Towards Enabling Education as a Service on High Performance Computing Resource

Weijia Xu
Texas Advanced Computing Center
The Univeristy of Texas at Austin
Austin, TX
xwj@tacc.utexas.edu

Hui Zhang
Department of Computer Sciences
University of Louisville
Louisville, KY
hui.zhang@louisville.edu

Abstract— **As big data analysis and advanced machine learning algorithms drastically increasing computing resource requirement, high performance computing (HPC) resources are actively adopted by wide variety domain fields. How to improve the accessibility and support of educational tasks on HPC have become an emerging problem for supercomputing centers. Here, "*educational tasks on HPC*" refer to those educational activities to be carried out on the supercomputer. Examples include educational task on how to efficiently using HPC resources, teaching computational methods and data sets with significant computing resource requirement which cannot be met using a virtual machine or a single workstation. While HPC service providers have successfully supported diverse user communities with shared HPC resources for years, supporting educational tasks presents several unique challenges. In this lighting talk, we share our experiences and practices using computing resources at Texas Advanced Computing Center to support education on HPC. We introduce an in-house web application framework development, IDOLS, which can lower the access barrier and improve the ease of the support for education on HPC. We will present our approach and demonstrate how it can be used with three different use cases: to support training workshops on efficiently usage of HPC; to teach parallel R workflow in the classroom; and to support large scale audio data collection analysis on HPC.**

I. Motivations and Challanges

There is an increasing demand of supporting educational activities directly with high performance computing (HPC) resources in recent years. This is in concert with the increasing adoption of data-driven analytics fueled by massive amounts of complex data produced by businesses, scientific applications, government agencies and social applications across various disciplinary fields. Data-driven analytics have the potential to help users gain new insights for decision support, scientific discovery [1]. However, data-driven analytics often require computing resources only available from remote HPC resources. Hence there is an increasing demand of learning and utilizing remote HPC resources along with an increasing demand of using HPC to facilitate classroom education across many disciplinary fields.

Since 2018, the authors have organized over a dozen training workshops at Texas Advanced Computing Center (TACC) and/or in conjunction with leading conferences such as Supercomputing conferences (2018 - 2020) using computational resources at TACC. The topics of those training workshops ranges including scalable machine learning, big data analysis tools, and distributed deep learning tutorials. The enrollments of those training events have been steadily increasing over the year.

During the same period, the number of requests to use HPC resource at TACC in the classroom is also increased dramatically. Those requests are motivated to support a wide variety of topics, from traditional computing fields, such as teaching distributed methods in computer science and data science program, to domain sciences such as introducing large data set and latest methods in biology and digital humanities. Therefore, education and training on effectively using HPC resources become essential requirements disciplinary fields.

There are many unique challenges to directly support educational tasks with the HPC resources. We summarize three main challenges: resource provisions, credential management, and configurable content as follows.

Resource Provisions Challenge Cyberinfrastructures are usually designed as shared computing resources. Computing nodes are provisioned dynamically upon individual *job request* by users. Each *job request,* once submitted*,* will be placed in a queue and managed by a job management software. This resource provisioning model is not well suited to support educational activities where all students need access and use the system for a short period time. Since each *job request* is submitted independently by each student, there is an uncertainty on when one's job can run especially when available computing resources are limited. Due to different levels of familiarity, background and learning pace, it is often the case that some students will wait longer than others, which is disruptive to the pace of the entire class.

Credential Management Challenge As a shared infrastructure designed for many users, it is necessary to have strict and complex access control mechanisms. Taking the resource hosted at TACC as an example, the process not only requiring each user to set up an account prior to the class, but also requiring the instructor to associate student accounts with an appropriate project allocation. Complicating this process further is the adoption of multiple-factor authentication requirement. These security measures require student and instructor preparation well ahead the class time. Otherwise, it will take significant amount time to setup individual accounts during the class. In some scenarios, such as given an open tutorial at a conference, it is impractical to set up accounts for participants on the spot.

Content Customization Need Education activities often need to be tailored and customized each time. Even for the same topic and content, small updates may be needed each time. Some examples include, to update data sets used, to update exemplar code and exercises, to make adjustment based on classroom size etc. It is common that one seemingly small change may require additional updates. On the other hand, there are also similarities on the abstracted tasks needed for different educational activities. Hence there is a potential need to improve the

robustness and re-usability through a modularized customizable design support.

II. Approaches and Novelty

To address those challenges and to better support various education tasks, we propose the development and utilization of a user driven configurable web application framework, named IDOLS [1]. An overview of the system architecture of our framework and startup workflow is shown in Figure 1. Here *application owner* refers to the instructor who will start the web application service which is used to support various educational activities on the HPC cluster. A*pplication user* refers to students who will access and utilize the web application service started by the instructors. Hence, students will interact with clusters through the web application interface set up by the instructors rather than access the computing cluster directly.

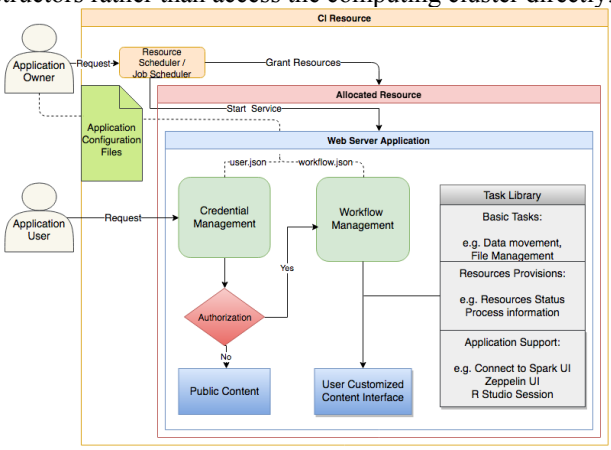

Figure-1 Overview of framework architecture and access workflow

There are four key components of this web application framework: credential management [2], application management, pre-defined task libraries and application configuration files. These components can address the resources provision challenge, credential management challenge, and customization needs described previously. On resources provision, instructors can request resources and start the web service application before the class. So, there is no need for students to request resources during the class. Since the web applications have its own dynamic credential management through which temporary credentials can be generated and assigned to the students on demand. The temporary credential can be linked to pre-generated training account on the cluster. The application configuration files are the central piece to enable further customization and updates. Most of the features and the web application content can be dynamically customized through configuration files [1]. For more details, please refer to [3].

III. Use Cases and Take Awyas

We have experimented with three use case scenariols: supporting training workshops with dynamic generated credentials; supporting of teaching parallel computing in the classroom; and supporting audio collection studies in digital humanities.

Supporting training workshops with dynamic generated credentials Many of our tutorial training sessions uses Jupyter Notebook to deliver examples and exercises to students. Preivously, we need hand out prepared training accounts to participants and ask student to start each Jupyter notebook session on their own. Each job will need go through the job queue and wait for resources to be ready. With the new approach, we can pre-start Jupyter Notebook sessions and assign one to each student upon request on demand. Both starting sessions, temporary credential generation and requests can be handled by pre-built tasks in IDOLS. This new approach can reduce the wait time for students and imporve roubstness.

Supporting of teaching parallel computing in the classroom Dr. Zhang has incorporate the framework with his class for teaching different parallel R models for undergraduate students at University of Louisville as well as a data science gateway for K-12 teachers to learn data science techniques and practice big data principles with R language. Most of the audiences are new to data analytics and are not yet equipped to take advantages of the advanced tools by HPC. Different parallel models are pre-built as modules for students to experiment. This approach can reduce the needs to introduce additional background knowledge on HPC environment and help students focusing on different parallel models [4].

Supporting audio collection studies in digital humanities This use cases concerns StoryCorps *Las Historias* audio collection. This audio collection could be used as a subject in different social science courses and studies. However, the collection is about 400GB makes it hard to distribute. We have built a web application interface using IDOLS to facilitate it access. In this model, the audio collection are stored on HPC, our web application serves as an interface for students and instructors to access. In addition to access and listen to the audio files, we also integrate with deep learning and natural language processing methods which can generate transcripts from audio and run meta-analysis for students. The interface can ease the access to these large data collections and also augmented with machine learning methods for social science students and researchers to interact [5].

Acknowledgment

This work has been supported by funding from National Science Foundation (Award# 1726816).

References

[1] Wang, Yige, Ruizhu Huang, and Weijia Xu. "Authentication with User Driven Web Application for Accessing Remote Resources." In *Proceedings of the Practice and Experience on Advanced Research Computing*, p. 2. ACM, 2018.

[2] Xu, Weijia, Ruizhu Huang, and Yige Wang. "Enabling User Driven Web Applications on Remote Computing Resource." In *2018 IEEE World Congress on Services (SERVICES)*, pp. 49-50. IEEE, 2018.

[3] Xu, Weijia, Ruizhu Huang, and Yige Wang. "Enabling User Driven Big Data Application on Remote Computing Resources." In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4276-4284. IEEE, 2018.

[4] R. Subramanian and H. Zhang, "Parallel R Computing on the Web," *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 3416-3423, doi: 10.1109/BigData47090.2019.9006311.

[5] W. Xu, M. Esteva, P. Cui, E. Castillo, K. Wang, H-R Hopkins, T. Clement, A. Choate, R. Huang (2020) "A Study of Spoken Audio Processing using Machine Learning for Libraries, Archives and Museums (LAM)," *2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 1939-1948, doi: 10.1109/BigData50022.2020.93784