

Overview. The research computing landscape – hardware, software, applications, and expectations – continues to evolve rapidly. Machine learning and big data analytics are creating new opportunities to extract research insight from experimental and computational data. Concurrently, cloud service providers are rapidly deploying new computing and data analysis services atop hardware configurations that are increasingly similar to those found in academic environments. All of this ferment is convolved with a new paradigm of streaming data from a world of scientific instruments and ubiquitous wireless sensors. Concurrently, new expectations for data management and retention, combined with resource constraints, are putting new pressures on research agencies and academic institutions to find cost-effective approaches that maximize scientific benefit while minimizing costs.

This project proposes to build, publish, and regularly update comparisons of the evolving total cost of ownership (TCO) of on premise HPC clusters and data storage systems relative to NSF-supported facilities and commercial cloud services such as Amazon AWS, Microsoft Azure, and Google Cloud Platform. Concretely, the team proposes to build and host a Jupyter notebook that enables the cyberinfrastructure community to add, modify, and explore total cost of ownership models based on a variety of usage patterns and performance expectations. This will facilitate community building, shared experimentation and comparisons.

Intellectual Merit. The shifting milieu of computing infrastructure raises several important questions about how best to support scientific research. For example, when might commercial clouds be more cost effective than university clusters in servicing researcher needs? How do staff costs, hardware utilization, capital depreciation, replacement costs, the value of money, return on investment, power and cooling costs, and bandwidth charges affect TCO? As science gateways hide resources behind web interfaces, when and where should such queries execute? With the Internet of Things (IoT), environmental and biomedical sensors, and quantitative social measurement systems, how can this data best be analyzed and correlated, while assuring adequate privacy and security?

Quite clearly, there is no single, simple answer to these questions. However, it is possible to develop comparative models that allow funding agencies, institutional resource providers, and individual users to evaluate the costs and benefits of specific choices and subsidies. These span, but are not limited to university HPC clusters and storage (centrally provisioned and condo/hotel models, and commercial computing and data management services).

Broader Impacts. Understanding technical computing TCO and provisioning alternatives has broad applicability to the national research community, institutional resource providers, and funding agencies. The team will collaborate with a wide range of academic, HPC and national organizations to create and share available data and capability for comparative analysis and resource acquisition by universities, research agencies, and the national community. These include the Advanced Cyberinfrastructure – Research and Education Facilitators (ACI-REF), the Coalition for Academic Scientific Computation (CASC), and academic CIOs. Because the value of such a resource rests on its historical (longitudinal) data and long-term sustainability, we will also engage the Association of American Universities (AAU) and the Association of Public and Land Grant Universities (APLU) senior research officers (SROs) and CIOs to promote long-term maintenance and data sharing. Simply put, the goal is to create a sustainable infrastructure for analysis. Consequently, the insights from this study will (a) have broad influence on the sharing of data on technical computing costs, configurations, and usage patterns, (b) help shape system configuration and pricing choices, and (c) facilitate national policy discussions about sustainable cyberinfrastructure.

Project Description: CYBER-INSIGHT: Evaluating Cyberinfrastructure Total Cost of Ownership

1. Introduction

The research computing landscape – hardware, software, applications, and expectations – continues to evolve rapidly. Machine learning and big data analytics are creating new opportunities to extract research insight from experimental and computational data, while heterogeneous multicore processors, GPUs and accelerators, burst buffers, and advanced networks are enabling new computational science. Amidst fierce price competition, commercial cloud service providers are rapidly deploying new computing and data analysis services atop hardware configurations that are increasingly similar to those found in academic environments. All of this ferment in technical computing is convolved with a new paradigm of streaming data from a world of scientific instruments and ubiquitous wireless sensors. Concurrently, new expectations for data management and retention, combined with resource constraints, are putting new pressures on research agencies and academic institutions to find cost-effective approaches that maximize scientific benefit while minimizing costs.

This shifting milieu raises several important questions about how best to support scientific research. For example, when might commercial clouds be more cost effective than university clusters in servicing researcher needs? Where and how should active data (i.e., used regularly) and cold data (i.e., important but rarely used) be stored to balance costs, access delays, and research productivity? How do staff support costs, hardware utilization, capital depreciation, replacement costs, and the value of money, return on investment, power and cooling costs, and bandwidth charges affect total cost of ownership (TCO)?

In addition, what are the benefits of elasticity and the costs of availability subject to uncertainty? How are evolving virtualization technologies (e.g., containers) affecting service availability and long-term maintainability? What are the long-term social, workforce, and technical implications of potentially outsourcing this expertise in the private sector? As science gateways hide resources behind web interfaces, when and where should such queries execute? What burdens do off-campus compute and storage resources place on network bandwidth and reliability? With the Internet of Things (IoT), environmental and biomedical sensors, and quantitative social measurement systems, how can this data best be analyzed and correlated, while assuring adequate privacy and security?

Quite clearly, there is no single, simple answer to these questions, as they depend on institutional circumstances, research domains, user characteristics, scientific and engineering expectations, and a complex set of policy issues. However, it is possible to develop comparative models that allow funding agencies, institutional resource providers, and individual users to evaluate the costs and benefits of specific choices and subsidies (condo/hotel models). As one example, the *XDMoD Value Analytics* project is already exploring the scientific return on investment (ROI) via cyberinfrastructure investments [6].

As a complement to scientific ROI, the proposed work would consider the TCO tradeoffs of on-premise and cloud infrastructure for batch-oriented computation, streaming data and gateway services, active data storage, and persistent, archival storage (i.e., infrequently used, but preserved for long periods). To validate this idea, we have already developed a series of Excel spreadsheets that show the TCO of NSF's leading edge HPC facilities (Blue Waters and XSEDE), commodity clusters, and cloud services, based on expected performance levels and utilizations. In this work, we propose to extend and generalize this approach.

This proposal outlines a plan to formalize and parameterize a TCO assessment process for community use. It includes building, publishing, and regularly updating comparisons of the evolving TCO of on premise HPC clusters and data storage systems relative to selected NSF infrastructure (HPC platforms and cloud systems such as Jetstream) and commercial cloud services such as Amazon AWS, Microsoft Azure, and Google Cloud Platform.¹ Concretely, we propose to build and host a Jupyter notebook [7, 8] that enables the cyberinfrastructure community to add, modify, and explore total cost of ownership models based on a variety of usage patterns and community needs

¹ We do not envision commercial services replacing NSF's leading edge facilities. Rather, we seek to evaluate the TCO for scientific analysis requiring moderate levels of parallelism, streaming data analytics, or other cloud-based services.

2. Cyberinfrastructure Cost and Performance Models

The total cost of ownership (TCO) of computing services in academia and government is sometimes difficult to capture accurately, as it often includes a variety of explicit and implicit subsidies. These include such things as utilities (though energy consumption (power and cooling) by HPC systems is now increasingly recognized), facility capital costs and debt service, network infrastructure and connection fees, software licenses, and staff salaries. In

addition, steady state operation must also include a budget for equipment replacement costs. By comparison, commercial computing services must incorporate all of these costs plus some profit margin, while remaining responsive to competitive market pressures.

Local Cluster Cost Components	Estimated Three Year Costs
Personnel	\$900,000
Facility	\$150,000
Cluster hardware	\$1,709,000
Energy infrastructure	\$50,000
Energy consumption	\$420,480
Storage	150 TB
I/O operations	20 billion/month

Below, we illustrate this complexity in an assessment of the TCO of both a local cluster and cloud services for a simple scientific benchmark, following by a broader discussion of approaches to quantifying TCO.

2.1 Cloud and Cluster Comparison Example

Consider a simple TCO analysis for the classic HPL (Linpack) [9], benchmark on both a local HPC cluster and a comparable cluster hosted in the cloud. In this case, we consider both reserved (prepaid) and on-demand cloud cluster instances. As a concrete example of a typical university research cluster, we used the University of Iowa's "Neon" system [10], a "condo model" shared campus cluster with 259 compute nodes, most of which are connected via QDR Infiniband and several of which include Intel Xeon Phi or NVidia Kepler accelerators. Based on campus measurements, the cluster's average utilization was roughly 70% (i.e., about 70 percent of the nodes were typically in use). In turn,

Table 1 shows the estimated cost for three years of the Neon cluster's operation, as reported by university IT staff.

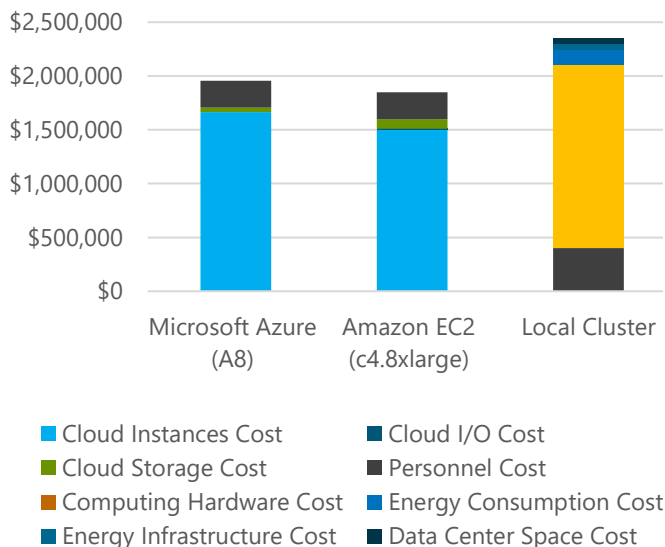


Figure 1 One Year TCO Comparison
AWS 3 year reserved full upfront, Azure on demand

Table 2 shows HPL performance for both the Neon cluster and two cloud cluster configurations, drawn from benchmarks and published data. From this, one can estimate the number of cores and cloud instances required to equal the HPL performance on the Neon cluster.² This data, together with published costs for on-demand, reserved, and spot instances³ of Microsoft Azure and Amazon AWS, suffice to estimate TCO under several scenarios, including the expected infrastructure usage period (years), average system utilization (fraction of nodes that are busy), and possible hardware replacement cost for the local cluster. In this comparison, one must remember that cloud service costs necessarily include hardware replacement costs plus vendor profit. Thus, a long-term

comparison to a campus cluster would include either cluster depreciation or the cost to refresh the hardware periodically.

² Linear scaling is, of course, incorrect, but it suffices to provide a rough scaling for cost analysis.

³ Spot instances are cloud resources acquired by online bidding and can be lost to a higher bidder

Cluster Hardware	Data Source	Core Count	Rmax (TFLOPS)	Rmax TFLOPS/core	Neon Equivalent Nodes
Amazon EC2 Cluster Intel Xeon E5-2666 v3 2.9 GHZ, 10G Ethernet (c4.8xlarge EC2)	Benchmark	180	2.96	0.016	162
Microsoft Azure Cluster Platform SL230 Gen8, Xeon E5-2670 2.6 GHz, Infiniband QDR (Azure A8 equivalent)	TOP500 November 2012	8,064	151.3	0.019	319
University of Iowa Neon HP Intel Xeon E5-2670 2.6 GHz, Infiniband QDR	Benchmark	4,240	47.8	0.011	530

Table 2 HPL Benchmark Comparison

With this backdrop, Figures 1-3 compare the TCO of the Neon cluster to an Amazon AWS cluster and a Microsoft Azure cluster under different assumptions. First, Figure 1 considers a single year of use, where the AWS cluster was reserved (full cost paid upfront) and the Azure cluster use was paid on demand. In this case, the upfront capital cost of an on premise cluster dominates because the hardware costs cannot be fully amortized (i.e., the useful lifetime of the on-premise hardware is greater than one year). Figure 2

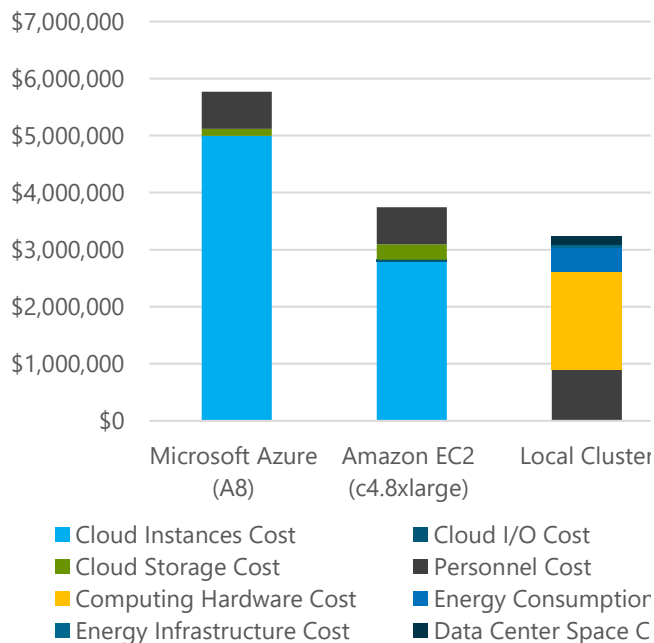


Figure 2 Three Year TCO Comparison
AWS 3 year reserved full upfront, Azure on demand

shows the same options but examines three years of usage where the cluster cost is fully amortized. Finally, Figure 3 compares the TCO of smaller cloud clusters whose nodes are always utilized to an on premise cluster where the nodes are busy only 70 percent of the time. This capacity planning question balances lower equipment capital costs against higher user queueing delays (i.e., the relative importance of equipment and user costs).

As even this simple example shows, many variables and assumptions affect TCO and provisioning decisions. Perhaps most importantly, the example illustrates the dramatic effect of expected demand and period of use on TCO optimization. Different assumptions about hardware utilization (percent of nodes busy), relative node performance on target applications, cloud costs, I/O volume and frequency, and network transfer volumes and unit costs, yield different ratios. It is precisely this complexity, along with shifting hardware and cloud service costs, which motivates our proposal to build a community infrastructure and service to support these comparisons.

2.2 Ecosystem Measurement and Consistency

Building and sustaining any measurement and comparison infrastructure is a balance between generality (i.e., including a wide variety of measurements and workloads) and longitudinal consistency (i.e., capturing comparable data over long periods for trend analysis). The persistence of the HPL benchmark [1, 9] illustrates this tension – it is a single figure of merit that demonstrably is not representative of application performance on current systems, but it is one of the few longitudinal data sets on HPC performance, system scale, and (more recently) energy consumption.

Modern cyberinfrastructure includes not only traditional, numerically intensive simulations but also scientific workflows and gateways [4, 8], complex data analysis and machine learning, streaming data processing and storage from sensor networks and scientific instruments, and persistent, archival storage of important data. Given this diversity, it is impossible for a single project, no matter how broad, to capture and analyze all possible resource demands and associated infrastructure costs.

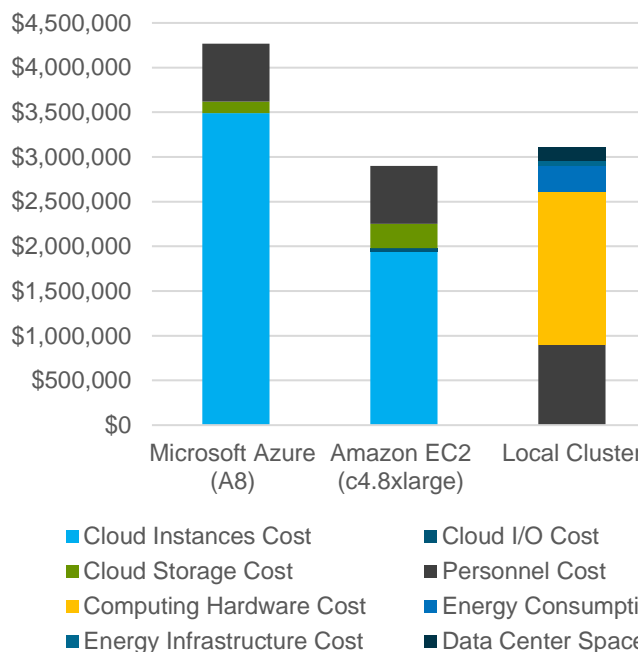


Figure 3 Three Year TCO Comparison
AWS 3 year reserved full upfront, Azure on demand
70% local nodes busy/100% cloud nodes busy

TCO assessment and infrastructure evaluation cannot become dominated by expensive benchmarks, or they will be neither valuable nor used; it must be simple and efficient to capture and record the needed data. *Consequently, our goal is not to create yet another benchmark suite, either for HPC or cloud services, nor is it to explore detailed configuration tuning.* Instead, our goal is to support comparison of the most common cyberinfrastructure demand types (i.e., those typical of those supported by institutional cyberinfrastructure) and costs, and allow addition of new demands as research and scholarship evolve.

For simplicity's sake, we propose to begin with a modest set of applications and services, for both community engagement and implementation effort. These will include four broad categories, each with an initial set of benchmarks and baseline infrastructure contexts

- High-performance computing (moderate, not leading edge)
- Input/output
- Data analytics
- Machine learning

Table 3 shows possible, initial targets in each of these four domains. We emphasize that these are only potential targets. This project can succeed only with community acceptance and participation. Thus, we propose to launch the project with a series of presentations and discussions to build consensus on an initial benchmark suite. With that backdrop, we offer rationales for some potential benchmarks.

First, HPL and HPCG [2] are well-known and widely used cluster benchmarks, providing a wealth of data for traditional scientific computing systems. Likewise, there has been extensive analysis of the “common workloads” on NSF’s Blue Waters system [19], and similar data is available for other NSF systems. Likewise, the scientific and open source communities have converged around a set of widely used tools for data processing and storage (e.g., Mosquitto (streaming data) and Cassandra or MongoDB (No SQL databases), as well as analysis (e.g., Hadoop and Spark). These are complemented by vendor cloud solutions for such things as sensor and IoT analysis (e.g., AWS Kinesis and Azure Event Hubs).

Assessment Area	Exemplar Capability Proxy Applications and Services
Scientific computation	HPL [1], HPCG [2], “common workloads” [19], XSEDE Gateways [3-5]
Input/output	File access, Mosquito, SQL, No-SQL (Cassandra, MongoDB)
Data analysis	Hadoop, Spark, Apache Kafka, AWS Kinesis, Azure Event Hubs
Machine learning	Google TensorFlow, Microsoft CNTK, Community engagement

Table 3 TCO Assessment Domains

The intersection of HPC and machine learning is evolving extremely rapidly, with both similarities to and distinct differences from many commercial applications. The rapid evolution highlights the importance of community collaboration to select appropriate benchmarks and tool suites.

Finally, based on experience and community feedback, we will also explore extending the benchmarks to include near-line and offline (cold) storage. This includes on-premise archival storage (e.g., via HPSS) and cloud-based storage systems such as AWS Glacier, Azure Cool Blog, and Amazon Coldline.

Because many of the benchmarks are large, complex software systems, deployment on any system can be a labor-intensive task. To simplify measurement and encourage rapid assessment, we propose to package each of the benchmarks as a container, drawing on existing containers for each benchmark wherever possible. As an example, CloudSuite [11, 12] includes benchmarks for data analytics, graph analytics, and (among others), data serving, all packaged as Docker containers. These containerized benchmarks are also being integrated with Google’s PerfKit Benchmarker [13], allowing cross-cloud comparisons.

Again, we emphasize that our focus is on TCO assessment and configuration planning, not on detailed, comparative performance benchmarking. Nor is it to assess the performance or TCO of nation-scale, leading edge computing infrastructure. Rather, our goal is to develop a community resource that will support analysis and understanding of common cyberinfrastructure tradeoffs.

2.3 TCO Community Building via Jupyter Notebooks

Jupyter notebooks [7, 8] have been adopted rapidly and widely as a mechanism to create and share documents that contain code, mathematics, visualizations, and text. With support for a variety of programming languages and tools, as well as modeling and data analytics packages, Jupyter notebooks allow communities to collaborate and share insights in new and important ways. In addition, notebook outputs can be exported in many formats, including web pages.

We propose to build on this community to create a Jupyter notebook that supports capture and analysis of TCO data from NSF facilities, campus cyberinfrastructure, and cloud services. Given their wide adoption, we believe the use of Jupyter notebooks will enable community building, shared experimentation, and resource comparisons. We expect the notebook to include, at a minimum, the following features:

- Data ingestion and storage for cloud services and campus infrastructure configurations
 - Infrastructure configuration (hardware and software specification)
 - Price/cost data components (equipment, software, facilities, utilities, staff, ...)
 - Benchmark results
- Interactive data exploration and visualization, as well as code-based analysis
 - Operational infrastructure lifetime
 - Expected utilization (frequency, counts, ...)
 - Cost or price changes
- Web page export for data sharing via Jupyter *nbviewer*

As Figure 4 illustrates, we will initially target Amazon AWS, Microsoft Azure, and Google Cloud Platform as commercial service exemplars. In addition, we will consider NSF cyberinfrastructure services, including both HPC platforms and cloud service infrastructure such as Jetstream. Both will complement data from campus infrastructure deployments. The Jupyter notebook will be web accessible for community

participation. In addition, the notebook and associated data will be packaged as a Docker container for rapid redeployment at other sites.

The efficacy of such a platform depends, quite obviously, on its community acceptance and use. To encourage community engagement, we plan to work with the

Advanced Cyberinfrastructure – Research and Education Facilitators (ACI-REF), the Coalition for Academic Scientific Computation (CASC), the Science Gateways Community Institute (SGCI), and academic CIOs to share information and develop models. We will also engage the Practice & Experience in Advanced Research Computing community (formerly XSEDE annual meeting).

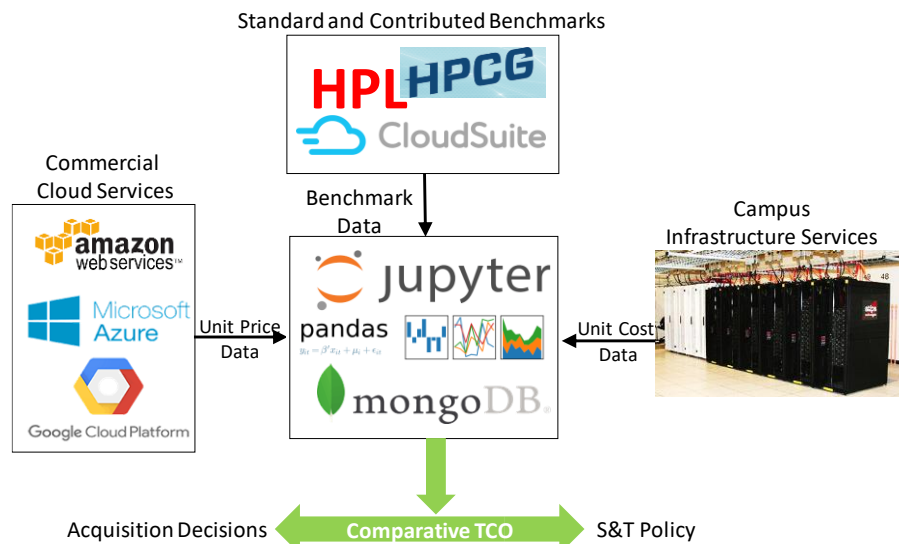


Figure 4 Cyber-Insight Infrastructure

Because the value of such a resource rests on its historical (longitudinal) data and long-term sustainability, we will also engage the Association of American Universities (AAU) and the Association of Public and Land Grant Universities (APLU) senior research officers (SROs) and CIOs to promote long-term maintenance and data sharing. Our goal is to create a sustainable infrastructure for analysis.

3. Broader Impacts

We expect the insights from this study to have broad influence on the sharing of data on technical computing costs, configurations, and usage patterns. As academic institutions feel increasing demands and expectations for cyberinfrastructure while also facing new funding pressures, there is great value and insight to be gained from data sharing and collaborative exploration of effective approaches. These benefits extend to a comparative analysis of on-premise and commercial cloud computing alternatives for both computing and storage access, as a function of workloads. Because we will be working with the academic cyberinfrastructure providers and commercial vendors, we believe our work will also influence national cyberinfrastructure policy.

In addition, integrating research, education and public engagement have long been hallmarks of our collaborations. During the project, we will involve faculty, staff, and students in all aspects of the infrastructure design, leveraging community tools. During his career, Professor Reed has worked with a wide variety of academic, government and commercial groups on cyberinfrastructure design, deployment, and operations, providing a basis for information sharing and community involvement. Consequently, the insights from this study will (a) have broad influence on the sharing of data on technical computing costs, configurations, and usage patterns, (b) help shape system configuration and pricing choices, and (c) facilitate national policy discussions about sustainable cyberinfrastructure.

4. Results from Prior NSF Support

For over thirty years, Professor Reed has been active in a wide variety of multidisciplinary research collaborations and community building initiatives supported by NSF. Examples include the Microsoft-NSF collaboration on cloud computing [14] and the Computing Research Association's (CRA) Community Computing Consortium (CCC) [15]. Within the past five years, Professor Reed has been supported by **NSF ACI-1349521 EAGER: Resilient, Energy Efficient HPC System Configuration, 9/2013-9/2016, \$298,828**. At a high level, this work investigated the cost and reliability of high-performance computing (HPC) building blocks, drawing on insights and ideas from containerized cloud computing deployments and

large-scale energy optimization. For additional details, see the website: hpc.research.uiowa.edu.Publications [16, 17].

Intellectual Merit. Our society increasingly depends on very large computing systems, whether for communication, information access, electronic commerce or access to government services. In industry and scientific research, large-scale computing is used to model complex phenomena and to gain insight into mathematical models. Thus, the reliability of these commercial cloud and high-performance computing systems is critical to our world. Likewise, as these systems grow in size, they consume ever-larger amounts of energy for their operation.

The objective of this research was to explore new approaches to reliability by adding extra hardware to systems for recovery after failures. In addition, the research also sought to quantify the total cost of ownership – hardware, facilities, personnel, and energy – required to operate these large systems. Using mathematical models and observational data, this work explored hardware resilience models and the cost effectiveness of over-provisioning versus just-in-time replacement of failing parts and human capital costs. Our research showed that the best design choices for reliability are defined by the relative costs of money, personnel costs and equipment type. We also showed that for large-scale research computing systems, operating costs are approaching the cost of equipment and facilities. Finally, we showed that for smaller scientific computing facilities with lower utilization, cloud computing may be an attractive economic alternative.

Broader Impact. This work drew on ideas and experiences from commercial cloud computing to inform the design of reliable, energy efficient high-performance computing platforms, based on Reed's experience as leader of Microsoft's cloud and data center futures team and the eXtreme Computing Group (XCG). He and his team developed integrated hardware and software prototypes for future data centers, questioning conventional approaches to server and infrastructure design [18]. More generally, the deployment of very large-scale computing systems, which target science and defense problems of critical national interest, is limited by system reliability and energy consumption. More reliable, energy efficient, high-performance computing (HPC) systems open access to advanced computing to individuals, groups, academic institutions and national laboratories.

5. Research Summary

The research computing landscape – hardware, software, applications, and expectations – continues to evolve rapidly. Concurrently, new expectations for data management and retention, combined with resource constraints, are putting new pressures on research agencies and academic institutions to find cost-effective approaches that maximize scientific benefit while minimizing costs. There is no single, simple solution to these challenges, as they depend on institutional circumstances, research domains, user characteristics, scientific and engineering expectations, and a complex set of policy issues. However, by sharing data on evolving cyberinfrastructure needs and implementation costs, we can develop comparative models that allow funding agencies, institutional resource providers, and individual users to evaluate the costs and benefits of specific choices and subsidies.

This project proposes to build, publish, and regularly update comparisons of the evolving TCO of on-premise HPC clusters and data storage systems relative to commercial cloud services such as Amazon AWS, Microsoft Azure, and Google Cloud Platform. Concretely, the team proposes to build and host a Jupyter notebook that enables the cyberinfrastructure community to add, modify, and explore total cost of ownership (TCO) models based on a variety of usage patterns and performance expectations. By pooling data and insights, the community can both help shape system configuration and pricing choices and facilitate national policy discussions about sustainable cyberinfrastructure.

References Cited

1. Dongarra, J.J., P. Luszczek, and A. Petitet, *The LINPACK Benchmark: Past, Present and Future*. concurrency and Computation: Practice and Experience, 2003. **15**(9): p. 803--820.
2. Heroux, M., J. Dongarra, and P. Luszczek. *HPCG Benchmark*. 2017; Available from: <http://www.hpcg-benchmark.org/index.html>.
3. (NSF), N.S.F. *Extreme Science and Engineering Discovery Environment (XSEDE)*. 2013; Available from: <https://www.xsede.org/allocation-policies>.
4. Thompson, C.S., S.M. Clark, and X.C. Song, *The XSEDE BLAST Gateway: Leveraging Campus Development for the XSEDE Community*, in *Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*. 2016, ACM: Miami, USA. p. 1-8.
5. XSEDE. *Science Gateways Listing*. 2017; Available from: <https://www.xsede.org/web/guest/gateways-listing>.
6. Fulton, B., et al., *XDMoD Value Analytics: A Tool for Measuring the Financial and Intellectual ROI of Your Campus Cyberinfrastructure Facilities*, in *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*. 2017, ACM: New Orleans, LA, USA. p. 1-7.
7. Milligan, M., *Interactive HPC Gateways with Jupyter and Jupyterhub*, in *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*. 2017, ACM: New Orleans, LA, USA. p. 1-4.
8. Shen, H., *Interactive Notebooks: Sharing the Code*. Nature, 2014. **515**(7524).
9. Dongarra, J.J., *The LINPACK Benchmark: An Explanation*, in *Proceedings of the 1st International Conference on Supercomputing*. 1988, Springer-Verlag New York, Inc.: Athens, Greece. p. 456-474.
10. *University of Iowa Neon Cluster Overview*. 2017; Available from: <https://wiki.uiowa.edu/display/hpcdocs/Neon+Overview+and+Quick+Start+Guide#NeonOverview+andQuickStartGuide-SystemOverview>.
11. Ferdman, M., et al., *Clearing the Clouds: A Study of Emerging Scale-out Workloads on Modern Hardware*, in *Proceedings of the Seventeenth International Conference on Architectural Support for Programming Languages and Operating Systems*. 2012, ACM: London, England, UK. p. 37-48.
12. Lausanne, É.P.F.d. *CloudSuite*. 2017; Available from: <http://cloudsuite.ch/>.
13. Google. *PerfKit Benchmark*. 2017; Available from: <http://googlecloudplatform.github.io/PerfKitBenchmark/>.
14. Barga, R., D. Gannon, and D. Reed, *The Client and the Cloud: Democratizing Research Computing*. IEEE Internet Computing, 2011. **15**(1): p. 72-75.
15. Bernat, A. and D. Reed, *NSF Selects CRA to Create Computing Community Consortium*, in *Computing Research News*. 2006, Computing Research Association.
16. Reed, D.A. and J. Dongarra, *Exascale Computing and Big Data*. Communications of the ACM, 2015. **58**(7): p. 56-68.
17. Geist, A. and D.A. Reed, *A Survey of HPC Scaling Challenges*. International Journal of High-Performance Computing Applications, 2015. **31**(1): p. 104-113.
18. Reed, D.A. and J.R. Larus, *Imaging the Future: Thoughts on Computing*. IEEE Computer, 2012. **45**(1): p. 25-30.
19. NSF Workload Analysis on Blue Waters, Final Report, ACI 1650758, https://drive.google.com/file/d/0B_EbrAJu57PsQi1rV0xPU1kwT1U/view, 2017