HPC Education and Training: an Australian Perspective

Maciej Cytowski
Supercomputing Specialist
Pawsey Supercomputing Centre
Kensington WA Australia
maciej.cytowski@pawsey.org.au

Luke Edwards
Data Specialist
Pawsey Supercomputing Centre
Kensington WA Australia
luke.edwards@pawsey.org.au

Mark Gray
Cloud Lead
Pawsey Supercomputing Centre
Kensington WA Australia
mark.gray@pawsey.org.au

Christopher Harris
Senior Supercomputing Specialist
Pawsey Supercomputing Centre
Kensington WA Australia
chris.harris@pawsey.org.au

Karina Nunez
Marketing and Event Manager
Pawsey Supercomputing Centre
Kensington WA Australia
karina.nunez@pawsey.org.au

Aditi Subramanya
Marketing & Events Officer
Pawsey Supercomputing Centre
Kensington WA Australia
aditi.subramanya@pawsey.org.au

ABSTRACT

The Pawsey Supercomputing Centre has been running a variety of education, training and outreach activities addressed to all Australian researchers for a number of years. Based on experience and user feedback we have developed a mix of on-site and online training, roadshows, user forums and hackathon-type events. We have also developed an open repository of materials covering different aspects of HPC systems usage, parallel programming techniques as well as cloud and data resources usage. In this paper, we will share our experience in using different learning methods and tools to address specific educational and training purposes. The overall goal is to emphasise that there is no universal learning solution, instead, various solutions and platforms need to be carefully selected for different groups of interest.

KEYWORDS

Online training, Self-guided learning, HPC training and outreach

1 INTRODUCTION

The Pawsey Supercomputing Centre is constantly evolving its training programs to build a critical mass of advanced computing knowledge in the research community. It will continue to engage in a broad range of activities to further grow the expertise of the next generation of supercomputing specialists and create a skilled workforce in Australia. The success of our HPC training program is driven by its diversity. From basic computer science training for non-experienced users, through introductory and intermediate supercomputing and cloud, to parallel programming courses, GPU hackathons and customised training for research groups.

In this paper, we will address some of the most recent challenges in delivering HPC training and describe our ideas and experiences in facing them. We will share our experience in using different learning methods and tools to address specific educational and training purposes. Starting from traditional on-site training, through self-guided training materials to online training and webinars. The overall goal of this paper is to emphasise that there is no universal learning solution; instead, various solutions and platforms need to be carefully selected for different groups of interest.

2 PAWSEY TRAINING PROGRAMME

The Pawsey Supercomputing Centre has been developing and offering its training program for over ten years. It consists of various training modules related to HPC, Data, Cloud and Visualisation which are being offered to Australian researchers within the National Training Programme. The list of the training modules together with a short description of the content is presented in Table 1.

National Training is traditionally delivered on-site at universities and research institutions across Australia. Currently, it is composed of 5 training modules (first five modules listed in Table 1) presented during two days. The target audience of the training is researchers (students, PhD students, postdocs) who are willing to use infrastructure services offered by HPC centres for their research. It is required that the participants have a basic understanding of computer science. After two intensive days of training the participants should be able to (among others):

- understand HPC, Data and Cloud services offered in computing centres,
- understand what a shell program is and use basic Unix shell commands,
- setup, manage and use VMs in the Cloud environment,

- submit various types of jobs on HPC systems with the use of schedulers,
- use compilation environment on HPC systems,
- understand the concept of parallel file systems and their efficient use.

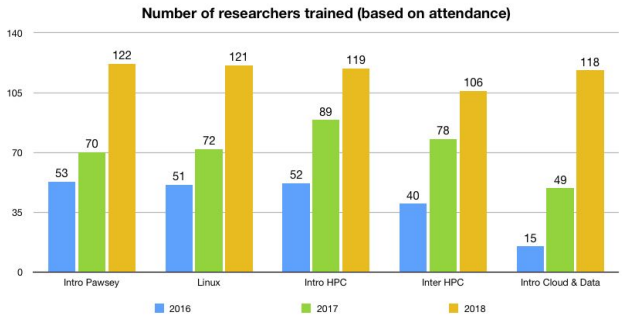

Figure 1. Attendance for the 5 core training modules.

We have gathered the attendance statistics for all training events provided over the years. Figure 1 presents attendance numbers for 5 core training modules of the National Training Programme over the past three years. Significant growth in attendance can be observed and we believe that this trend will continue in next years with some of the new online training offerings.

Training Module	Description of content and format
Introduction to Pawsey	Pawsey HPC, Data and Cloud resources. Usage scenarios. 30min presentation
Introduction to Unix	Intro to shell, navigating, pipes and filters, shell scripts, finding things. 1.5hr self-guided hands-on training
Introductory Supercomputing	Basic supercomputing concepts, supercomputing architectures, use of queueing systems. 3hr session: presentation + hands-on
Intermediate Supercomputing	Compiling codes on HPC systems, advanced workflows and queueing scripts, parallel file systems. 3hr session: presentation + hands-on
Using Nimbus: Cloud computing at Pawsey (Intro to Cloud)	Intro to cloud computing, How to create and launch a Nimbus VM (including making keypairs, security, attaching storage, managing instance). 3hr hands-on session
Introduction to Data Services	Intro to Pawsey Data services and good data management practices. 30min presentation
Introduction to pshell	Intro to command-line access to long-term data storage. 30min hands-on session
Introduction to Supercomputing Technology	Supercomputing building blocks: CPU architectures, memory hierarchy, nodes, interconnect, network topologies. 2hr session: presentation + hands-on
Remote Visualisation	Intro to remote visualisation tools offered at Pawsey. 3hr session: presentation + hands-on
Containerising Workflows	Intro to Docker, Discipline specific examples. 2-3hr session hands-on session
Optimising Serial Code	Programming languages, profiling codes, cache usage optimisations, basic loop transformations, vectorisation. 3hr session: presentation + hands-on
Parallelising your code with MPI	Distributed memory parallelisation concepts. Basic MPI: point-to-point communication, non-blocking communication, collectives. 3hr session: presentation + hands-on
Parallelising your code with OpenMP	Shared memory parallelisation concepts. OpenMP directives, library routines and environment variables. 3hr session: presentation + hands-on
GPU Programming Essentials	GPU parallelisation concepts. GPU architecture, developing codes with CUDA and OpenACC, GPU libraries. 2x 3hr sessions: presentation + hands-on

Table 1. Pawsey Training Programme (as of September 2018)

All hands-on training materials are available on Github: http://github.com/PawseySC
All presentations are available on Pawsey Centre user support pages: https://support.pawsey.org.au

Pawsey Supercomputing Centre has more than 1500 registered users affiliated with different research institutions across Australia. For that reason, Pawsey's National Training sessions are being offered in Australia's largest cities: Sydney, Melbourne, Perth, Brisbane and Hobart. As the travel costs associated with the organisation of those training programs are usually substantial, we try to

accommodate as many modules as possible within a single training. However, we have found that it is very hard to address different skill sets and interest of attendees with an increasing variety of topics presented during National Training. As a result, the most advanced training modules were usually less attended. Over the years we have also struggled to increase the turnout of our training events. Statistics for previous years are presented in Figure 2. Introducing small nominal registration fees was one of the considered solutions. This might potentially increase turnout, however, more organisation overhead is required and there is always a risk that the training might become less accessible, especially for students.

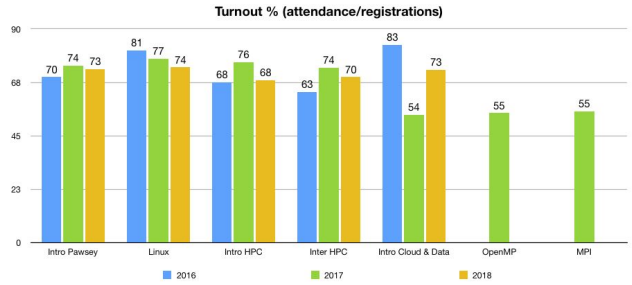

Figure 2. Turnout % (percentage calculated from actual attendance versus registration) for various training modules.

3 IN SEARCH OF THE NEW LEARNING FORMATS...

In the past, HPC developer courses were part of the National Training program. Training modules covering MPI, OpenMP and serial code optimisations were offered during the third day of the training. Unfortunately, we have found that with the proposed format it was particularly challenging to address different learners experience. Also, the number of registrations, as well as turnout of those training sessions, were often below expectations (Figure 2). To address those particular issues we have decided to experiment with a new online courses format designed especially for advanced HPC topics.

In the new format, parallel programming courses are divided into short webinar sessions (max. 90 minutes) organised throughout the week. Exercises are introduced to participants at the end of each session. Participants work on the solutions offline as access to HPC systems is granted to all participants throughout the whole week.

Interaction and communication are some of the most important aspects of training and education activities, which should be carefully addressed, especially in the case of online activities. Participants of the Pawsey online courses are encouraged to use chat room messaging during webinar sessions and a dedicated Slack channel for communication with Pawsey staff (or between themselves) while working offline on the solution of the exercises. This can be easily handled by two trainers, one presenting the material and the other keeping track and responding to chat messages.

We have found it very interesting that participants use both communication channels for exchanging their own experience and comments as well as helping themselves while working on the exercises. Participants can easily share their solutions by forking training materials repositories available on Github.

Online Course: Developing with MPI
Webinar 1 Introduction and Point-to-Point Communication **Exercise 1** Ping-Pong
Webinar 2 Non-blocking Communication and Communicators **Exercise 2** Message in a Ring
Webinar 3 Collectives, overview of other topics and next steps **Exercise 3** Game of Life

Table 2 Content and format of the Developing with MPI online course.

Online Course: GPU Programming Essentials
Webinar 1 GPU Computing at Pawsey **Exercise 1** Compile and run a GPU program
Webinar 2 Introduction to CUDA **Exercise 2** Host-device transfer and vector addition
Webinar 3 Programming with OpenACC **Exercise 3** Accelerating a Jacobi simulation
Webinar 4 GPU Libraries **Exercise 4** Using the cuBlas library

Table 3 Content and format of the GPU Programming Essentials online course.

We have run two HPC developer online courses as a pilot in June 2018. Contents and schedules of those courses are presented in Table 2 and Table 3. Both courses were very successful in terms of the number of registrations and participants. Although the turnout (actual webinar attendance versus registrations) was on the similar level as for the on-site training (around 50%), the registration numbers where much higher. We've got 54 registrations for the MPI course and 60 registrations for the GPU one. All registered participants received links to the webinar recordings. Therefore, potentially all registered participants

could follow the course even if, for any reason, they were not able to connect to the live webinar session.

The pilot run of two described online courses was followed by the “Overview of Containers in HPC” webinar which covered a brief introduction to containers, usage of containers on Pawsey systems as well as example workflows and benchmarks using containers. It was a huge success, with 87 overall registrations and more than 50 actual attendees.

4 SELF-GUIDED LEARNING

Basic computational science skills should be developed as part of a University’s curriculum, especially in computational science areas where computer simulations are a basic scientific tool. In practice, HPC centres still struggle to educate non-experienced computer users to use high-end computing infrastructure. We have identified self-guided learning platforms as particularly useful tools to address that problem. Specifically, our basic Unix training, which is based on Software Carpentry material [2] and uses the JupyterHub platform [3]. We deploy JupyterHub on our Pawsey cloud infrastructure (Nimbus), which provides us with the flexibility to give all users the same command-line environment to undertake hands-on exercises. This assists new users as they just need a browser and don’t need to be familiar with a shell / command-line environment. Limiting the diversity of command-line clients also reduces the troubleshooting issues faced by trainers and helps training stay on schedule.

Our Nimbus Cloud and Container training are also structured in the ‘Software Carpentry style’ which enables instructors to easily undertake hands-on and interactive training but also enables users to undertake training in their own time. We have also adopted the “One-Up, One-Down” Software Carpentry feedback approach [5] at the end of training each day. This approach involves the instructor asking the learners to alternately give one positive and one negative point about the day, without repeating anything that has already been said. We write the responses on a whiteboard and we have found this ‘encourages’ people to say things they otherwise might not, compare to post-training surveys. This has given us valuable feedback to improve our training.

5 COLLABORATIVE COURSES

As opportunities have arisen, Pawsey staff have collaborated with researchers and software developers to run various HPC and domain-specific courses, including areas such as quantum mechanics, radio astronomy, and fluid mechanics. The format of these courses has varied, from workshops over a couple of days to semester-long university courses.

Introductory HPC material is provided and delivered by Pawsey staff, and it is important that this occurs at the start of the course before the participants commence more specific learning that uses the systems. Our staff also manage processes for reserving resources and setting up accounts to support the coursework on our systems.

A critical factor in the success of these courses is the knowledge and expertise provided by the collaborating researcher or developer, and their willingness to provide effort to develop and deliver material.

6 BRIDGING GAPS

Carefully planned outreach activities can be extremely useful in bridging gaps between education programs and HPC training. The Pawsey Supercomputing Centre has not only continued to deliver training sessions to maximum capacity but has introduced a variety of complementary outreach activities addressed to different communities, students and research groups. These activities include roadshows, internship programs, careers nights, open days and community data centre tours [4]. During the talk, we will mention a few of those activities and share our experiences with running them.

Pawsey Roadshows are information sessions where researchers showcase their science and research to university students across Australia with focusing on how their outcomes have been positively impacted by high-compute power and staff expertise. This encourages the future scientists of the Nation to keep supercomputing at the forefront of their mind when the time comes for them to begin their research.

Though in previous years Pawsey Roadshows saw high attendance, 2018 numbers have been minimal to date. Although attendees found benefit in learning about supercomputing, it did not muster enough interest to increase attendance. Due to this, the Pawsey Uptake Group (PUG) and training committee trialled hosting roadshows within university events and research open days rather than stand-alone activities. This decision has proven to be successful with Roadshow attendance and engagement with Pawsey nearly doubling.

The Pawsey Supercomputing Centre Summer Internships is a good example of bridging gaps and developing skills for a new generation of HPC-ready researchers. The internship program runs for 10 weeks during the period November/December through to February. The internships are open to 3rd year, higher undergraduate students (including honours) or Masters students at Australian higher education institutions looking to

complement their discipline knowledge with hands-on HPC experience. Pawsey staff run an intensive week of hands-on training based on the existing Pawsey training program expanded to include other topics which would assist students in their projects, such as an introduction to Git and Python.

As part of our outreach activities, Pawsey hosted its first Careers Night in July in an endeavour to encourage year 10 students to pursue Science, Technology, Engineering and Mathematics (STEM) subjects in university. The careers night saw presentations from Pawsey staff, researchers and industry representatives - all with a story to tell and how studying a STEM subject have influenced their lives. After the presentations, the students spent their evening networking with the presenters and additional Pawsey staff to answer their curious questions.

Events such as the Careers Night reinforce Pawsey's focus on STEM and its importance in growing the minds of future scientists. It gives them insight into potential career paths and the ability of HPC to be part of any domain.

Similar to the Careers Night, every two years, Pawsey opens its doors to host a *Pawsey Open Day* as part of Australia's National Science Week. This day targets Perth's families of all ages and local government to raise awareness amongst the community about the benefits of supercomputers and the ground-breaking science Pawsey enables in Australia.

	2016	2017	2018
Pawsey Roadshow	56	139	26*
Student Summer Internship	16	14	12*
Career Night	-	-	52
Pawsey Open Day	300	-	500
Community Tours	28	66	43*

* as of October 2018

Table 4 Pawsey outreach activities to bridge gaps.

The Open Day includes back-to-back tours of the facility, 'designing a supercomputer' competition, researcher presentations and science activities (including child-friendly games that demonstrate parallelisation) to highlight the opportunities and encourage the ambition of young minds to work, research or be involved in HPC.

As a consequence of the first Open Day, Pawsey Community Tours are run every month and is an opportunity for the Perth residents to explore the Tier-1 facility that is sitting in their backyard. Any person over the age of 12 can tour through the Centre, which is predominantly guided by Pawsey's Head of Supercomputing. Tour groups are briefed with a *Science Showcase* of the work undertaken at the Centre, followed by a walkthrough of the supercomputing cell, I/O cell and tape cell.

Table 4 outlines the outreach activities discussed here and their corresponding attendance for the year.

7 SUMMARY AND FUTURE WORK

Pawsey has seen a marked change in the nature and size of its training and education program to better reflect the requirements of potential and existing Pawsey users. Through various feedback mechanisms, we seek to continue this change. In the future, we expect to increase the range of self-guided domain-specific training offerings, such as the recent bioinformatic training workshop we hosted on the 7th September 2018. This included specific material on running bioinformatics software from containers using our Nimbus cloud.

Similarly, we are planning to significantly increase the number and scope of training offered as online courses. Pawsey experts have found the organisation of those training very effective, both in terms of the attendance and the overall cost. We also recognise a large potential of those training activities. Although there is a number of similar webinars and online pieces of training available worldwide, most of them are hardly accessible due to the time difference.

We are also planning to continue the National Training Programme as this creates unique possibilities to reach out to research groups and to understand their particular interests and needs.

ACKNOWLEDGMENTS

We would like to acknowledge all past and present Pawsey Centre staff who actively contributed to the development and running of our training and outreach activities over the years.

REFERENCES

[1] Pawsey Supercomputing Centre Training Github Repository, https://github.com/PawseySC
[2] Software Carpentry, https://software-carpentry.org
[3] JupyterHub from Project Jupyter, http://jupyter.org/hub
[4] Pawsey Supercomputing Centre Annual Report 2016-2017
[5] Software Carpentry - Instructor training workshop https://carpentries.github.io/instructor-training/06-feedback/index.html