HPC Carpentry: Recent Progress and Incubation Toward an Official Carpentries Lesson Program

Andrew Reid
US National Institute of Standards and Technology
Gaithersburg, Maryland
andrew.reid@nist.gov

Trevor Keller
US National Institute of Standards and Technology
Gaithersburg, Maryland
trevor.keller@nist.gov

Alan O'Cais
University of Barcelona
Barcelona, Spain
alan.ocais@gmail.com

Annajiat Alim Rasel
BRAC University
Dhaka, Bangladesh
annajiat@bracu.ac.bd

Wirawan Purwanto
Old Dominion University
Norfolk, Virginia
wpurwant@odu.edu

Jane Herriman
Lawrence Livermore National Laboratory
Livermore, California
herriman1@llnl.gov

Benson Muite
Kichakato Kizito
Nairobi, Kenya
benson_muite@emailplus.org

ABSTRACT

The HPC Carpentry project aims to develop highly interactive workshop training materials to empower HPC novices to effectively leverage HPC to solve scientific and technical problems in their respective domains. Modeled after the Carpentries training programs, the project's goal is to develop foundational HPC skills and a sense of self-efficacy, rather than expertise. The workshop setting provides learners with hands-on experience that elicits confidence working with HPC systems and provides sufficient vocabulary to make subsequent self-study more effective. The project is the product of significant work over the past several years, incorporating valuable materials from many contributors.

Our most recent focus has been developing workshop materials for new HPC users. We begin with an introduction to the command-line shell (using Software Carpentry's Unix Shell lesson), followed by our Introduction to HPC lesson, covering remote access and resource management. We end with a newly developed lesson on HPC workflow management, which walks learners through the specification and execution of a scaling study on an HPC system, emphasizing both the benefits and limitations of HPC systems for domain applications. This workshop program was recently run in full at the Lawrence Livermore National Laboratory.

Next, the project will develop training resources for HPC developers, which will give workshop instructors the option of replacing the workflow lesson with a coding exercise in a parallel framework, such as MPI.

© 2024 Journal of Computational Science Education
https://doi.org/10.22369/issn.2153-4136/x/x/x

In a major milestone, the steering committee is leading the project through the formal incubation process towards becoming an official Carpentries lesson program alongside the existing Software, Data, and Library Carpentry programs.

KEYWORDS

Cyberinfrastructure, training, pedagogy, HPC, parallel computing, big data, non-degree training, hands-on

1 BACKGROUND

HPC Carpentry [2] is an informal training project with a mission to provide a set of lessons aimed at introducing the basic "know-how" of running applications on high-performance computing (HPC) resources to new audiences, including investigators from fields which are not traditional users of HPC systems, as well as novice users from fields in which HPC is commonly used. Eventually, the projet's goal is to empower HPC novices to effectively leverage HPC to solve scientific and technical problems in their respective domains. The project paves the way for the potential users from non-traditional HPC disciplines to tap into HPC resources for their data analysis, modeling, and simulation needs while remaining relevant for beginners from the traditional HPC disciplines. The current project is the product of significant work over the past several years, incorporating valuable materials from many contributors.

2 LESSON DEVELOPMENT EFFORTS

The recent focus of HPC Carpentry has been the development a complete workshop program for new HPC users. We begin with an introduction to the command-line shell using Software Carpentry's Unix Shell lesson [1], followed by our Introduction to High Performance Computing lesson [3], covering remote access and resource management. We end with a newly developed lesson on HPC workflow management [4], which walks learners through the specification and execution of a scaling study on an HPC system,

emphasizing both the benefits and limitations of HPC systems for domain applications. This set of three lessons is sufficient to offer a two-day hands-on workshop in a format similar to that of the Software Carpentry or Data Carpentry workshops.

The project plans to develop more advanced training resources for HPC developers, which will give workshop instructors the option to substitute the workflow lesson with a coding exercise in a parallel framework (such as MPI), for example. Several lessons had also been developed in the past, including an introduction of parallel programming using the Chapel programming language. Furthermore, we have received engagement from HPC community members at large who explore the potential of merging their in-house lessons into HPC Carpentry's lesson portfolio. This is still an ongoing effort and engagement with the community.

3 RECENT WORKSHOPS

The complete HPC Carpentry workshop program for new users was recently offered at the Lawrence Livermore National Laboratory in June of 2024. Previous workshops were held at University College Dublin, Brac University, Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS), University of Mauritius, Florida International University, Delft University of Technology, National Institute of Standards and Technology, and EPFL CECAM. **FIXME: Can we say briefly what the same things about these workshops and what we learned after offering these workshops?**

One of the important issue we are facing in offering HPC Carpentry workshops is the need for HPC infrastructure for learners to use during the workshop. While some HPC site operators have their own HPC systems to conduct their own workshop, others, particularly from under-resourced institutions, do not have their own HPC resources. We have attempted to create a small HPC Carpentry cluster in the cloud. More recently, we have acquired support from Jetstream 2 through ACCESS to set up a "standard" reference HPC Carpentry cluster in a virtual-machine-based environment. The cluster set-up has been prepared in an automated fashion using Terraform scripting. This effort could pave the way to allow instructors to set their own clusters, irrespective of the existence of a local HPC cluster in their own institution.

4 INCUBATION TO THE CARPENTRIES LESSON PROGRAM

In a major milestone, the steering committee is leading the project through the formal incubation process towards becoming an official Carpentries lesson program alongside the existing Software, Data, and Library Carpentry programs. This process is expected to last for about 18 months, with an expected target date of December 2025.

5 ACKNOWLEDGMENT

We acknowledge the infrastructure support hosted on Jetstream 2 cloud environment. Allocation to Jetstream 2 was provided through ACCESS, which is funded by the US National Science Foundation.

REFERENCES

[1] Gerard Capes, bkmgit, Jacob Deppen, G. A. Devenyi, Alexander James Ball, Piper Fowler-Wright, Alessia Visconti, Jeff Dusenberry, Jessica Vera, Randal Sean Harrison, Andreas Bilke, Jessica Nicole Welch, Kelly Thorp, Alfredo Hernandez, Ashkan Mirzaee, Benjamin Winjum, Chris Daley, Clay Wright, colinmorris, Dave George, Ephantus2017, Erik Myklebust, Frank Löffler, HariEpuri, Holger Wolff, Kairsten Fay, Luna Luisa Sanchez Reyes, Marius Politze, Maxim Belkin, Nathaniel Porter, nkicg6, Norman Ziegner, Sarah LR Stevens, Sean McCartney, Serah Njambi, ramisetti, Stacey Borrego, Andrew Christopher Brown, Ashley Cryan, mehrdadbn9, Md Intekhabul Hafiz, niketagrawal, Noah Benson, Aaron McDivitt, Aidan Budd, Amanda Stahlke, Andraš Tšitškan, Andrew Stewart, Becky Smith, Catherine Martlin, Christian Knüpfer, David McKain, David Wilby, Dimitra Salmanidou, Dave Turner, Edan Scriven, Edward Wallace, Elizabeth McAulay, Etienne Roesch, Frank Solinsky, Giordano Lipari, Hamish Starling, Iain Barrass, Isil Poyraz Bilgin, JSheffield159, James Acris, Jonathan Bradley, Matti Juvonen, Kathryn Napier, Kenton Ross, Kevin Ernst, Lukas Trombach, Martin Chorley, Melissa, Mike Lake, Mike Renfro, Mike Renfro, NJ, Natali, Nathan McKinlay, Nicola Soranzo, Pablo Rodríguez-Sánchez, Peter Wiringa, Petr Viktorin, Richard Rigby, rkm, Ryan S. Elliott, Samuel Lelièvre, Santiago Lacalle, Sujai Kumar, Tong Liang, Winfred Gatua, Yi Sun, cgmerrick, daking4, Deep Patel, erich333, karl holten, kathymd, laporpe, naveendangeti, nbehrnd, sophie, tbert, Tom Couch, Ram Krishna Shrestha, zzhang60, and "Eli. 2023. swcarpentry/shell-novice: Software Carpentry: the UNIX shell. https://doi.org/10.5281/zenodo.7886938

[2] The HPC Carpentry Project. 2024. HPC Carpentry: Teaching Basic Skills for High-Performance Computing. https://www.hpc-carpentry.org/

[3] The HPC Carpentry Project. 2024. hpc-intro: Introduction to High-Performance Computing. https://carpentries-incubator.github.io/hpc-intro/

[4] The HPC Carpentry Project. 2024. hpc-workflow: HPC Workflow Management with Snakemake. https://carpentries-incubator.github.io/hpc-workflows/