The USD Science Gateway:
A Bridge Between Research and Advanced Computing

Adison A. Kleinsasser
University of South Dakota
414 E. Clark Street
Vermillion, SD 57069
Adison.Kleinsasser@usd.edu

Joseph D. Madison
Mayo Clinic
200 1st Street SW
Rochester, MN 55905
Madison.Joseph@mayo.edu

Eroma Abeysinghe
Science Gateways Research Center
Pervasive Technology Institute
Indiana University
Bloomington, Indiana
eabeysin@iu.edu

Sudhakar Pamidighantam
Science Gateways Research Center
Pervasive Technology Institute
Indiana University
Bloomington, Indiana
pamadigs@iu.edu

Marcus Christie
Science Gateways Research Center
Pervasive Technology Institute
Indiana University
Bloomington, Indiana
machrist@iu.edu

Suresh Marru
Science Gateways Research Center
Pervasive Technology Institute
Indiana University
Bloomington, Indiana
smarru@iu.edu

Douglas M. Jennewein
University of South Dakota
414 E. Clark Street
Vermillion, SD
57069
doug.jennewein@usd.edu

Marlon Pierce
Science Gateways Research Center
Pervasive Technology Institute
Indiana University
Bloomington, Indiana
mapierc@iu.edu

ABSTRACT Science Gateways are virtual environments that accelerate scientific discovery by enabling scientific communities to more easily and effectively utilize distributed computing and data resources. Successful Science Gateways provide access to sophisticated and powerful resources, while shielding their users from the underlying complexities. Here we present updated work completed by the University of South Dakota (USD) Research Computing Group in conjunction with the Science Gateways Community Institute (SGCI) [1] and Science Gateways Research Center at Indiana University to set up a Science Gateway to access USD's high-performance computing resources. We also introduce improvements to the system since the previous presentation of our work. These resources are now available to both faculty and students and allow ease of access and use of USD's distributed computing and data resources. The implementation of this gateway project has been multifaceted and has included placement of federated user login, user facilitation and outreach, and integration of USD's cyberinfrastructure resources. We present this project as an example for other research computing groups so that they may learn from our successes and the challenges that we have overcome

in providing this user resource. Additionally, this project serves to exemplify the importance of creating a broad user base of research computing infrastructure through the development of alternative user interfaces such as Science Gateways.

CSS CONCEPTS

Computing methodologies→Simulation evaluation;
•Software and its engineering→Software design engineering;

KEYWORDS

Gateway, Apache Airavata, Keycloak, SciGaP

1 INTRODUCTION

At the University of South Dakota, high performance computing (HPC) resources are available to faculty and students for use in their research. However, a lack of knowledge of command lines and HPC functionality may create a barrier that prevents many from utilizing the resource to its full potential. In order to reduce this barrier, the USD Research Computing Group (USDRCG), in partnership with the Science Gateways Community Institute, has

developed a science gateway- a website designed to facilitate the use of supercomputing resources from a graphical user interface (GUI) rather than a command line.

2 GATEWAY DESIGN

2.1 USD Gateway with SciGaP & Apache Airavata

The USD Science Gateway is hosted and maintained as a SciGaP [2] gateway. SciGaP is the hosted gateway platform that provides gateways as a service with Apache Airavata [3] middleware as the underlying framework. SciGaP services include, but are not limited to, hosting the Science Gateway, maintaining it, and providing storage for its user data. The USD Science Gateway is a web-based gateway with a graphical interface for HPC job submissions. The gateway has a flexible framework that enables development of customizable and accessible interfaces to the traditional command line applications prominent in HPC. This allows adaptability to the varying use cases that present themselves both at USD as well as to users at other universities in the state university network.

2.2 Gateway User Management

Gateway user authentication and authorization is maintained by Keycloak [4], https://www.keycloak.org/. The Science Gateway provides multiple methods for user access. USD users can use their university account to gain access to the gateway through federated access provided by CILogon, http://www.cilogon.org/, through Keycloak. Users from other universities and organizations can either use their organizational accounts or create a local account in the gateway. Once the gateway account is created, gateway administrators can decide the level of access to grant for each user (e.g. gateway admins, basic users or group leaders).

2.3 Gateway User-space

In the Science Gateway, users create 'experiments' in order to submit computational jobs to the USD supercomputer: Lawrence (Figure 1). The gateway supports SLURM job submissions, including monitoring jobs through completion. Users can submit experiments from the gateway and view real time status updates of the jobs as well as view the results through the gateway portal. Other features include downloading results, sharing results with other gateway users, cancelling running jobs on hosts, and cloning existing jobs.

The gateway also provides the ability to collect users into specific named groups. Group membership may be employed in a classroom setting to present a specific set of applications to students. Groups can also be used to manage access to software that is licensed for a limited number of people; by adding the users to a group and giving the group access to the software, the software can be utilized through the gateway, while fulfilling the requirements of the license.

2.4 Gateway Administration

The USD Science Gateway provides a dedicated set of web interfaces for gateway administrators in the form of an admin section, which provides interfaces required for managing gateway configurations and users. Admin features include interfaces to manage users, add software tools, maintain credential store keys for secure SSH communications [5], and add HPC configuration information. Job statistics are also available through the Admin Dashboard which includes gateway usage summaries, user specific software utilization, and status of HPC jobs and failures. These statistics are essential for a gateway administrator in order to best serve the gateway user community in resolving their issues and advising them with their HPC usage. Experiment statistics can be listed based on a date and time period provided by the administrator and further filtered by username, HPC resource, and software. With recent changes to the new system, administrators are better able to add new applications in a streamlined approach. The application setup is condensed into one page divided by tabs, (previously three pages) and contains more advanced customization options for application set up.

Figure 1. The USD Lawrence cluster supports general computational tasks as well as high-memory and GPU jobs.

2.5 Gateway Theme Updates

The Science Gateway uses a default gateway client provided by the SciGaP platform with localizations and other customizations provided by the USD team. Some of the updated customizations include logon options (USD users, other university users, and non-university users), a "Cite Us" page with streamlined citations and link to our previous abstract [6] (on which this abstract is based), and an "About" page.

Wagtail [7], a content management system, now manages the gateway login page as well as the supplemental pages. This gives USDRCG direct access to editing and maintaining these pages, as well as the ability to add more should the need arise.

2.6 Outreach

The USD gateway is currently focused on outreach events to establish the user base. Outreach is organized into two phases. The first phase focuses on recruiting users from academic and research groups within USD, and the second phase is focused on intrastate university collaborations and research centers. Upcoming conferences will be used as outreach events for introducing the gateway to the state. Planned and ongoing outreach events are in the form of workshops, tutorials, webinars and conferences along with help from SciGaP gateway team members.

3 APPLICATIONS

The USD Science Gateway enhances the research capabilities of the institutional cyberinfrastructure; this includes both USD and other institutions in South Dakota. Recent improvements to the Science Gateway enable USDRCG to add applications more efficiently and with a greater degree of customization.

3.1 Research

The purpose of the Science Gateway is to provide researchers with an easy-to-use method of utilizing software through an online graphical user interface. One available application is Mothur, an open-source project designed to analyze the rRNA sequences in microbial samples [8]. Ordinary interaction with Mothur is through a command line interface and requires, at minimum, filesystem navigation to a batch script and input files, followed by issuing a command to start the script. Without the gateway, a researcher using the USD Lawrence supercomputer must also know how to log in to Lawrence and be familiar with the Lawrence file system layout.

Alternatively, the Science Gateway's graphical interface provides a more accessible and approachable way to run computational jobs. In this interface, the user can access the Science Gateway's submission page (Figure 2) to upload the necessary data files along with a script containing the desired Mothur commands. The gateway uses the concept of an *experiment* to abstract away the concept of a computational job (which traditionally entails steps such as data staging and batch script editing), incorporating all necessary information for running an instance of an application into a single "New Experiment" web form. Once the job has been submitted, users can view it, and previous job submissions (Figure 3).
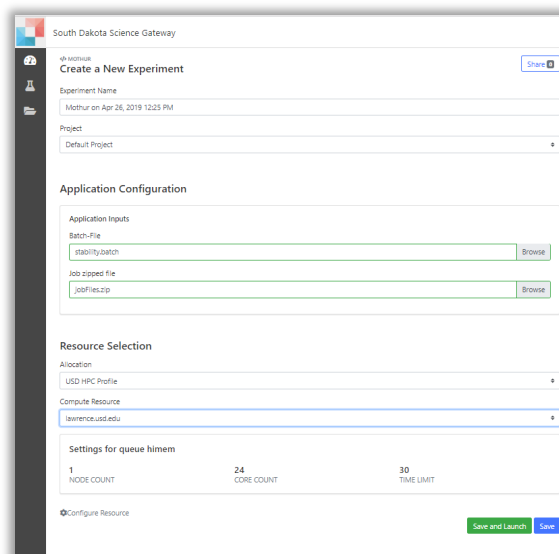

Figure 2: Users can submit a job via the "New Experiment" form.
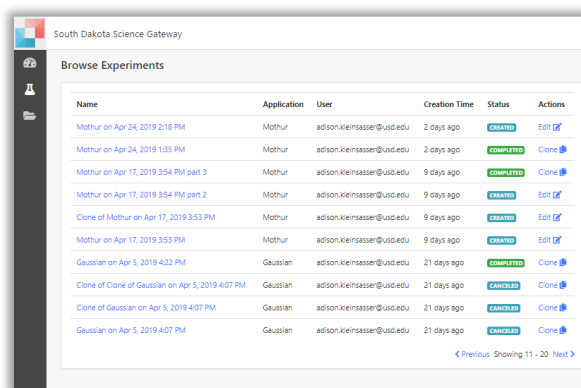

Figure 3. Job status reports for submitted job requests

When the computation is finished, the summary of the job contains links to download the output files, as well as the input files used (Figure 4).
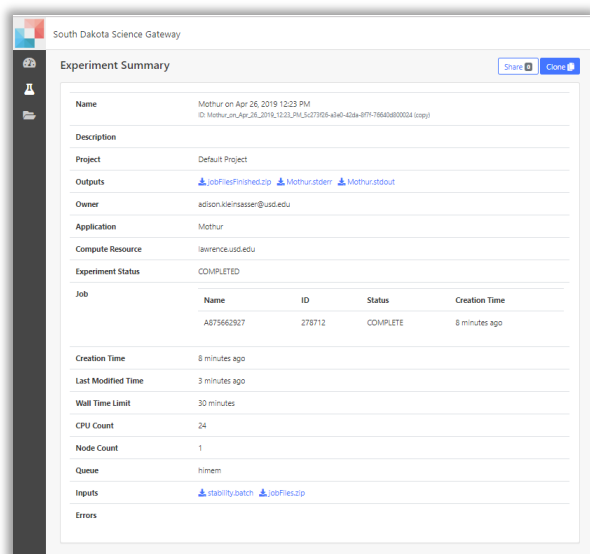
Figure 4. Users can download the job output data from the job summary.

3.2 Intrastate Institutional Access

The USDRCG Science Gateway has been setup to allow multi-institutional access. USD currently operates significant cyberinfrastructure resources in South Dakota, while many smaller state universities do not have these resources. Due to this limitation and USD's status as one of the main state providers of research cyberinfrastructure, access through the Science Gateway is being made available to all public and private universities, and their collaborators, in South Dakota.

4. FUTURE WORK AND CONCLUSIONS

To date, the USDRCG science gateway has been developed as a tool primarily for new users. Improvements and updates are ongoing, with additional applications expected to be added regularly.

Workshops and training are periodically held through the Research Computing Group. Integrating the Science Gateway into the workshops and training will provide users with exposure to the gateway and its capabilities, as well as to the computing resources available.

REFERENCES

[1] Lawrence, K.A., Zentner, M., Wilkins-Diehr, N., Wernert, J.A., Pierce, M., Marru, S. and Michael, S., 2015. Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community. Concurrency and Computation: Practice and Experience, 27(16), pp.4252-4268.

[2] Marlon Pierce, Suresh Marru, Mark A Miller, Amit Majumdar, and Borries Demeler. 2013. Science Gateway Operational Sustainability: Adopting a PlatformasaService Approach. (2013).

[3] Suresh Marru, Lahiru Gunathilake, Chathura Herath, PatanachaiTangchaisin, Marlon Pierce, Chris Mattmann, Raminder Singh, Thilina Gunarathne, Eran Chinthaka, Ross Gardler, et al. 2011. Apache airavata: a framework for distributed applications and computational workflows. In Proceedings of the 2011 ACM workshop on Gateway computing environments. ACM, 21–28.

[4] Christie, M. A., Bhandar, A., Nakandala, S., Marru, S., Abeysinghe, E., Pamidighantam, S., & Pierce, M. E. (2017). Using Keycloak for Gateway Authentication and Authorization.

[5] Kanewala, T. A., Marru, S., Basney, J., & Pierce, M. (2014, May). A credential store for multi-tenant science gateways. In *Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on* (445-454). IEEE.

[6] Madison J.D., Abeysinghe E., Pamidighantam S., Marru S., Christie M., Jennewein DM, and Pierce M. 2018. Science Gateway Implementation at the University of South Dakota: Applications in Research and Education. In Proceedings of the Practice and Experience on Advanced Research Computing (PEARC '18). ACM, New York, NY, USA, Article 97, 4 pages. DOI: https://doi.org/10.1145/3219104.3229265.

[7] Adithela, S.P., Christie, M., Marru, S. and Pierce, M., 2018, July. Django Content Management System Evaluation and Integration with Apache Airavata. In *Proceedings of the Practice and Experience on Advanced Research Computing* (p. 86). ACM.

[8] Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J. and Sahl, J.W. (2009). Introducing mothur: open-source, platform-independent, community- supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537-7541.