TAMU HPRC Portal: Leveraging Open OnDemand for Research and Education

Ping Luo
High Performance Research Computing
Texas A&M University
College Station, TX, USA
pingluo@tamu.edu

Donald McMullen
High Performance Research Computing
Texas A&M University
College Station, TX, USA
mcmullen@tamu.edu

T. Mark Huang
High Performance Research Computing
Texas A&M University
College Station, TX, USA
tmarkhuang@tamu.edu

Shaowen Mao
Department of Petroleum Engineering
Texas A&M University
College Station, TX, USA
shaowen.mao@tamu.edu

Michael Dickens
High Performance Research Computing
Texas A&M University
College Station, TX, USA
cmdickens@tamu.edu

Marinus Pennings
High Performance Research Computing
Texas A&M University
College Station, TX, USA
pennings@tamu.edu

Yang Liu
High Performance Research Computing
Texas A&M University
College Station, TX, USA
yangliu@tamu.edu

Abstract—**The Texas A&M University High Performance Research Computing (TAMU HPRC) Portal is a local installation and adaptation of Open OnDemand (OOD) on the HPRC clusters. The Portal provides an advanced cyberinfrastructure that enables HPRC users to utilize the High Performance Computing (HPC) resources for their research. It also serves as an educational platform for faculty and staff for teaching and training.**

In this paper, the implementation of the TAMU HPRC Portal based on OOD is described, including customization of the Matlab OOD interface to meet our users' needs and implementation of a dynamic form generation scheme to simplify app deployment and management. To understand the impact of the Portal on our users, we analyzed Portal access data and conducted a survey among users. Most of the 148 responses acknowledge that the Portal is useful and majority of users would recommend it to other users. Two use cases are provided to further demonstrate how the Portal has enhanced research and training at TAMU HPRC.

Overall, the TAMU HPRC Portal has provided a robust and simple approach for both novice and experienced users to accessing HPC resources. It is a valuable addition to the traditional command line based approach.

Keywords—cyberinfrastructure, portal, Open OnDemand, HPC, survey, training

I. Introduction

Open OnDemand (OOD), developed by the Ohio Supercomputing Center (OSC), is an open source web-based client portal for HPC centers. It was released to public in March 2017. The project was funded by NSF and is now in its second stage with extended NSF support [1].

OOD provides applications (apps) for general system access and job management. It also provides a central location for hosting any number of server and GUI apps [2]. Using OOD for an HPC web portal has many benefits. First, it provides a single access point to all the HPC resources and apps on a web-based portal and can greatly simplify HPC workflows. Second, it provides an intuitive user interface that significantly reduces the barrier between users and HPC working environments. Third, the extensible and scalable design makes it easy to accommodate a growing number of users and apps.

OOD is the first general and comprehensive web portal for HPC centers. It differs from other existing science gateways, such as Galaxy [3], HUBZero [4], and Globus Connect [5]. Galaxy is a domain specific portal designed for researchers in the bioinformatics area to reuse and share their datasets and workflows with a simple user interface and readily accessible tools. HUBZero is a platform for the development of specific gateways for any science domains. Globus Online is an online portal for data transfers. Before we installed OOD, TAMU HPRC implemented its own remote visualization portal [6] based on GSI-OpenSSH. None of above mentioned portals can satisfy as many needs of HPC centers as OOD does.

OOD continues to draw increasing attention from academia and industry. TAMU HPRC is among the earliest adapters of OOD. We started evaluating OOD in June 2017 and collaborated with OOD developers to port the software on the HPRC Ada cluster. We assisted the OOD developers in identifying bugs, as well as provided them with feedback on the features of OOD. The TAMU HPRC Portal was opened to all HPRC users in March 2018 and has since been well received by our users. In February 2019, OOD was ported on the HPRC Terra cluster and integrated into the TAMU HPRC Portal.

In this paper, the implementation of the TAMU HPRC Portal based on OOD is described, including how Matlab is integrated into OOD, and how a dynamic form generation scheme is implemented to simplify app deployment and management. Statistics from the portal logs are presented to show how the portal has been used by HPRC users. Results from a user survey are provided to describe our users' experience with the portal. Our paper is one of the first that describes the experience with OOD from an HPC site outside of OSC. Our experience could assist other universities and organizations in helping them make informed decisions on adapting OOD.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. TAMU HPRC PORTAL

A. Setup and Improvement

The homepage of the TAMU HPRC Portal provides links to the Portal user guide and the OnDemand portal for each cluster: the Ada OnDemand Portal and the Terra OnDemand Portal, as shown in Fig. 1. The Ada OnDemand Portal is a local adaptation of OOD on a login node of the Ada cluster. Likewise, the Terra OnDemand Portal is a local adaptation of OOD on a login node of the Terra cluster. We configured the access to the TAMU HPRC Portal and the two cluster portals with Central Authentication Service (CAS) provided by TAMU IT. When a user visits the Portal homepage, the user will be redirected to the TAMU CAS page and prompted to enter their TAMU ID and password. Once authenticated, any user who has a valid account on our clusters will be able to log on to the Portal homepage, where the user can access each of the two cluster portals without additional authentication due to the single-sign-on feature of CAS. This feature is very convenient for users who want to use both cluster portals at the same time.

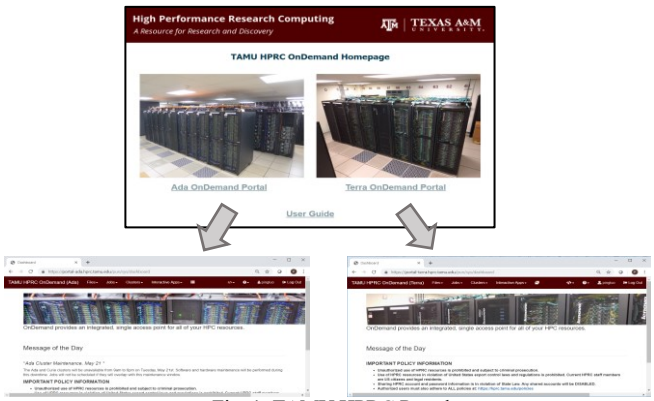

Fig. 1. TAMU HPRC Portal

Besides the out-of-the-box features provided by OOD, we have enhanced the Matlab app in our cluster portals. Unlike most apps, where the user needs to specify hardware resources such as cores and nodes, we used a different approach for the Matlab app where the user specifies Matlab specific resources such as computational threads and workers. We did this for two main reasons. First, many users might not be familiar with hardware resources. Second, we want to provide a consistent interface for our users with other Matlab resources we provide. In addition, Matlab uses automatic multithreading for many of its operators so we need to ensure the number of resources requested matches the number of resources Matlab actually uses.

As more and more apps are hosted in each cluster portal, and software modules on our clusters are frequently updated, it is necessary to automatically generate the software module list shown in the form for each app. To address this need, we designed and implemented a dynamic form generation scheme. In this scheme, we use form.yml.erb to replace the original form.yml file. Each app is able to start with the same form.yml.erb and change only the variable app_name to its own name. The embedded Ruby script reads an external file to generate the module list in the form for the app. The external file contains a list of modules for the app and will be generated periodically by automatically querying LMOD which is a software module management environment. The dynamic form generation scheme makes app deployment and management much easier.

B. Training and Outreach

The TAMU HPRC Portal provides a number of commonly used GUI and server apps, such as Abaqus, Ansys, LS-PrePost, Matlab, RStudio, and Jupyter. The Portal has been used in several workshops to teach Abaqus and LS-PrePost, as well as by the Microscopy and Imaging Center at TAMU for one of their workshops. We also used the Portal to teach a short course Spark for Big Data. The success of these workshops and short course stems from the ideal training environment provided by OOD, since each student can launch their own copy of a given software package in a web browser from a personal laptop or lab computer with a consistent work environment and set of computing resources.

However, there are some aspects that could be improved to facilitate training for a broader group of students. First, since the Portal is configured with CAS, it cannot be used by non-TAMU users who do not have a TAMU ID. At the same time, it requires an active HPRC account on the cluster and hence cannot be used by non-HPRC users who have no HPRC account. These two ID requirements prevent the Portal from being used for courses and workshops opened to a broader audience other than HPRC users only. If the Portal could support two different authentication methods at the same time, one for CAS and one for cluster users only, then we would be able to manage user access more gracefully. Second, the portal is tightly tied to the batch system on the cluster. Running any GUI and server app requires submitting a job to the batch system, which may be overkill for a training program, especially if the cluster is busy and the jobs are queued for an extended period of time. For the second problem, we usually reserve compute nodes in advance for the training program. While this works well, we hope OOD could provide a new functionality to launch the apps without the batch scheduler. This will eliminate the need to reserve nodes.

III. ACCESS STATISTICS AND ANALYSIS

We analyzed the user access log files generated between January 2018 and April 2019. The total number of users of the Portal has been steadily increasing over time to more than five hundreds as shown in Fig. 2(a). We define returning users as those who accessed the portal in at least two different days, while non-returning users as those who accessed the portal on only one day. Figure 2(b) shows the number of returning and non-returning users. There were 62.3% returning users who have accessed the Portal for their work on different days. The number of returning-users shows that the majority of Portal users consider the Portal to be useful. Figure 2(c) displays the classification of Portal users by colleges. The Portal users come from a diverse background. About 89% of Portal users are from six colleges: the College of Engineering, College of Agriculture and Life Science, College of Science, College of Geosciences, Veterinary Medicine and Biomedical Sciences, and TAMU-Galveston. The remaining 11% of Portal users are from College of Liberal Arts, College of Architecture, Mays Business School, Institute of Quantum Science, etc.

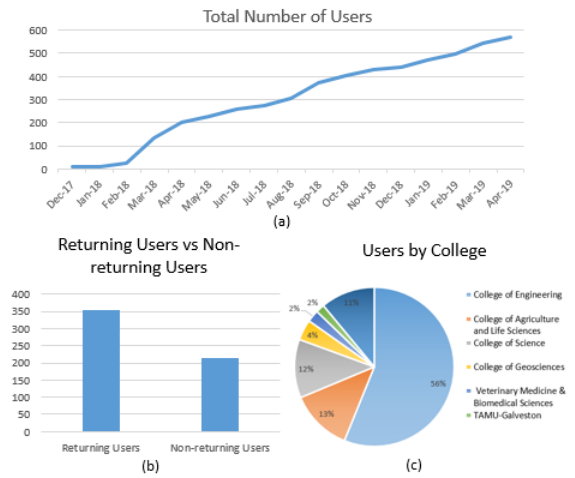

Fig. 2. User Statistics

We also collected the access data for the apps between January 2018 and April 2019. As shown in Fig. 3(a), the number of apps launched each month was relatively stable before December 2018, but increased every month since February 2019, most likely because (1) more users learned about the portal and started using it, and (2) repeat users became more familiar with the portal interface and used it more often. Fig. 3(b) shows the total number of accesses since January 2018 for the following apps: Files, File Editor, Shell Access, Active Jobs, Job Composer, as well as Interactive Apps. Job Composer allows users to construct and submit jobs in a browser, which seems to be an attractive functionality. But Fig. 3(b) shows Job Composer is the least popular app and has drawn only 362 launches. From our experience with our users, we think there are three reasons. First, Job Composer can be useful to new users, but is unlikely to attract users who are used to job submission from the command line. This is different from the other apps that simplify the traditional command line approach of doing the same thing, and hence are attractive to both new and experienced users. Second, Job Composer handles the job working directory differently from the traditional job submission at the command line. Job Composer submits jobs from a directory pre-determined by OOD, while the traditional job submission can occur from any directory specified by the user. To use Job Composer, the user must either prepare all input files in the pre-determined job working directory or modify the job script to include a command to change to the actual working directory where all things are setup before any other commands in the job script. If a job can be submitted from any directory through Job Composer, as we have suggested to the OOD team, it will make Job Composer easier to use. Third, the interface is not as intuitive as the other apps and requires several steps to accomplish job creation and submission. New users, who often have no concept of what a batch job is, will easily get lost.

The category of Interactive Apps in Fig. 3(b) includes all the web server apps and GUI apps, and had about half of the total launches of all apps. In Fig. 3(c), we further breakdown the access of Interactive Apps by interactive app names. The VNC app allows users to launch any GUI program on a remote desktop and thus was used the most. LS-PrePost was used the least due to its small user base.

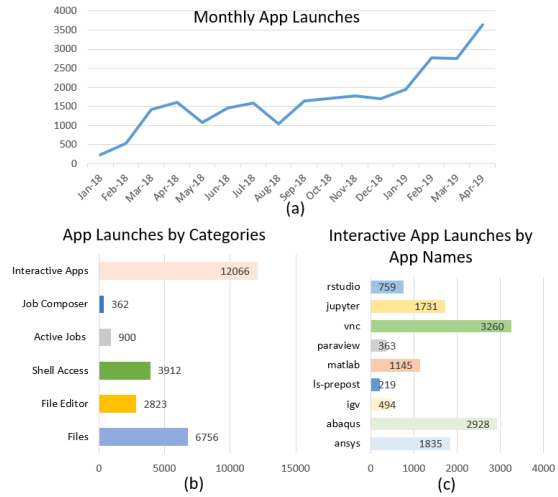

Fig. 3. App Statistics

IV. User Survey

We sent out a survey to 544 portal users out of 1730 total HPRC users, who had accessed the portal between March 22, 2018 and April 24, 2019, and received 148 responses. The respondents have a variety of backgrounds, as shown in Fig. 4(a). The pie chart shows respondents by colleges, which has similar patterns as Fig. 2(c). The similarity in the diverse backgrounds of the respondents indicates that the survey responses are representative of our Portal users.

Initially we assumed that OOD would attract more users who do not have strong Linux skills. Therefore our survey asked the respondents to describe their Linux skills by selecting one from the six answers: never used, very little, some, good, very good, and expert. The distribution of the responses is shown in Fig. 4(b) which reveals that the majority of the respondents (60.1%) have some or good Linux skills, since the Portal enables them to be more effective at using the clusters with its graphical interface. Only 16.9% of the respondents have little or no Linux skills. Our survey results disagreed with our initial assumption. This may be because users with fewer Linux skills do not have basic understanding of Linux and HPC and hence have difficulties to use the various OOD features. For example, the Job Composer requires users to understand concepts such as cores, wall time. This barrier can be removed by hands-on workshops. In fact 89% of the respondents suggested that hands-on workshops will help them become more familiar with the portal.

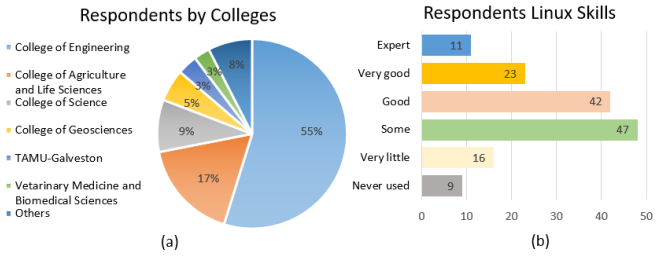

Fig 4. Background of the Respondents

The survey asked two additional questions in order to collect users' opinions about the Portal, which is our first goal of the survey. The first question is how the user evaluates the Portal

with five options: very good, good, neutral, not so good, and bad. Among all 148 responses, 82.4% are positive, including 56 very good, 66 good, 23 neutral, and only 2 not so good. The second question is how likely the user will recommend the Portal to other users, on a scale of 1 to 10. The average score for the question is 8.11 and the median is 8. These data show that most users acknowledge that the Portal is useful and will very likely recommend the Portal to others.

The second goal of the survey is to learn about our users' opinion of Job Composer. The responses on Job Composer include 38.5% no opinion (never used it), 15.5% extremely useful, 31.75% very useful, 11.5% somewhat useful, 2.7% not so useful, and 0% not at all useful. However, in an open question that asked the users whether they thought Job Composer had helped them in batch job submission, most of them said they had never used it, even though some of them said that Job Composer extremely useful or very useful. Such a contradiction probably indicates that Job Composer is a promising app provided by the Portal since it was considered to be useful by most users, but there is more room for improvement to encourage more users to use it.

The final goal of the survey is to ask users for suggestions on improving the Portal. Out a total of 148 respondents, 72 skipped the open question, and 25 answered no suggestions and/or the Portal was already pretty good. In summary, most current users are very satisfied with the Portal. There were still good suggestions for further improvements of the Portal. The top suggestion from 11 respondents requested for improved documentation, tutorials, videos, and hands-on workshops. Other suggestions included more advertisement of the Portal and adding features to improve specific apps, such as sorting of files in the File Explorer and improving the interface of Job Composer.

V. Use Cases

A. Investigating Proppant Distribution in a Complex Fracture Geometry

Due to pre-existing natural fractures and bedding layers in reservoirs, complex fracture networks are generally created upon multistage horizontal fracturing, to which a mixture of proppant-fluid is added. The overall objective of the research is to investigate proppant distribution in a complex fracture geometry that consists of primary fractures connected to constricting secondary fractures. An advanced computational model called Multiphase Particle-in-cell Method (MPPIC) will be employed to investigate how proppants transport in hydraulic fractures. The fluid and proppant interactions can be efficiently and fully considered. Proppant transport will be investigated in the field-scale fracture geometry and for field-scale injection time. MPPICFoam solver in OpenFoam is used to do the numerical simulation [7].

Since we want to simulate the proppant transport behavior for field scale fracture geometry and field scale injection time, the simulation requires a lot of computational resources. Terra with its strong computational power enables us to do the parametrical studies of proppant transport in the real field scale. By using the Terra portal, we can directly modify the files in the web browser. It saves us much time in modifying the files and setting up different cases. Also, the Job Composer improves our work efficiency to a large extent and Active Jobs allows us to monitor the job working conditions easily. When jobs are done, we can easily visualize the results using ParaView. The portal truly provides a one-stop shop for research in need of HPC.

B. Teaching Spark for Big Data

The Portal provides an ideal training environment for many purposes. One example of this is a short course we teach on getting started with the Spark big data package. A typical class of thirty students use OOD to launch a Jupyter Notebook with their own copies of exercises and sample data to learn Spark. Software version and Python packages are managed through the modules system behind the scenes to provide a reproducible learning environment which can be updated easily for all learners. OOD's in-built terminal emulator and file manager allow students to rapidly switch from a heavily scaffolded Jupiter Notebook environment to a command line environment for more advanced apps such as launching Spark clusters through the batch system.

VI. Conclusion and Future Work

The TAMU HPRC Portal based on OOD has provided a simple and robust method to both novice and experienced users at TAMU HPRC to access HPC resources and services. It is a valuable addition to the traditional command line approach. Future work includes enhancing the Job Composer interface to make it more user friendly, supporting two-way authentication, improving documentations, and hosting additional workshops.

Acknowledgment

We thank OOD developers Eric Franz and Jeremy Nicklas from OSC for helping us porting OOD on Ada and Terra. We also thank HPRC system manager Francis Dang for hardware support and HPRC director Honggao Liu for supporting this project.

References

[1] D. Hudak, D. Johnson, A. Chalker, J. Nicklas, E. Franz, T. Dockendorf, B. L. McMichael, "Open OnDemand: A web-based client portal for HPC centers," Journal of Open Source Software, 3(25), 622, May, 2018.

[2] D. Hudak, T. Bitterman, P. Carey, D. Johnson, E. Franz, S. Bradly, P. Diwan, "Open OnDemand: Transforming Computational Science Through Omnidisciplinary Software Cyberinfrastructur," In Proceedings of the Xsede16 Conference on Diversity, Big Data, and Science at Scale, 43:1–43:7. XSEDE16. New York, NY, USA: ACM.

[3] B. Giardine, C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah , Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent, A. Nekrutenko, "Galaxy: a platform for interactive large-scale genome analysis," *Genome Research*, 15(10):1451-5, October, 2005.

[4] M. McLennan, R. Kennell, "HUBzero: A Platform for Dissemination and Collaboration in Computational Scienceand Engineering," *Computing in Science and Engineering* 12(2), pp. 48-52, March/April, 2010

[5] I. Foster, "Globus Online: Accelerating and democratizing science through cloud-based services," *IEEE Internet Computing*(May/June):70-73, 2011

[6] P. Luo, D. Chakravorty, D. McMullen, F. Dang, "TAMU HPRC Remote Visualization Portal," poster presented at PEARC 2018.

[7] S. Mao, K. Wu, "An Efficient Three-dimensional Multiphase Particle-in-cell Model for Proppant Transport in the Field Scale," In: The Unconventional Resources Technology Conference, Denver, Colorado, USA, 2019