Short Paper: Managing Computational Gateway Resources with XDMoD

Jeanette M. Sperhac*, Robert L. DeLeon, Thomas R. Furlani, Steven M. Gallo, Martins Innus, Matthew D. Jones, Jeffrey T. Palmer, Abani Patra, Benjamin D. Plessinger, Ryan Rathsam, Nikolay Simakov, Joseph P. White, Thomas Yearke
Center for Computational Research, University at Buffalo, State University of New York, Buffalo, NY, 14203, USA
*Corresponding author email: jsperhac@buffalo.edu

***Abstract*—The U.S. National Science Foundation (NSF) has invested heavily in research computing, funding XSEDE to integrate supercomputers with science gateways. XD Metrics on Demand (XDMoD), an NSF funded tool that collects and presents detailed data about computing resources, was developed to manage such high performance computational resources. XDMoD metrics describe accounting and performance data for computational jobs including resources consumed (computation, memory, disk, network, etc.), wait times, and quality of service for these resources. XDMoD can provide information on individual jobs, or data aggregated over an ensemble of jobs. Its web interface offers centralized charting, exploration, and reporting of these metrics, for user-selected time ranges, and across all resources. Gateways users and administrators can also benefit from the tool. In this short paper, we introduce XDMoD, describe the state of its support for gateway resources, and outline our plans to further enhance its capabilities.**

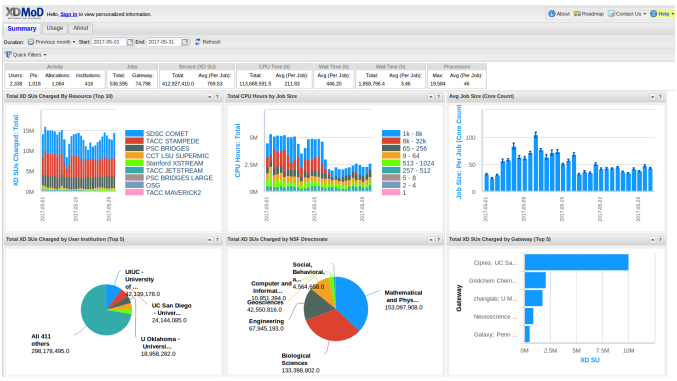


Fig. 1. XDMoD web interface

I. Introduction

The XD Metrics on Demand (XDMoD) tool provides stakeholders with ready access to data about utilization, performance, and quality of service for High Performance Computing (HPC) resources. [1]–[5] This comprehensive tool was originally developed to support resources for the National Science Foundation (NSF) XSEDE program; it was later open-sourced and made available to general HPC resources at universities, government laboratories, and commercial entities. [6] XDMoD enables users, managers, and operations staff to monitor, assess and maintain quality of service for their computational resources. To do this, XDMoD harvests data from the various resources and displays the resulting job, usage, and accounting metrics over any desired timeframe, using the XDMoD web interface and its rich array of visual analysis and charting tools. See Figure 1.

As XSEDE diversifies its computational portfolio, a variety of new resources, including gateways, come online. While XDMoD already supports aggregated gateways data and metrics, The XDMoD team is actively enhancing the tool with new capabilities to better report on scientific gateway usage. With these changes, we aim to improve our support for the scientific gateway community, both inside and outside XSEDE.

The remainder of this paper is organized as follows: Section II discusses the open-source version; Section III summarizes its core capabilities, as well as modules for integrated Application Kernels and performance-level job metrics; Section IV describes XDMoD's current support for gateways resources and their users; and Section V outlines future enhancements for gateways. Finally, Section VI provides a summary.

II. Open XDMoD

Though XDMoD was developed for the XSEDE program, the widely-adopted Open XDMoD enjoys active support and development, and is continually being enhanced with features requested by our users. This open-source version of XDMoD helps us manage our own computational center, the Center for Computational Research at the State University of New York at Buffalo. Open XDMoD can be installed on computational resources of any size, and configured to collect and represent data for academic or business concerns. [7] Installations of Open XDMoD number in the hundreds, and include industrial and academic centers and government labs in Australia, Brazil, Canada, France, Germany, India, Norway, Poland, Saudi Arabia, Spain, Turkey, the United Kingdom, the U.S., and elsewhere.[1]

Aside from XSEDE-specific customizations, the functionality of the XSEDE and Open XDMoD versions is the same; in this paper, the two will be referred to as "XDMoD".

[1]Institution and location are determined at download and upgrade time.

Presented at Gateways 2017, University of Michigan, Ann Arbor, MI, October 23-25, 2017. https://gateways2017.figshare.com/

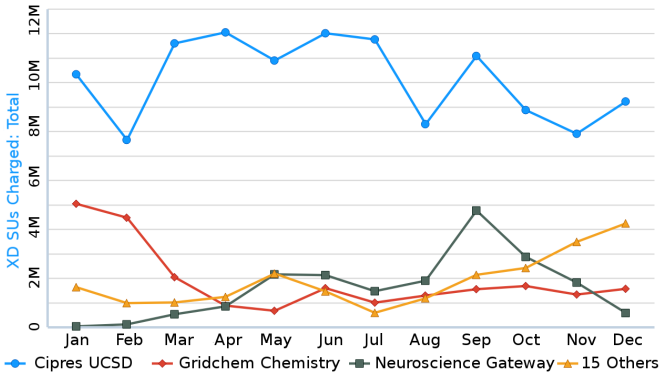


Fig. 2. *Total* XD SUs for Science Gateways Grants, 2016: Cipres (blue circles), Gridchem (red diamonds), Neuroscience (grey squares), All 15 Others (yellow triangles).

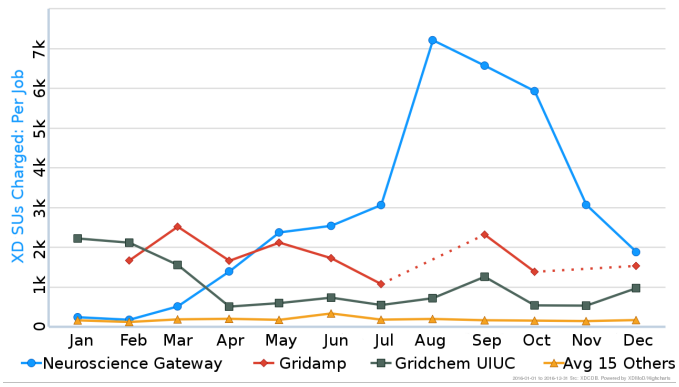


Fig. 3. XD SUs *per Job* for Science Gateways Grants, 2016: Neuroscience (blue circles), Gridamp (red diamonds), Gridchem (grey squares), Avg. of 15 Others (yellow triangles).

III. XDMoD Capabilities

XDMoD mines log files from computational resources to provide a multitude of job metrics that describe resource usage and utilization, including number of jobs, CPU hours, wall times, wait times, and job sizes. These metrics may be examined for a single job or aggregated over any desired time frame, in order to monitor and report collective job activity on a resource. XDMoD supports extensive data analytic functions such as filtering, grouping and drill-down. It provides reporting capabilities that include charting, data export, and custom report generation. Additional modules, briefly described below, extend XDMoD's base capabilities with quality of service and job level performance metrics.

The metrics collected by XDMoD are assembled into groups called Realms, based on the type of information they measure. For example, the Jobs realm metrics are gleaned largely from job accounting data, while the SUPReMM Realm includes metrics from job-level performance data.

A. Application Kernels Module

Application Kernels enable quality-of-service monitoring for HPC resources. [8], [9] The application kernel remote runner (AKRR) periodically runs computationally lightweight benchmarks, thus establishing baselines for application performance. These benchmarks help center personnel pinpoint underlying problems when a resource performs poorly.

Application Kernels are presently running on these XSEDE resources: SDSC Comet and LSU SuperMIC; PSC Bridges and TACC Stampede2 are also slated for inclusion. AKRR may also be configured to run with installations of Open XDMoD, and as such we run it at our own center.

B. SUPReMM Module

The SUPReMM module consists of components that collect data from system hardware counters and present job-level data for viewing and analysis in the XDMoD user interface. [10] Software running on compute nodes collects a wealth of performance information, such as memory usage, filesystem usage, interconnect fabric traffic, and CPU performance. [11]–[13] These data are supplemented with job submission script information, then ingested into the XDMoD database. The web interface presents these performance data in the SUPReMM realm.

The XDMoD web interface enables users to probe these performance data for individual jobs, or to aggregate over an ensemble of jobs. The Job Viewer presents details about a job's executable, its accounting data, job scripts, application information, and timeseries plots of individual metrics such as CPU user, flops, parallel file system usage, and memory usage.[2] These data reflect the details of each job and enable support personnel and resource users to troubleshoot individual job performance or evaluate job efficiency. [14] Alternatively, job performance data can be aggregated over a customizable time range to obtain an assessment of the performance of the resource as a whole.

Job performance data are presently collected on these XSEDE resources: TACC Stampede2, SDSC Comet, NICS Darter, and LSU SuperMIC. This job performance data pipeline may also be configured for use with installations of Open XDMoD, and as such we run it at our own center.

IV. XDMoD and Gateways

In its present form, XDMoD provides a number of metrics and dimensions useful to XSEDE gateway users, administrators, and center staff based on existing job reporting into the XSEDE Central Database. Gateway use is typically identified with the community user account used to run jobs on associated resources; the data returned provide a largely-complete summary of gateway usage on the resource, albeit aggregated by gateway.

A. Job metrics and dimensions

Metrics available in the Jobs realm of XDMoD entail counts of the jobs run and measures of the resources consumed by

[2]Individual jobs data are not publicly available; they are accessible to the user that submitted the job, and those such as support personnel and principal investigators who are considered managers of the job.

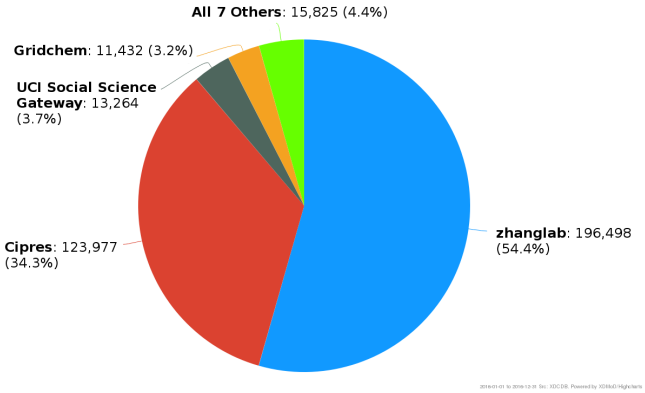


Fig. 4. Number of gateways jobs run on Comet, 2016

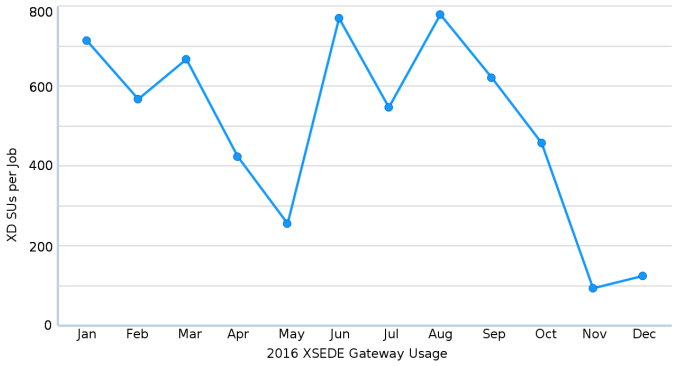


Fig. 5. XD SUs per job for Science Gateways Grants, 2016

these jobs. See Figures 2 through 5. These metrics, which incorporate gateways data, include:

- CPU hours
- Job size
- Node hours
- Number of jobs
- Wait hours
- Wall hours

Other metrics found in the Jobs realm describe XSEDE allocations and accounting, such as:

- Allocation usage rate
- XSEDE Normalized Units (NUs) charged [15]
- Number of allocations
- Number of institutions
- XD Service Units (SUs) charged [15]
- XSEDE utilization

Some metrics are reported in several different ways. For example, Number of Jobs is determined for running, ended, started, and submitted jobs. Similarly, CPU Hours and Service Units (SUs) Charged are reported both as totals and per job. To illustrate, Figure 2 displays total XD SUs, or the *total use* of gateways, over the 2016 year; Figure 3 shows XD SUs per job, or typical 2016 usage on the gateway, *per job*.

In Figures 2 and 3, XD SUs are illustrated, rather than CPU Hours, for a meaningful comparison of usage across resources

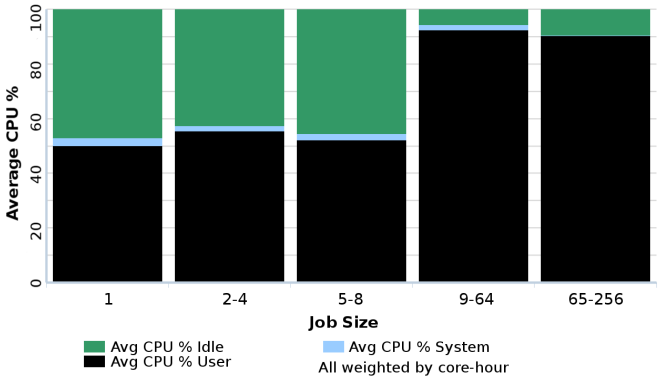


Fig. 6. Cipres Gateway 2016 CPU% by Job Size (core count): 1, 2-4, 5-8, 9-64, and 65-256.

with varying configurations. The XD SUs metric reports computational resources consumed in units standardized across XSEDE; the conversion is derived from High-Performance LINPACK (HPL) benchmarks on each resource.[3] [15]

Using XDMoD, Jobs metrics can be grouped by different dimensions for plotting and further analysis. Available dimensions in the Jobs realm include Gateway, Grant Type, Job Size, PI, Resource, and User.

B. Performance metrics and dimensions

XDMoD users can easily chart a wide variety of metrics on aggregate gateway jobs and usage, provided that the gateway's underlying resources collect and report job performance data. Performance metrics include:

- Total memory
- CPU user
- Memory bandwidth
- I/O bandwidth
- Block read/write rate

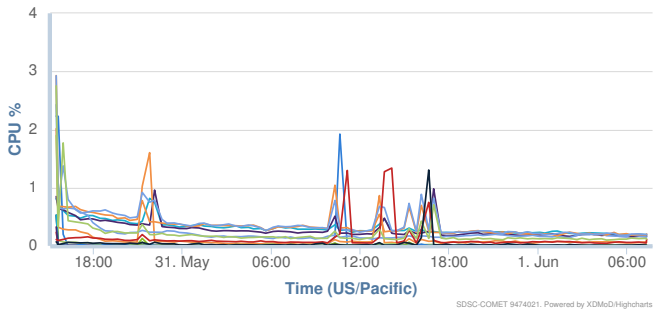
Performance realm dimensions can then be used to group metrics values. These dimensions include Application, CPI Value, Data source, Exit Status, Job Size, Queue, Share Mode, and User. Refer to Figure 6 for an example plot.

In order to view metrics by Gateway in the SUPReMM realm, the user can filter by User, selecting *Community User*; Gateways jobs are currently reported to XDMoD by their associated resources as Community User jobs.

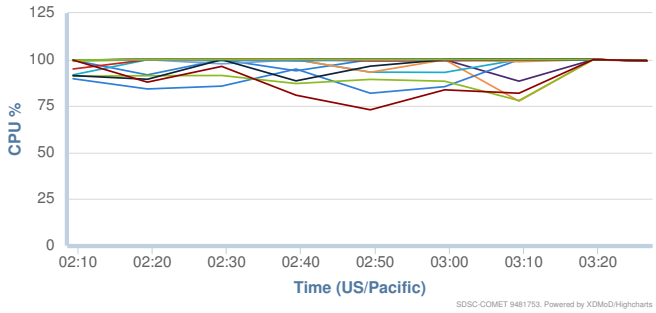
The XDMoD Job Viewer tab affords a window into the job efficiency metrics that performance data provide. There, users may view individual jobs' detailed statistics and job data timeseries plots, by drilling down on SUPReMM metrics.[4] In Figure 7 we contrast two jobs running the same executable, each in exclusive mode. The figure clearly displays a substantial difference in resource utilization between the two jobs; Figure 7a shows inefficient use of the requested resources, with CPU User less than 3% for 24 cores; Figure 7b, meanwhile,

[3]An XSEDE SU is defined as one CPU-hour on a Phase-1 DTF cluster; a Phase-1 DTF SU is equal to 21.576 NUs. NUs measure resources consumed as charged to a roaming allocation.

[4]Where user permissions allow.



(a) Inefficient resource utilization: sub-3% CPU User for 24 cores.



(b) Efficient resource utilization: near 100% CPU User for 24 cores.

Fig. 7. Job Viewer timeseries plots compare CPU User % (the average percentage of time spent in CPU user mode) for two different Cipres Gateway jobs on Comet. Both jobs are running the same executable. Note the scale difference of the y-axes. Each line represents one of 24 cores.

shows CPU User near 100% for 24 cores.[5] The Job Viewer lets user support staff and end users examine their job data in detail, enabling improved identification and diagnosis of inefficient jobs, and better use of resources.

C. Simple gateways plots in XDMoD

Gateways-relevant plots of various kinds are available in the XDMoD metric catalog. For instance, a resource manager might be interested in *a)* the number of gateway jobs run on their resource; *b)* the aggregated computational resources used per gateway job, over some timeframe; or *c)* the computational resources consumed, by different gateways.

To illustrate these cases, Figure 4 shows the total number of gateway jobs submitted to Comet in 2016. Figure 5 displays XSEDE XD SUs per job across all science gateways in 2016.[6] To show the breakdown of jobs per gateway, drill down into the metric by clicking on the plot and select *By Gateway*. Figure 3 thus shows XSEDE gateway XD SUs per job in 2016.

The XDMoD user interface offers extensive plot customization. The user can plot a metric as a timeseries or as an aggregate; change metric time windows and time aggregations; overplot multiple datasets; and tailor the number of individual dimensions shown. In addition, the user can represent the data as a line, bar, scatter, or pie chart; scale axes; edit labels; and customize the plot's data and its appearance.

[5]These jobs are floating point intensive, and use low memory bandwidth.

[6]Performance metrics are currently visible for jobs you submitted or manage.

To demonstrate performance realm plots in aggregate, we present a plot showing CPU usage on the Cipres gateway over a full year. Refer to Figure 6. This plot displays the average percent of CPU that was idle, used by the system, and devoted to user commands, weighted by core hours, for all Cipres jobs executed in 2016. The results are provided as a function of job size (core count). The plot suggests that larger Cipres jobs appear to use the CPU more efficiently, when viewed in aggregate.

V. Future work

Although XDMoD presents aggregated data relevant for gateways users and managers, several current shortcomings of XDMoD for gateways data are evident: *a)* Since individual gateway user IDs are not yet processed by XDMoD, gateway users cannot yet be directly searched or filtered on in the web interface. *b)* Since gateway managers lack access to data across all resources to which their gateways submit jobs, they cannot access individual job performance data, which could help troubleshoot problematic gateways jobs.

To address these gaps, the XDMoD team plans to develop a dedicated gateway data realm that will treat gateways as their own entities, much as service providers and computing resources are treated now. In this new realm, individual gateway users will be directly linked to the jobs they run on a resource, regardless of whether these users have their own logins on the resource. This enhancement will allow drilldown, filtering, and searching by gateway user. The realm will also permit gateway managers to access individual job data pertinent to their own gateways, and likewise allow gateway users to view the details of their own gateway jobs. This proposed development will treat gateways as first-class citizens and facilitate truly granular access to gateway user behavior and job management.

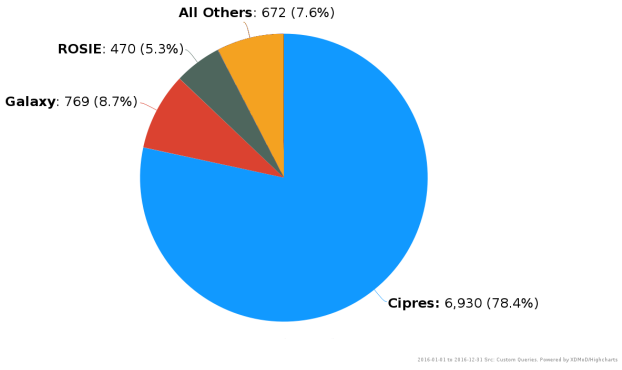


Fig. 8. Unique Gateway Users, 2016. Number of unique gateway users who ran jobs during 2016, listed by gateway.

The new gateways realm will be prototyped in-house at CCR using VIDIA [16], a gateway that supports data-intensive computation for university campuses that lack access to HPC resources. As such, VIDIA enables access to CCR's HPC cluster, where VIDIA users ran more than 2000 jobs in the first 9 months of 2016. The Open XDMoD instance running at CCR collects and aggregates information on the VIDIA

gateway, and will serve as a testbed for collecting, ingesting, and aggregating data on gateway resource access and user jobs. We are confident that better integration of gateways data in XDMoD will help us manage our own gateway, and will similarly assist XSEDE and open-source gateway managers and users.

In upcoming XSEDE XDMoD release 7.1, a new set of aggregate gateway plots will be available to all users under the Custom Queries tab. This series of plots shows counts of jobs or users across gateways and resources, with adjustable timeframes, using data attributes not yet fully exploited by XDMoD. See Figure 8. The thirteen gateway-specific queries are plotted as pie charts, such as:

- Unique gateway users by resource and/or gateway
- Number of gateway jobs by resource and/or gateway
- Number of gateway jobs submitted by unknown user
- Gateway XDSU vs. total XSEDE XDSU

Until the gateways realm is available, we offer this new view of gateway-specific data on the XSEDE systems. Meanwhile, the XDMoD team requests feedback from the gateway user community about other enhancements that would make the tool more useful and relevant.

VI. Conclusion

XDMoD, a comprehensive tool for managing HPC resources, collects and presents numerous metrics that describe computational jobs, resources consumed (computation, memory, disk, network, etc.), wait times, and quality of service. The XDMoD web interface offers tools for charting and visualizing these metrics, enabling interactive drill-down to expose and plot related information. Gateways users and administrators, whether they use XSEDE or Open XDMoD, can benefit from the data and metrics offered in XDMoD. We aim to further enhance the tool for the gateways community.

Acknowledgments

This work is supported by the National Science Foundation under grant numbers OCI 1025159 (Technology Audit Service), OCI 1203560 (SUPReMM), and ACI 1445806 (XD Net Metrics Service for High Performance Computing).

References

[1] Center for Computational Research, University at Buffalo, "XD Metrics on Demand (XDMoD)," https://xdmod.ccr.buffalo.edu, 2017.

[2] J. C. Browne, R. L. DeLeon, A. K. Patra, W. L. Barth, J. Hammond, M. D. Jones, T. R. Furlani, B. I. Schneider, S. M. Gallo, A. Ghadersohi, R. J. Gentner, J. T. Palmer, N. Simakov, M. Innus, A. E. Bruno, J. P. White, C. D. Cornelius, T. Yearke, K. Marcus, G. von Laszewski, and F. Wang, "Comprehensive, open-source resource usage measurement and analysis for HPC systems," *Concurrency and Computation: Practice and Experience*, vol. 26, no. 13, pp. 2191–2209, 2014, cPE-14-0027.R1. [Online]. Available: https://dx.doi.org/10.1002/cpe.3245

[3] T. R. Furlani, M. D. Jones, S. M. Gallo, A. E. Bruno, C.-D. Lu, A. Ghadersohi, R. J. Gentner, A. K. Patra, R. L. DeLeon, G. von Laszewski, L. Wang, and A. Zimmerman, "Performance metrics and auditing framework using applications kernels for high performance computer systems," 2012.

[4] T. R. Furlani, B. I. Schneider, M. D. Jones, J. Towns, D. L. Hart, S. M. Gallo, R. L. DeLeon, C. Lu, A. Ghadersohi, R. J. Gentner, A. K. Patra, G. Laszewski, F. Wang, J. T. Palmer, and N. Simakov, "Using XDMoD to facilitate XSEDE operations, planning and analysis," in *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery (XSEDE '13)*. ACM, 2013, p. 8.

[5] J. C. Browne, R. L. DeLeon, C. Lu, M. D. Jones, S. M. Gallo, A. Ghadersohi, A. K. Patra, W. L. Barth, J. Hammond, T. R. Furlani, and R. T. McLay, "Enabling comprehensive data-driven system management for large computational facilities," in *Proceedings of SC13: International Conference for High Performance Computing, Networking, Storage and Analysis (SC '13)*. ACM, 2013, p. 11.

[6] J. T. Palmer, S. M. Gallo, T. R. Furlani, M. D. Jones, R. L. DeLeon, J. P. White, N. Simakov, A. K. Patra, J. M. Sperhac, T. Yearke, R. Rathsam, M. Innus, C. D. Cornelius, J. C. Browne, W. L. Barth, and R. T. Evans, "Open XDMoD: A tool for the comprehensive management of high-performance computing resources," *Computing in Science and Engineering*, vol. 17, no. 4, pp. 52–62, 2015. [Online]. Available: https://dx.doi.org/10.1109/MCSE.2015.68

[7] J. P. White, M. Innus, M. D. Jones, R. L. DeLeon, N. Simakov, J. T. Palmer, S. M. Gallo, T. R. Furlani, M. Showerman, R. Brunner, A. Kot, G. Bauer, B. Bode, J. Enos, and W. Kramer, "Challenges of Workload Analysis on Large HPC Systems: A Case Study on NCSA Blue Waters," in *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*, ser. PEARC17. New York, NY, USA: ACM, 2017, pp. 6:1–6:8. [Online]. Available: http://doi.acm.org/10.1145/3093338.3093348

[8] N. A. Simakov, J. P. White, R. L. DeLeon, A. Ghadersohi, T. R. Furlani, M. D. Jones, S. M. Gallo, and A. K. Patra, "Application kernels: HPC resources performance monitoring and variance analysis," *Concurrency and Computation: Practice and Experience*, vol. 27, no. 17, pp. 5238–5260, 2015, cPE-14-0402.R1. [Online]. Available: https://dx.doi.org/10.1002/cpe.3564

[9] T. Furlani, M. Jones, S. Gallo, A. Bruno, C. Lu, A. Ghadersohi, R. Gentner, A. Patra, R. DeLeon, G. Laszewski, L. Wang, and A. Zimmerman, "Performance metrics and auditing framework using application kernels for high-performance computer systems," *Concurrency and Computation: Practice and Experience*, vol. 25, p. 918, 2013.

[10] Center for Computational Research, University at Buffalo, "XDMoD SUPReMM Module," http://supremm.xdmod.org/, 2017.

[11] J. P. White, R. L. DeLeon, T. R. Furlani, S. M. Gallo, M. D. Jones, A. Ghadersohi, C. D. Cornelius, A. K. Patra, J. C. Browne, W. L. Barth, and J. Hammond, "An analysis of node sharing on HPC clusters using XDMoD/TACC_Stats," in *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, ser. XSEDE '14. New York, NY, USA: ACM, 2014, pp. 31:1–31:8. [Online]. Available: https://dx.doi.org/10.1145/2616498.2616533

[12] T. Evans, W. L. Barth, J. C. Browne, R. L. DeLeon, T. R. Furlani, S. M. Gallo, M. D. Jones, and A. K. Patra, "Comprehensive resource use monitoring for HPC systems with TACC_Stats," in *Proceedings of the First International Workshop on HPC User Support Tools*, ser. HUST '14. Piscataway, NJ, USA: IEEE Press, 2014, pp. 13–21. [Online]. Available: https://dx.doi.org/10.1109/HUST.2014.7

[13] R. L. DeLeon, T. R. Furlani, S. M. Gallo, J. P. White, M. D. Jones, A. K. Patra, M. Innus, T. Yearke, J. T. Palmer, J. M. Sperhac, R. Rathsam, N. A. Simakov, G. von Laszewski, and F. Wang, "TAS view of XSEDE users and usage," in *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, ser. XSEDE '15. New York, NY, USA: ACM, 2015, pp. 21:1—-21:8. [Online]. Available: https://dx.doi.org/10.1145/2792745.2792766

[14] Center for Computational Research, University at Buffalo, "XDMoD Single Job Viewer," https://www.xsede.org/documents/10157/1125659/XDMoD_OverviewDemo_Mar2016.pdf/b92eb25d-db12-4710-a1f8-3cf19acf0c0c.

[15] XSEDE (Extreme Science and Engineering Discovery Environment), "XSEDE SU Converter," https://www.xsede.org/su-converter, 2017.

[16] J. Sperhac and S. Gallo, "VIDIA: a HUBzero Gateway for Data Analytics Education, plus a demo of the VIDIA Gateway: a Virtual Infrastructure for Data Intensive Analysis," 1 2016. [Online]. Available: https://figshare.com/articles/Gateways2016_Sperhac_Gallo/4488308