lesson Program incubation process. Incubation is a 3 phase process with the final phase being adoption as an official lesson program of The Carpentries (alongside Software Carpentry, Data Carpentry and Library Carpentry).

3.2 Development of Additional Lessons

Aside from the lessons intended for workshop integration, we also plan to continue to welcome and disseminate contributions on general HPC topics. We are aware that HPC is used for many things, so we can imagine application-specific lessons for popular HPC applications, or lessons for more compact and expressive languages for HPC applications.

3.3 Engagement with HPC Education and Training Communities

As we enter a new incubation phase with The Carpentries, we wish to also increase our engagement with the wider HPC education and training communities. The continuous improvement model we use relies on a steady stream of instructors, workshops and integration of feedback from those workshops. Connection with this group can help to streamline this process.

4 CONCLUSION

HPC Carpentry is entering an exciting phase, where seeds that were planted years ago are starting to bear fruit. HPC Carpentry has honed an initial lesson program to the point where it is being considered as an official lesson program of the Carpentries. It has addressed, and found workable solutions for, complex issues such as access to HPC resources for training purposes, and site-specific customisation of community-maintained lessons.

ACKNOWLEDGMENTS

The authors would like to thank The Carpentries for their ongoing engagement with this effort. Alan Ó Cais is supported by Digital Europe Programme under Grant No. 101100604 (BioNT).

REFERENCES

- [1] Inigo Aldazabal Mensa, Harriet Alexander, James Allen, Areej Alsheikh-Hussain, Daniel Baird, Piotr Banaszekiewicz, Pauline Barmby, Rob Beagrie, Trevor Bekolay, Evgenij Belikov, Jason Bell, Kai Blin, John Blischak, Simon Boardman, Maxime Boissoneault, Jessica Bonnie, Andy Boughton, Ry4an Brase, Amy Brown, Dana Brunson, Orion Buske, Abigail Cabunoc Mayes, Daniel Chen, Kathy Chung, Gabriel A. Devenyi, Emily Dolson, Jonah Duckles, Rémi Emonet, David Eyers, Filipe Fernandes, Hugues Fontenelle, Francis Gacenga, Matthew Gidden, Ivan Gonzalez, Norman Gray, Varda F. Hagh, Michael Hansen, Emelie Harstad, Adina Howe, Fatma Imamoglu, Damien Irving, Mike Jackson, Emily Jane McTavish, Michael Jennings, Dan Jones, Alix Keener, Kristopher Keipert, Tom Kelly, Jan T. Kim, W. Trevor King, Christina Koch, Bernhard Konrad, Sherry Lake, Doug Latornell, Philip Lijnzaad, Eric Ma, Joshua Madin, Camille Marini, Kunal Marwaha, Sergey Mashchenko, François Michonneau, Ryan Middleson, Jackie Milhans, Bill Mills, Amanda Miotto, Sarah Mount, Lex Nederbragt, Daiva Nielsen, Aaron O’Leary, Randy Olson, Adam Orr, Nina Therkildsen, Kirill Palamartchouk, Adam Perry, Jon Pipitone, Timothée Poisot, Hossein Pourreza, Timothy Poval, Adam Richie-Halford, Scott Ritchie, Noam Ross, Halfdan Rydbeck, Mahdi Sadjadi, Pat Schloss, Bertie Seyffert, Genevieve Shattow, Raniere Silva, Sarah Simpkin, John Simpson, Byron Smith, Nicola Soranzo, Ashwin Srinath, Daniel Standage, Meg Staton, Peter Steinbach, Marcel Stimberg, Bartosz Telenczuk, Florian Thoele, Tiffany Timbers, Stephen Turner, Jay van Schyndel, Anelda van der Walt, David Vollmer, Jens von der Linden, Andrew Walker, Josh Waterfall, Ethan White, Carol Willing, Greg Wilson, Donny Winston, Lynn Young, and Lee Zamparo. 2016. Software Carpentry: The Unix Shell. <https://doi.org/10.5281/zenodo.57544>
- [2] Alan Ó Cais and Daniel Smith. 2018. Breakout 8: HPC Carpentry. <https://github.com/carpentries/carpentrycon/tree/master/CarpentryCon-2018/Sessions/2018-05-31/05-Breakout-8-HPC-Carpentry>. Breakout session at CarpentryCon 2018.
- [3] The Carpentries. 2023. The Carpentries teaches foundational coding and data science skills to researchers worldwide. <https://carpentries.org/>.
- [4] HPC Carpentry. 2023. HPC Workflow Management with Snakemake. <https://carpentries-incubator.github.io/hpc-workflows/>.
- [5] HPC Carpentry. 2023. Introduction to High Performance Computing. <https://carpentries-incubator.github.io/hpc-intro/>.
- [6] HPC Carpentry. 2023. Introduction to High-Performance Computing in Chapel. <https://www.hpc-carpentry.org/hpc-chapel/>.
- [7] Pawsey Supercomputing Centre. 2023. Containers on HPC and Cloud with Singularity. <https://pawsey-sc.github.io/singularity-containers/>.
- [8] Félix-Antoine Fortin, etienneDub, Frédéric Fortier-Chouinard, Maxime Boissoneault, Charles Coulombe, Erik Sundell, Simon Guilbault, Adam Huffman, jboschee, R. P. Taylor, Étienne Parent, Chris Want, Darren Boss, Justin Lagüe, Kenneth Hoste, Marie-Helene Burle, Quoding, and Pier-Luc St-Onge. 2023. ComputeCanada/magic_castle: Magic Castle 12.6.0. <https://doi.org/10.5281/zenodo.8096727>
- [9] Christina Koch and Peter Steinbach. 2018. Workshop 5: HPC Carpentry. <https://github.com/carpentries/carpentrycon/tree/master/CarpentryCon-2018/Sessions/2018-06-01/05-Workshop-5-HPC-Carpentry>. Workshop at CarpentryCon 2018.
- [10] Felix Mölder, Kim Philipp Jablonski, Brice Letcher, Michael B. Hall, Christopher H. Tomkins-Tinch, Vanessa Sochat, Jan Forster, Soohyun Lee, Sven O. Twardziok, Alexander Kanitz, Andreas Wilm, Manuel Holtgrewe, Sven Rahmann, Sven Nahsen, and Johannes Köster. 2021. Sustainable data analysis with Snakemake. *F1000Research* 10 (April 2021), 33. <https://doi.org/10.12688/f1000research.29032.2>
- [11] Andrew Reid, Trevor Keller, Alan Ó Cais, and Annajiat Alim Rasel. 2021. HPC Carpentry: Introducing New Users to HPC. <https://sc21.supercomputing.org/app/uploads/2021/11/SC21-Final-Program-211114.pdf>. BoF session at SC21.
- [12] Peter Steinbach. 2021. psteinb/hpc-parallel-novice: fork to work. <https://doi.org/10.5281/zenodo.4525377>
- [13] Andrew Turner, Christina Koch, Tracy Teal, Robert Freeman Jr, Chris Bording, and Martin Callaghan. 2017. HPC Carpentry - Practical, Hands-On HPC Training. https://sc17.supercomputing.org/index.html%3Fpost_type=page&p=5407&id=bof125&sess=sess359.html. BoF session at SC17.