Delivering Digital Skills across the Digital Divide

An accessible self-paced on-demand HPC virtual training lab.

[†]Bryan Johnston
CSIR
Cape Town
South Africa
bjohnston@csir.co.za

[†]Lara Timm
CSIR
Cape Town
South Africa
ltimm@csir.co.za

[†]Mabatho Hashatsi
CSIR
Cape Town
South Africa
mhashatsi@csir.co.za

† Advanced Computer Engineering Lab (NICIS/CHPC)

ABSTRACT

The training of new and existing HPC practitioners is recognized as a priority in the HPC community[1][2]. The traditional approach to delivering HPC System Administrator training has historically been through physical face-to-face workshops[3][4][5], traditionally making use of a cloud-based service or remote hardware to provide compute resources to emulate an HPC system deployment [6]. There are several challenges associated with the traditional in-person approach to HPC System Administrator training, including limits to the class size, limits to available compute resources, and the disruption to typical work duties to allocate time to participate during work hours. The HPC Ecosystems Project [7]has endeavored to mitigate these challenges by following lessons learned from MOOC methodology on developing HPC Training [8], [9] . Through the development of a reproducible, offline-capable, self-paced HPC Virtual Training Lab that emulates a basic 3-node compute cluster on a trainee's local machine, we have created an HPC training lab that does not need any high-end computing resources or cloud infrastructure and can be accessed on-demand on the participant's preferred schedule.

CCS CONCEPTS

•Applied computing~Education~E-learning•Human-centered computing~Accessibility~Accessibility theory, concepts and paradigms•Social and professional topics~Professional topics~Computing education~Adult education

KEYWORDS

SC Proceedings, High Performance Computing, NICIS, CHPC, HPC Ecosystems Project, OpenHPC, Workforce development

ACM Reference format:

Bryan Johnston, Lara Timm and Mmabatho Hashatsi. 2023. Delivering Digital Skills across the Digital Divide: Creating an accessible online/offline self-paced HPC Virtual Training Lab. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (SC23). ACM, Denver, CO, USA, 2 pages.*

1 INTRODUCTION

1.1 Digital Skills and the Digital Divide

The HPC Ecosystems Project was established in South Africa in 2016 to address the national and regional need for localized advanced research computing resources [7]. Its mandate is to facilitate the deployment of HPC systems in institutions throughout Africa [5] and to equip personnel with the digital skills required to operate and maintain these systems. The development of hands-on technical skills in implementing an HPC system and administering the HPC system software stack are considered as essential components of Digital Skills in this regard.

Notwithstanding the large appetite for advanced research computing, many African institutions lack the available resources and infrastructure to operate an HPC system. Furthermore, the teams expected to one day deploy and maintain these HPC systems lack the required digital skills and have no means of accessing resources to upskill themselves. These factors have led to a notable digital divide in the field of HPC for Africa.

1.2 The Method

As of June 2023, the HPC Ecosystems Community consists of 37 officially supported sites in Africa and many more prospective sites, with several hundred members. There are seven HPC practitioners tasked with supporting these sites and personnel. Owing to the scale and scope of the community needs, following the traditional HPC training model of conducting onsite training to all the partner institutions is no longer feasible. To address the growing community's demands for training and the lack of available HPC infrastructure traditionally necessary to conduct the training, the HPC Ecosystems Project team has developed an on-demand virtual HPC Training Lab. The development of the Lab took into account best practices in the implementation of educational technology and followed lessons learned in this area in an effort to provide a robust and effective learning platform for HPC training [10]–[15].

Another contributor to the digital divide is the lack of access to reliable high-speed internet to take advantage of any remotely available training resources. The Virtual HPC Lab has been designed to allow the entire Lab to be run using either the live online content or in an offline mode with the downloaded source material or using a pre-packaged USB flashdrive.

1.3 Choosing OpenHPC

OpenHPC is a widely used HPC software stack with an active and growing community of users and contributors [5], [16]–[19]. To take advantage of the wide community base and to promote utility beyond the African HPC community, the Virtual HPC Lab uses OpenHPC as its base software stack.

2 THE TRAINING MODEL

2.1 The Traditional Model

Prior to the Virtual HPC Lab, all HPC Ecosystems Technical training was conducted on-site at partner institutions by a team of two to three trainers. The training would take place over the course of one week and all participants would be required to attend full-time, taking them away from their usual work duties during the training. Typically, this would mean, if not all, a significant portion of the site's Computer Support team would be unavailable during the training. Due to the scale of demand from the growing HPC Ecosystems Community and the limited impact on a small team of five or six trainees per site training workshop, the on-site training became unsustainable, leading to the adoption of virtual training workshops to improve accessibility and optimize the impact of a single in-person training workshop by facilitating multiple sites in one virtual workshop.

Traditional virtual HPC training workshops rely on cloud-based resources to simulate an HPC environment and these require a persistent network connection. In Africa, most institutions are prone to connectivity or resource provisioning issues. Additional delays to the training can occur from relatively trivial issues such as cloud credential problems or network latency. Furthermore, once the training workshop is concluded, the cloud-based infrastructure is destroyed, and participants are unable to return to their virtual training cluster.

2.2 The HPC Ecosystems Training Model

The Virtual HPC Lab developed by the HPC Ecosystems Project differs from traditional virtual training workshops in six distinct ways:

1. The Virtual HPC components do not use a cloud, they are deployed using Oracle VirtualBox onto the trainee's local machine, with the 3-node Virtual HPC cluster being configured to operate on any device with at least 4GB of RAM. The configuration is completely automated and reproducible to ensure consistency.
2. Content is available in two formats - video tutorials and as an interactive guide, available through GitLab.
3. The training can be performed entirely offline (all videos are downloadable to any device, the interactive guide can be exported to PDF, and the GitLab repository can be freely cloned).
4. The training can be performed self-paced, with no threat of computing resources being decommissioned or expiring, as is common with cloud resources.
5. The Virtual HPC cluster remains on the participant's machine and is available for revision or further development until the participant chooses to destroy it.
6. The end-product Virtual HPC Management Node can be easily modified to manage physical infrastructure.

2.3 The OpenHPC Training Model

While OpenHPC is offered as a deployment guide, the material assumes that the reader has established knowledge of HPC systems, which makes the material difficult for newcomers to consume. At the time of development of the Virtual HPC Lab, the only available formal OpenHPC training, except for site-specific internships, were offered as half-day tutorials at HPC conferences such as PEARC[20]. In summary, there was no meaningful resource for OpenHPC training besides participating in a tutorial at one of the mainstream conferences – an impossible and unaffordable task for nearly every member of the African HPC community, confirming the value in pioneering the first public on-demand Virtual OpenHPC training lab.

3 RESULTS AND IMPACT

Since publishing the Virtual HPC Lab in October 2020, there have been 226 participants trained in six formal online workshops hosted to date, with more than 5,500 views of the online training videos. Notably, the Virtual HPC Lab has attracted participants from outside of Africa, attesting to the global relevance of the first and only on-demand virtual OpenHPC Training Lab.

4 FUTURE WORK

An updated version of the Virtual HPC Lab (OpenHPC 2.x) is ready to launch in 2023Q3. Further work is underway to develop additional HPC modules that can be treated as standalone or 'bolt-on' training courses to the foundational virtual 3-node cluster that is deployed in the Virtual HPC Lab. Future HPC modules will serve as standalone training labs on various HPC Software Stack tools, including OpenOnDemand [21].

CONCLUSION

The move to virtual content delivery for HPC System Administrator training has enabled the HPC Ecosystems Project to reach a wider and larger audience of trainees in a fraction of the usual time. We have observed a significant number of participants (including outside of Africa) not only participate in the online training but also successfully deploy virtual HPC systems. By offering a virtual training lab for a globally adopted HPC software stack (OpenHPC) that is not only accessible on demand but also self-paced, we believe we have provided a meaningful mechanism towards bridging the digital divide to deliver HPC digital skills not only in Africa but in any Resource Constrained Environment.

ACKNOWLEDGMENTS

John Poole, Mopeli Khama, Anton Limbo, Rammolenyane Lethaha, Cyrus Kamau, Sakhile Masoka, Israel Tshililo, Matthew Cawood, Themba Hlatshwayo, Zintle Sanda, Zama Mtshali.

REFERENCES

[1] J. Kunkel *et al.*, "Towards an HPC Certification Program," *The Journal of Computational Science Education*, vol. 10, no. 1, pp. 88–89, Jan. 2019, doi: 10.22369/ISSN.2153-4136/10/1/14.

[2] M. R. Sancho, N. Alexandrova, and M. Gonzalez, "Addressing HPC skills shortages with parallel computing MOOC," *Proceedings of 2015 International Conference on Interactive Collaborative and Blended Learning, ICBL 2015*, pp. 86–93, Jan. 2016, doi: 10.1109/ICBL.2015.7387641.

[3] S. Sarajlic, N. Edirisinghe, Y. Wu, Y. Jiang, and G. Faroux, "Training-based Workforce Development in Advanced Computing for Research and Education (ACoRE)," 2017, doi: 10.1145/3093338.3104178.

[4] D. Akin *et al.*, "Linux Clusters Institute Workshops: Building the HPC and Research Computing Systems Professionals Workforce ACM Reference format," 2017, doi: 10.1145/3155105.3155108.

[5] "View of Namibia's first high performance computer." https://www.sacj.org.za/index.php/sacj/article/view/1189/472 (accessed Aug. 05, 2023).

[6] S. Harrell, A. Younts, and S. L. Harrell, "Teaching HPC Systems Administrators," 2020, doi: 10.22369/issn.2153-4136/11/1/16.

[7] B. Johnston, "HPC Ecosystems Project: Facilitating advanced research computing in Africa," in *ACM International Conference Proceeding Series*, 2019. doi: 10.1145/3332186.3333264.

[8] M. Guzdial, "Limitations of MOOCs for Computing Education- Addressing our needs," *Ubiquity*, vol. 2014, no. July, pp. 1–9, Jul. 2014, doi: 10.1145/2591683.

[9] J. Mullen, W. Filinger, L. Milechin, and D. Henty, "The Impact of MOOC Methodology on the Scalability, Accessibility and Development of HPC Education and Training," *The Journal of Computational Science Education*, vol. 10, no. 1, 2019, doi: 10.22369/issn.2153-4136/10/1/11.

[10] T. Jones and K. Cuthrell, "Computers in the Schools YouTube: Educational Potentials and Pitfalls," 2011, doi: 10.1080/07380569.2011.553149.

[11] "(PDF) Moving from andragogy to heutagogy: implications for VET." https://www.researchgate.net/publication/37357847_Moving_from_andragogy_to_heutagogy_implications_for_VET (accessed Aug. 05, 2023).

[12] O. Abiona, C. Onime, S. Cozzini, and S. Hailemariam, "Capacity building for HPC Infrastructure setup in Africa: The ICTP experience", Accessed: Aug. 05, 2023. [Online]. Available: https://www.infona.pl//resource/bwmeta1.element.ieee-art-000006107383

[13] D. Akin *et al.*, "Linux Clusters Institute Workshops: Building the HPC and Research Computing Systems Professionals Workforce ACM Reference format," 2017, doi: 10.1145/3155105.3155108.

[14] E. Shoop *et al.*, "Virtual Clusters for Parallel and Distributed Education," 2012, Accessed: Aug. 05, 2023. [Online]. Available: http://www.virtualbox.org/

[15] M. Cytowski *et al.*, "HPC Education and Training: an Australian Perspective Mark Gray," 2019, doi: 10.22369/issn.2153-4136/10/1/8.

[16] C. Langin, "From BigDog to BigDawg Transitioning an HPC Cluster for Sustainability," 2019, doi: 10.1145/3332186.3332188.

[17] C. Simmons, K. Schulz, and D. Simmel, "Customizing OpenHPC," p. 1, Jul. 2020, doi: 10.1145/3425306.3444765.

[18] S. Husodo, J. Chappell, V. Gazula, L. Pike, and J. Griffioen, "Slicing and Dicing OpenHPC Infrastructure: Virtual Clusters in OpenStack", doi: 10.1145/3332186.3332214.

[19] K. W. Schulz *et al.*, "Cluster Computing with OpenHPC," Nov. 2016, doi: 10.1145/1235.

[20] J. L. Wofford and C. Lueninghoener, "The Supercomputing Institute: A Systems-Focused Approach to HPC Training and Education," 2019, doi: 10.22369/issn.2153-4136/x/x/x.

[21] R. Hudak *et al.*, "Open OnDemand: A web-based client portal for HPC centers," vol. 43, pp. 1–43, 2013, doi: 10.21105/joss.00622.