Let's Get Our Heads Out of the Clouds

A scalable and sustainable approach to HPC Training Labs for Resource Constrained Environments and anyone else stuck in the Clouds

Bryan Johnston
Advanced Computer Engineering Lab
CSIR
bjohnston@csir.co.za

Lara Timm
Advanced Computer Engineering Lab
CSIR
ltimm@csir.co.za

Eugene de Beste
Advanced Computer Engineering Lab
CSIR
edebeste@csir.co.za

Mabatho Hashatsi
Advanced Computer Engineering Lab
CSIR
mhashatsi@csir.co.za

ABSTRACT

In this paper, we present an approach to hands-on High Performance Computing (HPC) System Administrator training that is not reliant on high performance computing infrastructure. We introduce a scalable, standalone virtual 3-node OpenHPC-based training lab designed for Resource Constrained Environments (RCE's) that runs on a participant's local computer. We describe the technical components and implementation of the virtual HPC training lab and address the principles and best practices considered throughout the design of the training material.

KEYWORDS

SC Proceedings, High Performance Computing, CSIR, NICIS, CHPC, HPC Ecosystems Project, OpenHPC, Workforce development, Virtual Training, System Administrator Training.

1 INTRODUCTION

The training of new and existing HPC practitioners is recognized as a priority in the global HPC community [23]. The *HPC Ecosystems Project* is an initiative of the South African Department of Science and Innovation to address this priority. This is achieved through the distribution of repurposed HPC resources in Africa and subsequent training of HPC System Administrators to manage these systems [18].

Within the global HPC community, the common approach to delivering HPC System Administrator training is through physical face-to-face engagements. The hands-on component of HPC training is facilitated through leveraging existing HPC infrastructure or cloud-based services that simulate an HPC environment [2, 7–10, 40].

During its early stages (2014-16), the HPC Ecosystems Project followed the same approach: face-to-face training engagements using HPC resources allocated from a production HPC system or cloud-based platform[1]. Over time, the HPC Ecosystems Project has migrated much of its training material to digital delivery platforms[2] and has favored the use of virtual resources, in the form of local Virtual Machines (VMs) for the hands-on training component. This adaptation was made in response to the challenges experienced with the traditional approach towards HPC System Administrator training; namely relating to the use of physical HPC infrastructure or remotely provided HPC resources. We prioritize the delivery of HPC System Administrator training that does not require a constant internet connection and aims to provide downloadable resources to facilitate offline learning.

This paper will describe our approach towards offering HPC System Administrator training that overcomes these challenges and presents a more sustainable and scalable solution.

We emphasize that this is a training lab intended to teach HPC System Administrators how to configure and build an OpenHPC cluster from scratch. There are numerous localized virtual OpenHPC cluster solutions available publicly, but these are pre-configured virtual clusters that are not suitable for learning how to deploy an OpenHPC cluster. Going forward, we will refer to any allocated HPC resource, whether it is a remotely accessible HPC or a cloud-hosted service, collectively as a 'hosted HPC resource'; since these resources offer the same hands-on HPC experience for System Administrator training.

2 MERITS OF VIRTUAL MACHINES AS A TEACHING RESOURCE

A Virtual Machine (VM) is a software-simulated computer environment that behaves like a physical computer system that accesses the host computer's resources from software called a hypervisor. Virtualization allows the creation and running of multiple VMs on a single physical computer system. The hosting computer system for VMs can be a single laptop computer or a cloud-based system. Since the revival in popularity of VMs in the mid-2000s owing to the ability to operate VMs on commodity hardware [32, 33], significant research has been done on evaluating the impact of VMs as teaching resources.

Staubitz et al. [38] evaluated the use of VMs for hands-on exercises in courses and concluded that properly managed VMs can provide many benefits for scalable and online teaching, including:

- Reducing support costs ("The amount of effort and time the course provider has to invest to deliver the exercise or to help

https://doi.org/10.22369/issn.2153-4136/15/1/13

[1] The ACE Lab's internal private OpenNebula (pre-2020) and OpenStack cloud (2020+)

[2] HPC Ecosystems GitLab Training Repository; HPC Ecosystems YouTube Training Videos; OpenHPC 101 GitLab Pages

users with the setup") by utilizing mainstream desktop hypervisors and automated VM deployment tools;

- Reducing setup costs by preparing a 'golden image' in advance with all tools and software tailored for the course coordinator's needs which can then easily be replicated on demand for participants. The course coordinator can control the behavior of the tools that are curated in the central VM image.
- Locally hosted VM's can reduce costs for the course provider since computation is performed on the participants' machines.

Staubitz et al. [38] found that VMs are an effective way to simulate a real environment at greater scale for training labs. Students are trained for real world applications because they experience and use a system as it would typically be used by a professional. The adoption of VMs as a teaching tool for hands-on technical training has long been accepted as an acceptable and effective tool.

3 HPC SYSTEM ADMINISTRATOR TRAINING

3.1 Classifying the HPC System Administrator Audience

For our case study, we considered HPC System Administrator Training to be classified into two target categories: for Resource Constrained Environment (*RCE*), or for traditional scientific computing environments (*non-RCE*). We adopted the metrics below to classify an RCE or non-RCE. Sites classified as an RCE are shown to experience significant challenges when following a traditional training approach. Within this classification, an RCE lacks at least two of the following:

- Access to advanced computing infrastructure that can be purposed for scientific computing workloads.
- Computing resources that are available on-premises or via remote access (such as institutional partnerships or remote hosting).
- Access to stable and performant internet and infrastructure.
- Technology service levels that are considered reliable and generally uninterrupted (such as network connectivity and power supply).

In alignment with the goals of the HPC Ecosystems Project, the designed training approach aims to overcome the challenges faced when providing HPC System Administrator training in RCEs, where a traditional approach is not always appropriate or effective.

3.2 Hosted HPC Resources for Training

Through our own experience in both delivering and participating in many international HPC training programs that facilitate training through hosted HPC resources, we have identified several underlying limitations:

- Training class size is still limited by the limited availability of compute resources;
- The training lab's existence is dependent on the timed availability of advanced computing infrastructure (of any measure);
- The teaching resources and practical compute environments are ephemeral;
- The transitory life of the training resources limits time flexibility for participants before the resources are inevitably destroyed and the learning opportunity is over.

In particular, the class capacity restrictions associated with limited compute resources reduce the effective scalability of the HPC System Administrator training. Likewise, the reliance on timed availability of HPC resources along with the ephemeral properties of these training resources limits accessibility and long-term impact. Where hosted HPC resources are integrated into the training, any failure with the cloud host, while rare, will impact the entire class of participants. Additionally, it is undesirable that a participant's hard-fought gains during the training are at risk of being destroyed on the remotely hosted environment before they have concluded working with the resources. There is generally no straightforward mechanism for participants to reference their original training environment, or even to progress further on their original training once the training period is officially concluded. Simply put, to address the need for scalable and sustainable HPC training, short-lived cloud labs are short-lived answers.

3.3 Training HPC System Administrators in Resource Constrained Environments

As the HPC Ecosystems Project expanded into more than thirty African partner sites, so the need for scaling out of the training into these partner countries expanded. Specifically, many of the international partner sites were receiving their first HPC deployments and required HPC System Administrator training [35]. The traditional model of provisioning HPC or cloud-hosted training resources was not a viable approach to facilitating HPC technical training, given that many partner institutions met the classification of an RCE.

- Sites did not have HPC or cloud resources to use for provisioning lab resources.
- Many sites lacked reliable internet connectivity to facilitate remote access to hosted HPC resources in South Africa.

Our virtual HPC Training Lab is designed to provide an effective scalable and sustainable HPC training platform with these constraints in mind.

3.4 Training HPC System Administrators in non-Resource Constrained Environments

It is reasonable to assume that non-RCE's enjoy a richer and more effective training experience because they have access to advanced computing infrastructure, be it on-premises or remotely; however, we have indicated several shortcomings to the learning experience that are explicitly related to hosted HPC resources, such as the ephemeral nature of the resources and the limits to the class size that are directly linked to the resource capacity of the remote resource. We note that non-RCE's are not shielded from these training shortfalls.

4 THE CASE FOR A TAKEAWAY LAB

There is a need for truly scalable and sustainable HPC training to meet the training priorities of the HPC community [2, 23]. Based on our experiences in aspiring to deliver sustainable and scalable HPC training through hosted HPC resources, we have learned that the answer is not going to lie exclusively "in the cloud." While our solution was originally developed for a specific HPC community within a specific geographic region, we believe there is global utility for the principles of an on-host virtual computing cluster to facilitate HPC training. Our involvement with international HPC workshops suggests that many trainees globally would benefit from an HPC lab resource that lives on indefinitely after a course is concluded. Specifically, it would be preferable for participants to have the freedom to choose whether they wish to keep their HPC resource to continue their learning experience after a training course is concluded.

To overcome the fundamental barriers to effective HPC training labs associated with the traditional hosted model, we have considered lessons learned from E-Learning and Massive Open Online Course (MOOC) methodologies to develop an effective HPC training lab alternative. The resultant training lab is a virtual HPC lab that is reproducible, self-paced and emulates a basic three-node computing cluster on a trainee's local machine with an indefinite lifespan and without the need for any high-end computing resources or cloud infrastructure.

5 A MODEL FOR HPC TRAINING

The HPC Ecosystems virtual HPC lab does not require high-end computing resources or cloud infrastructure and is available on demand in the participant's preferred schedule because it is all hosted on their personal computer.

The development of the virtual lab considered best practices in the implementation of educational technology and followed previous lessons learned to provide a robust and effective learning platform for HPC System Administrator training [1, 2, 8, 11, 19, 36, 38].

Our intention is not to present our virtual HPC lab as the definitive ensemble of technical tools for all future HPC System Administrator training. Rather, we wish to showcase the principles underpinning our approach to promote a new model for scalable HPC System Administrator training.

5.1 Online Delivery vs. Online Training

Digital delivery has enabled our training solution to become available, scalable, and globally accessible. Although the training material is hosted online, this is simply to facilitate delivery of the training on demand and for free. The hands-on virtual lab is, in principle, capable of being modified for offline distribution and operation[1] and is designed to mitigate needing special HPC systems or infrastructure for training [13, 26].

Research indicates that online content delivery can be as good as, if not better than, on-premises course participation [8, 21]. "Online students learned as much as students in the Traditional version" [20] and [21] observed "the online students actually achieved superior learning outcomes despite spending less time."

The virtual training lab is intended to be easily scaled up to a virtual classroom environment to allow for blended learning delivery where needed. Without proposing dramatic changes to the existing in-person content, simply including an online rendition of the existing presentations and workshops could potentially be as effective as the current on-premises approach but open accessibility to a wider audience [20–22].

Given that the existing traditional face-to-face delivery mechanisms are time constrained with defined start and stop time windows, the potential for improved learning in shorter time with online content would suggest online delivery leads to improved impact of teaching HPC System Administrator skills [21].

6 TECHNICAL COMPONENTS

Commensurate with accessibility and sustainability principles, our virtual HPC lab uses commonly available opensource resources and differs from traditional HPC workshops in several ways:

1. Content is available in two formats - video tutorials and an interactive guide, available through YouTube and GitLab respectively;
2. The lab uses the OpenHPC[2] software stack to create the HPC environment.
3. *There is no cloud* – the HPC components are provisioned by Hashicorp Vagrant[3] and deployed using Oracle VirtualBox to emulate a 3-node virtual cluster that can run on any computer with at least 8GB of RAM.[4] Vagrant ensures consistency and reproducibility of the lab VMs;
4. The training lab has the potential to be modified to operate entirely offline (all components are downloadable);
5. The training is self-paced and is available on demand, with no cut-off deadline for access to the HPC resources;
6. The training lab has fewer restrictions to infrastructure modifications
7. There is a freedom to fail – the localised lab can use snapshots to preserve development states;
8. The virtual cluster remains on the participant's computer until they choose to remove it;
9. The finished virtual management node can be tweaked to manage physical HPC infrastructure.

The virtual HPC lab is available to anyone wishing to configure a basic 3-node HPC system or to practice / learn the OpenHPC software stack. Since all resources are run on the trainees' local machines, there is no need to troubleshoot remote connectivity or to create cloud credentials. Significantly, since every HPC resource is localized to each participant, there is no limit to the number of participants able to partake in the training.

Perhaps the strongest arguments in favor of a remotely hosted HPC resource are the comparatively poor performance of a locally hosted VM and the requirement for participants to have access to local computing systems with sufficient memory and disk space to run the virtual HPC labs. We do not consider performance to be a reason for concern since we are focused on delivering an accessible learning experience rather than a performant HPC cluster. Participant feedback indicates 4GB RAM is sufficient to operate the HPC Lab – the VirtualBox VM's are collectively allocated 7GB of RAM, and the finished lab occupies 6GB of local disk space; we have not yet encountered a participant who has been unable to complete the virtual lab – if necessary, the virtual HPC cluster can be reduced to a 2-node cluster, which will allocate 4GB of RAM through VirtualBox.

6.1 Support Materials

A scalable training lab ensures that content is available to any and all numbers of participants. We took note that MOOCs provide a means for scalable education but require some attention to potential challenges in delivering successfully at scale [29]. We also

[1] We have prototyped an offline solution but it has not yet formally been released.

[2] OpenHPC

[3] Vagrant by HashiCorp (vagrantup.com)

[4] Each compute node consumes 3GB of RAM and the **smshost** consumes 1GB of RAM.

considered [31] insights into the use of a cloud-based virtual lab to "develop scalable, maintainable, and shareable contents that minimize technical hurdles while still exposing students to critical concepts in cluster computing". When delivering our training, we have selected delivery services that support scalable delivery and downloadable content for offline consumption.

6.1.1 Virtual Lab Guide
The primary lab guide instructs users on how to deploy 3-node virtual cluster using Vagrant, VirtualBox, and OpenHPC. In line with established education principles, the efficacy of training content is enhanced with additional documentation beyond the core deployment steps for the virtual HPC [3]. To this end, the guide includes general high-level principles and context-relevant supplementary information and elaborations of the HPC software stack components. We use GitLab to host the training guide and training material and deliver the guide using GitLab Pages.[1]

6.1.2 Virtual Lab Video Guide
The video content and the documentation are constructed with similar detail such that the media can be used independently or collectively, depending on the specific needs of the participant. While the primary objective is to deliver a virtual HPC lab, it is evident that the most effective delivery of the content will have accompanying reference documentation [3]. When considering an appropriate delivery method for the video content, we noted [30] observations that many of the conferencing technologies they tested had significant constraints on numbers of connections. Noting that YouTube offers free access, free hosting, and offline download capabilities, YouTube is a potential platform to deliver training [5]. In line with these findings we chose to host the video content on YouTube to offset any delivery constraints.

YouTube allows for videos to be downloaded for offline viewing and is designed to be compatible with most modern devices (phones, tablets, mobile devices, etc.). The option to provide for offline video and document downloads is vital for accessibility for African partners, especially where electricity and network reliability is uncertain. Additionally, YouTube caters for automated captions in multiple languages, broadening accessibility for non-English speaking participants [19, 27]. While we have employed YouTube videos to enhance the learning experience for the virtual lab, discussion around the specifics of the format and construction of the video content is beyond the scope of this case study.

6.2 OpenHPC
OpenHPC is the standard HPC software stack for the HPC Ecosystems Community and the selection of OpenHPC is not arbitrary – it is a popular HPC software stack due to its simplification of implementing traditional HPC software components[34]. It is widely used with an active and growing community of users and contributors [15, 24, 34, 35, 37].

6.3 No Cloud
The motivation for a solution that is not dependent on a hosted HPC environment has already been explored in previous sections; suffice to say, we sought a solution that would provide at least an equivalent technical experience to one delivered by a remote resource. We settled on a combination of VirtualBox managed by *Hashicorp Vagrant*[2] that emulates many of the benefits of a traditional hosted delivery platform while avoiding the problems related to remote / hosted resources indicated earlier. Since all resources are provisioned through automated deployment and all run on the trainees' local machines, there is no need to troubleshoot remote connectivity or to create cloud credentials, or to spend any time with configuring or provisioning remote compute resources. Since everything is local to each participant, there is no limit to the number of participants that can undertake the training in the same cohort. From a training perspective, any catastrophic-level failures in the localized virtual HPC resource will be limited to an individual participant and not an entire class.

The lab has been verified to work on host systems with at least 8GB of RAM (each compute node requires 3GB of RAM and the management node requires 1GB of RAM). Survey results from participants indicate that some have reported success with deploying the 3-node virtual cluster with 4GB host systems. The deployed virtual cluster occupies 6GB of local storage.

6.3.1 VirtualBox and Vagrant
When selecting a type-2 Hypervisor that can be installed on participants' local machines, VirtualBox was selected because it is a widely used Hypervisor with support for the primary computer Operating Systems (Linux, Windows, Mac). We were cognisant of three problems observed by Helsing et al. [12] when implementing a localized virtual lab with VirtualBox –

1. the lack of user skills;
2. the long latency in installing and configuring the environment;
3. and the lack of resources and material to help solve technology issues in the virtual lab.

Noting that the participants will have heterogeneous computer environments, we introduced Vagrant as a virtual machine environment management tool that ensures configuration parity to address the three problems raised by Helsing and Staubitz [12, 38] in addition to managing the unpredictability of heterogeneous environments. Vagrant offers a solution to the "configuration drift". Since it is free and supports all major platforms, Vagrant ensures parity for all virtual lab participants and significantly reduces the complexity of installing and configuring the virtual machine environment for the virtual lab [4, 38]. Staubitz et al. showed that Vagrant "reduces the friction for creation, distribution, setup and update of Virtual Machines" and "can reduce hosting costs, improve the user experience for learners compared to traditional virtualization software and thus reduce support efforts on the course provider's side."

6.4 Potential for Offline Delivery
The HPC lab's virtual machine environment is deployed and stored locally on the participant's computer and the virtual cluster nodes are fully accessible without any internet connection. In principle, with the necessary Linux repositories synchronized offline and the GitLab repository cloned to the local machine, the entire training lab can plausibly run without an internet connection.

6.5 Self-paced and On Demand
Training labs—either in-person or remote—have additional accessibility and availability considerations with respect to the cost and time factors associated with the training. Even where training

[1] GitLab Pages | GitLab

[2] Vagrant | HashiCorp Developer

may be offered for free, there might be a cost factor associated with traveling to the event, or the time cost of participating during work hours. HPC tutorials that allow remote participation may resolve the issue of travel costs, but these tutorials will still be governed by a time schedule (and in some cases, running across a different time zone). The need for on demand material is borne from the fact that people seeking training may not find any in a time zone suitable to them [29]. Accordingly, our lab material is available on demand and is time zone agnostic, with participants able to access the material at their convenience, and at no cost.

6.6 Fewer Restrictions

Since the virtual machine training environment is stored locally, the participants have the control to modify the VM infrastructure if desired, such as network interface parameters, RAM, CPU core count, etc. On machines with sufficient resources, additional virtual nodes can be added to the virtual cluster – none of this would be possible on a cloud-hosted HPC lab without an internet connection or without appropriate administrator permissions.

6.7 Freedom to Fail

A benefit of many virtual environments is the ability to perform snapshots—an automated point-in-time backup of the virtual infrastructure that typically captures the state of the virtual resources which can be restored at a later stage. Besides the fact that most remotely served HPC labs disable the snapshot feature, it would in any event be of limited benefit since the remote resources are ephemeral. Since the HPC Ecosystems virtual HPC lab is hosted on local machines, participants can secure their workload with snapshots whenever they wish, which encourages the principle of "freedom to fail" [25, 39].

6.8 Persistent Image

Beyond the benefits of the snapshot feature, the virtual HPC lab remains persistent on the local machine indefinitely. Participants can perform the training at their own pace, or to explore further upon completion of the official training. Notwithstanding the utility in having a persistent image for further exploration, we acknowledge that it is advisable to keep images up to date and we advise participants to regularly update their software stack.

6.9 Virtual to Physical

The transition from the virtual HPC cluster to a live physical system is trivial—participants can either start over and repeat the process on physical infrastructure or they can take the documented steps to link the completed virtual management node to physical resources, which is a relatively straightforward procedure. The ability to transition from the virtual lab to a production system with only slight modifications to the virtual cluster brings substantial value to the persistent image.

7 EXISTING ALTERNATIVES

In addition to tutorials offered at HPC conferences (such as ISC, SC, and PEARC), there are many documented training and learning opportunities for HPC System Administrators facilitated in the Northern Hemisphere. For example:

- The TACC Institute Series
- Linux Cluster Institute workshops
- PRACE Training Portal

By contrast, it is difficult to locate documented public HPC System Administrator training opportunities occurring on the African continent. Under these conditions, accessibility to face-to-face HPC System Administrator training is affected by geographical location. When training events may include an online component, the issue of time zone differences may also be a factor.

7.1 OpenHPC Training

The official OpenHPC guide states that it is intended for *experienced Linux System Administrators* [34], with current training for OpenHPC conducted as short hands-on workshops during conferences hosted primarily in the Northern Hemisphere. Based on the information presented in Table 1, there have been approximately **fifteen hours of hands-on tutorials** conducted by the OpenHPC leadership community to date, with each session usually lasting approximately three hours of dedicated hands-on training.

Some other limited training on OpenHPC is provided at research institutions through internships or workshops. For instance, Wofford et al. [40] describes a 10-week internship program hosted at Los Alamos National Laboratory that includes aspects of hands-on OpenHPC experience. Outside of the engagements listed and the limited training at research institutions, OpenHPC support is informally provided through the OpenHPC virtual group.

Table 1: List of Known Formal Training for OpenHPC

Event	Duration in hours	Year & Month	Description
SC20	3	2020-11	Tutorial
SC19	1	2019-11	Birds of a Feather
PEARC19	*3.5*	2019-07	Tutorial
ISC 2019	*1*	2019-06	Birds of a Feather
HPCKP'19	0.5	2019-06	Presentation
SC18	1 unknown	2018-11	Birds of a Feather Tutorial
DAAC 2018	unknown	2018-11	Presentation
Open Source Summit 2018	3	2018-08	Tutorial
MVAPICH'18	unknown	2018-08	Presentation
DevConf.CZ 2018	0.25	2018-01	Presentation
SC17	1	2017-11	Birds of a Feather
MVAPICH'17	unknown	2017-08	Presentation
PEARC 17	3.5	2017-07	Tutorial
ISC 2017	1	2017-06	Birds of a Feather
HPCKP 2017	0.5	2017-06	Presentation
SC16	1	2016-11	Birds of a Feather
MVAPICH 2016	unknown	2016-08	Presentation
ISC 2016	1	2016-06	Birds of a Feather
FOSDEM 2016	0.45	2016-01	Presentation

8 RESULTS AND IMPACT

Since the launch of the OpenHPC1.3.x Virtual 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 accompanying online training videos. Notably, the Virtual HPC

Lab has attracted participants from outside of Africa, attesting to the global relevance as on demand virtual OpenHPC Training Lab.

8.1 Evaluation & Assessment

Regular participant assessment is identified as a key component to enhancing online learning efficacy [29]. Similar HPC Training workshops collect "attendee evaluations for each workshop session to ensure topical relevance, assess the instructors, and support a continual improvement process for the instructional materials." [2] or daily 'sticky note' surveys and end-of-term longer, formal surveys [40] To measure the effectiveness of the Training Lab, both quantitative and qualitative evaluation must be measured.

8.1.1 Quantitative Feedback

Quantitative evaluation of the HPC Ecosystems Virtual HPC Lab can be measured by:

- The number of Virtual Clusters successfully activated on the software layer monitoring agent that is associated with the Virtual Lab;
- The number of engagements with GitLab, YouTube, and the OpenHPC Virtual Lab guide.

Although the listed measures provide definitive *values*, the metrics themselves may not be meaningful without additional information, such as the number of *unique* users attempting the training material, or the number of attempts (and revisits) to the training material by a single participant over time. To track each participant of the training, a registration form is incorporated for the formal online training events, but this does not include measures for informal ad hoc consumption of the material. We intend to incorporate additional quantitative metrics to measure time-to-completion as well as the success rate of the virtual cluster deployments.

8.1.2 Qualitative Feedback

Participants are encouraged to provide regular feedback to enhance the quality of the content being delivered in future iterations (and possibly to allow for quick-response adjustments to issues with the course).

8.2 Risks

When the HPC System Administrator training is not conducted face-to-face, it will be prone to the common risk factors associated with exclusively online courses, such as high drop-out rates, isolation in an online course, lack of proximity to physical hardware, and lack of dedicated technical support for troubleshooting [6, 16, 17, 20, 25, 26, 29]. Findings by Michinov et al. [28] indicate that there is a correlation between lack of participation in discussion forums and low performance in online learning environments; to see an improvement, participants should be encouraged to participate in online discussions.

With both online and virtual learning, participants do not have the opportunity to see, touch or interact with actual hardware components [26]. Cahill et al. [6] advise providing additional instructional resources along with the online content to reduce limitations of online-based training.

Further risks relate to our specific approach to the digital delivery of the training material; the training model requires for the resources to be installed on a local machine:

- Participants are expected to have computers with adequate specifications;
- An internet connection is necessary to connect to the digital delivery;
- Where internet connection is not available, the coordination of shipping the modified offline virtual HPC lab potentially to international sites.

9 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 in a fraction of the usual time. In our formal training events, we have observed that the localised virtual HPC lab enables many participants that do not complete the training during the allotted period to successfully deploy their virtual HPC systems later in their own time.

Cytowski et al. [8] observe that there is no universal learning solution – "various solutions and platforms need to be carefully selected for different groups of interest" – accordingly, much of what is asserted in this case study was initiated with the narrow project scope of HPC Ecosystems community members in mind. We acknowledge that hosted HPC resources are also helpful in facilitating HPC training and there is certainly merit in adopting a hosted platform for certain types of training or audiences.

In all considerations, in order to deliver a solution that meets the requirements of the HPC Ecosystems audience (who are all considered RCE's), emphasis has been placed on scalability, sustainability, and self-sufficiency.

While cloud-computing resources can be used for virtual labs [2] these can prove costly and inaccessible for participants with internet connectivity constraints, such as those present in RCE's. Catering for offline interaction with a virtual lab not only expands accessibility but offers limitless scalability since each participant will be hosting their own virtual cluster infrastructure.

Through the development of a virtual training lab that does not rely on a remote cloud for HPC resources, we believe we have identified a sustainable and scalable solution for practical HPC training that reaches further than the HPC Ecosytems community: an accessible on-demand self-paced virtual HPC lab where the HPC resources remain available indefinitely on a user's local machine.

10 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 additional standalone courses or used as 'bolt-on' modules to the foundational virtual 3-node cluster that is deployed in the Virtual HPC Lab.

Future planned OpenHPC modules include OpenOnDemand [14]. We currently have a group of supporters and contributors from numerous countries who are offering time and resources towards developing our future content and we always welcome more!

ACKNOWLEDGMENTS

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

REFERENCES

[1] Olatunde Abiona, Clement Onime, Stefano Cozzini, and Sebsibe Hailemariam. 2011. Capacity building for HPC Infrastructure setup in Africa: The ICTP experience. In *2011*

IST-Africa Conference Proceedings, May 2011, Gaborone, Botswana, 1-8.
[2] David Akin, Fang Cherry Liu, Amiya Kumar Maji, Henry Neeman, Resa Reynolds, Andrew Sherman, Michael Showerman, Jenett Tillotson, John Towns, George William Turner, et al. 2017. Linux clusters institute workshops: Building the HPC and research computing systems professionals workforce, In *EventHPC Systems Professionals Workshop, Proceedings of HPCSYSPROS 2017*, Denver Colorado, November 2017.
[3] Maha Bali. 2014. MOOC pedagogy: Gleaning good practice from existing MOOCs. *MERLOT Journal of Online Learning and Teaching 10*, 1.
[4] Abhishek Bhattacharya, Shagun Paul, and Arunima Jaiswal. 2014. Operating systems: Basic concepts and challenges against virtualization. *International Journal of Computer Applications 102*, 4, 975–8887.
[5] Sloane C. Burke, Shonna Snyder, and Robin C. Rager. 2009. An assessment of faculty usage of YouTube as a teaching resource. *Internet Journal of Allied Health Sciences and Practice 7*, 1, 8.
[6] Katharine J. Cahill, Scott Lathrop, and Steven Gordon. 2017. Building a community of practice to prepare the HPC workforce. *Procedia Computer Science 108*, 2131-2140.
[7] Carolyn Connor, Amanda Bonnie, Gary Grider, and Andree Jacobson. 2017. Next generation HPC workforce development: The computer system, cluster, and networking summer institute. *In Proceedings of EduHPC 2016: Workshop on Education for High-Performance Computing,* Salt Lake City Utah, November 2016, 32-39.
[8] Maciej Cytowski, Christopher Harris, Luke Edwards, Karina Nunez Mark Gray, and Aditi Subramanya. 2019. HPC education and training: An Australian perspective. *Journal of Computational Science Education 10*, 1, 48-52.
[9] Alex Younts and Stephen Lien Harrell. 2020. Teaching HPC systems administrators. *Journal of Computational Science Education 11*, 1, 100-105.
[10] Stephen Lien Harrell, Hai AhNam, Veronica Vergara Larrea, Kurt Keville, and Dan Kamalic. Student cluster competition: A multi-disciplinary undergraduate HPC educational tool. In *EduHPC '15: Proceedings of the Workshop on Education for High-Performance Computing*, Austin Texas, November 2015, 1-8.
[11] Stewart Hase and Chris Kenyon. 2001. Moving from andragogy to heutagogy: implications for VET. In *Proceedings of Research to Reality: Putting VET Research to Work: Australian Vocational Education and training Research Association* (AVETRA), Adelaide SA, March 2001.
[12] Joseph Helsing, Paulette Lewis, and Edward Warga. 2013. Teaching tools, applications, and infrastructure for digital curation through the use of a virtual lab. In *iConference 2013 Proceedings*, University of North Texas, February 2013. 781–784.
[13] Violeta Holmes and Ibad Kureshi. 2015. Developing high performance computing resources for teaching cluster and grid computing courses. *Procedia Computer Science 51*, 1, 1714–1723.
[14] Dave Hudak, Doug Johnson, Alan Chalker, Jeremy Nicklas, Eric Franz, Trey Dockendorf, and Brian L. McMichael. 2018. Open OnDemand: A web-based client portal for HPC centers. *The Journal of Open Source Software* 43, 1–43.
[15] Satrio Husodo, Jacob Chappell, Vikram Gazula, Lowell Pike, and James Griffioen. 2019. Slicing and dicing OpenHPC infrastructure: Virtual clusters in OpenStack. In In *PEARC '19: Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines, Chicago Illinois,* July 2019, 1-6.
[16] Alexandru Iosup and Dick Epema. 2014. An experience report on using gamification in technical higher education. In *SIGCSE '14: Proceedings of the 45th ACM technical symposium on Computer science education,* Atlanta Georgia, March 2014, 27–32.
[17] Evelyn S. Johnson, Michael J. Humphrey, and Keith W. Allred. 2017. Online learning and mentors: Addressing the shortage of rural special educators through technology and collaboration. *Rural Special Education Quarterly 28*, 2,17–21.
[18] Bryan Johnston. 2019. HPC Ecosystems Project: Facilitating advanced research computing in Africa. In *PEARC '19: Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines, Chicago Illinois,* July 2019, 1-3.
[19] Troy Jones and Kristen Cuthrell. 2011. YouTube: Educational potentials and pitfalls. *Computers in the Schools 28*, 1.
[20] David Joyner. 2018. Toward CS1 at scale: Building and testing a MOOC-for-credit candidate. In *L@S '18: Proceedings of the Fifth Annual ACM Conference on Learning at Scale,* London, June 2018, 1-10.
[21] David Joyner and Melinda Mcdaniel. 2019. Replicating and unraveling performance and behavioral differences between an online and a traditional CS course. In *CompEd '19: Proceedings of the ACM Conference on Global Computing Education*, Chengdu China, May 2019, 157–163.
[22] David A. Joyner, Charles Isbell, Thad Starner, and Ashok Goel. 2019. Five years of graduate CS education online and at scale. In *CompEd '19: Proceedings of the ACM Conference on Global Computing Education,* Chengdu China, May 2019, 16–22. DOI:https://doi.org/10.1145/3300115.3309534.
[23] Julian Kunkel, Kai Himstedt, Nathanael Hübbe, Hinnerk Stüben, Sandra Schröder, Michael Kuhn, Matthias Riebisch, Stephan Olbrich, Thomas Ludwig, Weronika Filinger, Jean-Thomas Acquaviva, Anja Gerbes, and Lev Lafayette. 2019. Towards an HPC certification program. *The Journal of Computational Science Education 10*, 1, 88–89.
[24] Chet Langin. 2019. From BigDog to BigDawg transitioning an HPC cluster for sustainability. In *PEARC '19: Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines*, Chicago Illinois, August 2019.
[25] Joey Lee and Jessica Hammer. 2011. Gamification in Education: What, How, Why Bother? *Academic Exchange Quarterly 15,* 2, 146.
[26] Pedro López and Elviro Baydal. 2017. On a course on computer cluster configuration and administration. *Journal of Parallel and Distributed Computing 105*, 127–137.
[27] Justin R. Louder, Suzanne Tapp, Larry Phillippe, and Jackie Luft. 2016. Video captioning software to help comply with

ADA accessibility requirements. *Community College Enterprise 22*, 1, 71–76.

[28] Nicolas Michinov, Spohie Brunot, Olivier Le Bohec, Jacques Juhel, and Marine Delaval. 2011. Procrastination, participation, and performance in online learning environments. *Computers & Educatyion 56,* 1, 243-252.

[29] Julia Mullen, Weronika Filinger, Lauren Milechin, and David Henty. 2019. The impact of MOOC methodology on the scalability, accessibility and development of HPC education and training. *The Journal of Computational Science Education*. 10, 1, 67-73. DOI:https://doi.org/10.22369/issn.2153-4136/10/1/11.

[30] Henry Neeman, Horst Severini, Dee Wu, and Katherine Kantardjieff. 2010. Teaching high performance computing via videoconferencing. *ACM Inroads*. 1, 167–71.

[31] Linh B. Ngo and Jeff Denton. 2019. Using CloudLab as a scalable platform for teaching luster computing. *Journal of Computational Science Education 10,* 1, 100-106.

[32] Mendel Rosenblum. 2004. The reincarnation of virtual machines: Virtualization makes a comeback. *Queue 2*, 5, 32-40.

[33] Mendel Rosenblum and Tal Garfinkel. 2005. Virtual machine monitors: Current technology and future trends. *Computer 38*, 5, 39–47. DOI:https://doi.org/10.1109/MC.2005.176.

[34] Karl W. Schulz, C. Reese Baird, David Brayford, Yiannis Georgiou, Gregory M. Kurtzer, Derek K. Simmel, Thomas L. Sterling, Nirmala Sundararajan, and Eric Van Hensbergen. 2016. Cluster computing with OpenHPC. Retrieved from https://api.semanticscholar.org/CorpusID:63310875.

[35] Jimmy Shapopi, Anton Limbo, and Michael Backes. 2023. Namibia's first high performance computer. *South African Computer Journal 35*, 1, 4–18. DOI:https://doi.org/10.18489/SACJ.V35I1.1189.

[36] Elizabeth Shoop, Malcolm Kane, Richard Brown, Devry Lin, Eric Biggers, and Maura Warner. 2012. Virtual clusters for parallel and distributed education. In *SIGCSE '12: Proceedings of the 43rd ACM Technical Symposium on Computer Science Education*, Raleigh North Carolina, March 2012, 517-522.

[37] Christopher Simmons, Karl Schulz,. and Derek Simmel. 2020. Customizing OpenHPC. In *PEARC '20 Companion: Proceedings of the Conference on Practice and Experience in Advanced Research Computing*, Portland Oregon, July 2020.

[38] Thomas Staubitz, Maximilian Brehm, Johannes Jasper, Thomas Werkmeister, Ralf Teusner, Christian Willems, Jan Renz, and Christoph Meinel. 2016. Vagrant virtual machines for hands-on exercises in massive open online courses. *Smart Innovation, Systems and Technologies*. 59, 363–373.

[39] Andrew Stott and Carman Neustaedter. 2013.Analysis of gamification in education. Retrieved from http://clab.iat.sfu.ca/pubs/Stott-Gamification.pdf.

[40] J. Lowell Wofford and Cory Lueninghoener. 2019. The Supercomputing Institute: A systems-focused approach to HPC training and education. *Journal of Computational Science and Education 11,* 1, 73-80.