

Usage of Indiana University computation and data cyberinfrastructure in FY 2011/2012 and assessment of future needs

*Matthew R. Link
David Y. Hancock
Kurt Seiffert
Stephen Simms
Scott Michael
Craig A. Stewart*

Indiana University
PTI Technical Report PTI-TR13-004
2013

Citation:

Link, M.R., Hancock, D.Y., Seiffert, K., Simms, Stephen, Michael, S., and Stewart, Craig A. "Usage of Indiana University computation and data cyberinfrastructure in FY 2011/2012 and assessment of future needs," Indiana University, Bloomington, IN. PTI Technical Report PTI-TR13-004, Jun 2013.
<http://hdl.handle.net/2022/16604>



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services
Pervasive Technology Institute

- This research was supported in part by:
 - The Pervasive Technology Institute, Indiana Metabolomics and Cytomics Initiative, and the Indiana Genomics Initiative. All of these initiatives have been supported in part by Lilly Endowment, Inc.
 - Grant number 1U24AA014818-01 from NIAAA/NIH. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIAAA/NIH.
 - National Science Foundation under Grants CDA-9601632, EIA-0116050, ACI-0338618, OCI-0451237, OCI-0535258, and OCI-0504075, CNS-0723054, and CNS-0521433.
 - Shared University Research grants from IBM, Inc. to Indiana University.
- Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies represented above.

Table of Contents

1. Executive summary	1
2. Introduction	1
3. History and impact of IU strategic planning for information technology.....	2
4. Indiana University's cyberinfrastructure - 2012.....	7
4.1. Physical facilities.....	7
4.2. Overall structure and support of IU's advanced research cyberinfrastructure.....	7
4.2.1. High performance computing (HPC) systems.....	7
4.2.2. Data storage systems	9
5. Recent patterns of IU cyberinfrastructure use	11
5.1. Use of IU HPC systems	11
5.2. Use of IU research storage systems	14
5.2.1. "Scratch" and short-term file system usage	14
5.2.2. Home directories.....	15
6. Satisfaction with support for research computation and data storage cyberinfrastructure....	16
7. Key benefits to IU as a result of IU's advanced cyberinfrastructure.....	16
8. Role of local vs. national and commercial facilities.....	19
8.1. Cloud computing as compared to high throughput, loosely coupled clusters.....	19
8.2. Tightly coupled, massively parallel computing systems – local, national, and experimental computing systems.....	20
9. Analysis of current and future researcher needs in computation and storage	21
9.1. Survey	21
9.2. Interviews.....	22
9.2.1. Common themes from interviews regarding computational systems	22
9.2.2. Specific and specialized needs	23
9.2.3. Results from interviews regarding user needs for file systems	23
10. Appendix – results of 2012 survey of computational and data usage patterns and needs..	25

Table of Tables

Table 1. Summary of physical facilities at Indiana University	7
Table 2. Summary of computational resources at Indiana University	9
Table 3. Summary of data storage resources available at Indiana University.	10
Table 4. Key usage metrics for IU cyberinfrastructure systems for calendar 2011.....	11
Table 5. Characteristics (campus, department, teraflops-hours used, % of total system hours) of the 25 individual top users of Big Red during CY2011.....	14

Table of Figures

Figure 1. Growth in total CPU hours utilized on Big Red and Quarry over time.....	11
Figure 2. Quarry usage for calendar year 2010 and 2011.....	12
Figure 3. Big Red usage for calendar year 2010 and 2011.....	12
Figure 4. Usage of Big Red and Quarry by group for CY2011 for top groups.	13
Figure 6. Data Capacitor and Data Capacitor-Wide Area Network (DC-WAN) usage in the first half of calendar year 2012. Data points represent the % of capacity of disk space used at peak during each month.	15

1. Executive summary

This report details the past and current cyberinfrastructure resources that have been deployed by the Research Technologies (RT) division of University Information Technologies Services to support research and scholarly activities at IU. This report also presents data and detailed analysis of system usage and services supported by RT for the FY 2011/2012 period, projects future usage trends based on these data, and provides several recommendations for the most effective ways to meet the growing need for high performance computing resources in research and scholarly endeavors.

The major findings presented in this report can be divided into three major categories. First, IU has been, and continues to be, at the forefront of innovation in high performance computing in both basic research and operational deployment. This is a result of the incredibly productive synergy between RT and the other Pervasive Technology Institute (PTI) centers, which allows the PTI Research Centers to focus on basic research and development and the PTI Service Centers to focus on production-quality operational deployments that serve the entire university. The leadership that IU has shown has resulted in a competitive edge for the university in the areas of research and scholarly endeavors as well as in securing grant funding.

Secondly, current demand for computational and storage resources from IU scholars meets or exceeds RT's existing supply. All of the systems and services that IU provides to the university are operating at or near 100% capacity and researchers say, both anecdotally and in formal surveys, that additional resources would enable them to conduct projects that are even more ambitious than the ones in which they are currently engaged. In addition, we find that the usage scenarios for the systems supported by RT run the gamut of possibilities from a large group of users running many jobs requiring a high throughput environment to, at the other end of the spectrum, another set of users running relatively few jobs where each job requires a large amount of resources.

Finally, our analysis suggests that, for the time being, the most effective way to support the breadth of uses cases represented by IU researchers is to continue to provide and support high performance computing resources, both loosely coupled clusters and tightly coupled parallel computing systems, as a local resource for IU researchers and scholars. Though cloud computing and national cyberinfrastructure resources like the Extreme Science and Engineering Discovery Environment (XSEDE) can be beneficial for a number of use cases, there are many that can be supported only by local resources or where local resources are the best option, both operationally and economically. Even in the cases when cloud or grid computing are viable options, in many cases support staff from RT are needed to aid researchers in getting up and running on those resources and in porting their applications to work in the new environment.

2. Introduction

Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.¹ This report summarizes the current status and patterns of usage of IU's advanced research cyberinfrastructure; describes some of the benefits that IU enjoys as a result of this investment; and describes analysis of current and future needs felt by the IU research community. This report details activities from the 2011/2012 fiscal year in particular (1 July 2011 to 30 June 2012), but also includes some information on earlier years. This report focuses on the research cyberinfrastructure delivered by the

¹ Stewart, C.A., S. Simms, B. Plale, M. Link, D. Hancock and G. Fox. What is Cyberinfrastructure? In: *Proceedings of SIGUCCS 2010*. (Norfolk, VA, 24-27 Oct, 2010). <http://portal.acm.org/citation.cfm?doid=1878335.1878347>

Research Technologies division of University Information Technology Services, which is affiliated with the IU Pervasive Technology Institute² as a Cyberinfrastructure and Service Center. The mission of the Research Technologies division is “to develop, deliver, and support advanced technology solutions that improve productivity of and enable new possibilities in research, scholarly endeavors, and creative activity at Indiana University and beyond; and to complement this with education and technology translation activities to improve the quality of life of people in Indiana, the nation, and the world.” This mission fits well into the mission of the IU Pervasive Technology Institute, which is “to improve the quality of life in the state of Indiana and the world through novel research and innovation and service delivery in the broad domain of information technology and informatics. As a world-class organization, PTI pairs fundamental academic computational research with the widely known strengths of Indiana University through innovations and service delivery in networking and high performance computing. By means of organization into research and service centers, PTI encourages collaboration that crosses center boundaries, where practice informs the science, and science advances the practice, the results of which advance the university, state, and nation as a whole.”

3. History and impact of IU strategic planning for information technology

Over the past fifteen years, IU has made significant investments in its high performance computing resources and related cyberinfrastructure. These periodical acquisitions have brought substantial return to IU’s research mission and external funding. Some highlights include:

- 1997 – IU acquires a CAVE (CAVE Automatic Visualization Environment) immersive 3D room-scale virtual reality environment
- 1997 – Acquisition of an SGI Origin2000 (the first university-owned Origin2000)
- 1999 – IU joins High Performance Storage Systems (HPSS) Consortium and implements a tape-based massive data storage system based on use of HPSS with a tape library located at IU Bloomington
- 2000 – Replacement of the SGI Origin2000 with a Sun E10000
- 2000 – First deployment of geographically distributed High Performance Storage System (HPSS) tape archive system, allowing data replication between IU tape storage silos in Bloomington and Indianapolis. Greatly enhanced the security of data storage for IU and was made possible by code written and contributed to HPSS by IU staff members
- 2001 – The IU IBM Research SP is the first university-owned supercomputer capable of 1 teraflops (debuted in 50th place on Top500 list in November of 2001)
- 2003 – The AVIDD Linux cluster is the first distributed cluster to exceed 1 teraflops achieved performance in the HP Linpack benchmark suite
- 2003 – IU develops and commercially licenses the John-e-Box (lab-scale stereoscopic visualization system) and uses National Science Foundation funds to deploy nine systems to labs and studios across IUB, IUPUI, IUE, and IUN
- 2004/05 – IU deploys its first ultra-resolution display wall and reconfigurable VR Theater at IUPUI, recognizing the complementary demands of ultra-high resolution and stereoscopic immersion
- 2005 – IU installs BARCO 3D MoVE Lite Virtual Reality Theatre at IUPUI; highest resolution VR system in operation in the US

² pti.iu.edu

- 2006 – IU deploys Data Capacitor as 535 terabyte (TB) file storage system
- 2006 – IU's Big Red debuts in 23rd place on June 2006 Top500 list – the fastest academic supercomputer in the western hemisphere
- 2009 – IU develops IQ-wall design for lower-cost, easy-to-use ultra-resolution display walls
- 2010 – Joint federal and IU funding provides for Visualization and Collaboration Theater in IU Innovation Center at IUB; merges immersive visualization and virtual reality capabilities with high-end videoconferencing, collaboration, and presentation space
- 2011 – Data Capacitor is upgraded to more than 1 petabyte (PB)
- 2011 – 10' x 30' 3D display installed in Innovation Center; 24' x 9' 2D tiled display installed in Cyberinfrastructure Building (CIB)
- 2012 – AVL installs new mid-scale (3' x 5' up to 10' x 10') visualization systems and IQ interactive systems in School of Informatics and Computing, CIB, and Innovation Center, bringing the total number of such mid-scale facilities installed at IU to 10

IU has invested in other areas of computation such as large memory systems (Mason), virtual machines and high throughput systems (Quarry), and experimental hardware (including cell processors). These investments have consistently positioned IU as one of the top universities in the nation in terms of advanced cyberinfrastructure. Externally, IU is perceived as a high-quality and highly effective advanced computing center – more like a national supercomputing center or government lab than a university. IU has managed this success in external perception and competition for grant awards while still focusing advanced and innovative IT services towards the scholarly and creative communities so the IU community feels it is well served by IU's investment in advanced IT. This allows IU faculty to engage in cutting edge research while simultaneously improving IU's overall intellectual and artistic leadership. IU's advanced cyberinfrastructure facilities create new capabilities in research and artistic creation, aiding faculty recruitment and enhancing grant competitiveness. A recent analysis showed that the pattern of investment IU had made for decades – periodic investment in systems appearing on the Top500 list – boosts research productivity and funding success³. IU computer scientists and information technologists can develop new instruments to perform computer-based experiments and analyses, putting them among the first to perform massive simulations or make new discoveries. This meets a goal set in 1998 by then Vice President McRobbie at the time the first IU Information Technology Strategic Plan was created. IU has persisted in keeping information technology as a strategic asset, recognizing the interplay between research capabilities and productivity and innovation in information technology. We have held steady in this strategy, and it is interesting to see General Motors moving IT back into its own core operations after concluding that outsourcing IT was holding back its ability to innovate.⁴

The depth and breadth of use of IU's advanced cyberinfrastructure is such that more than 30% of the IU research community (faculty, research staff, and graduate students) use some service provided by the Research Technologies (RT) division of UITs. More members of the IU research community make active use of IU's advanced cyberinfrastructure (Big Red, Quarry, Data Capacitor) than the total number of people who have accounts on the Extreme Science and Engineering Discovery Environment (XSEDE)⁵.

The maturation of the Pervasive Technology Institute (PTI) and the progressively deeper involvement of Research Technologies in PTI have been critical in the development, deployment, and utility of the

³ Apon, A., S.C. Ahalt, V. Dantuluri, C. Gurdgiev, M. Limayem, L. Ngo and M. Stealy. "High Performance Computing Instrumentation and Research Productivity in U.S. Universities". *Journal of Information Technology Impact*, 10(2), 87-98. 2010. Available from: <http://www.jiti.net/v10/jiti.v10n2.087-098.pdf>

⁴ <http://www.distilinfo.com/itadvisory/2012/07/11/general-motors-will-slash-outsourcing-in-it-overhaul/>

⁵ See <http://www.indiana.edu/~uitssur/> for information on overall usage of Research Technologies services by IU community

cyberinfrastructure hardware investments made by IU. The current structure of PTI and its organization into *Research Centers* (Center for Applied Cybersecurity Research, Center for Research in Extreme Scale Technologies, Data to Insight Center, Digital Science Center) and *Cyberinfrastructure & Service Centers* (Research Technologies, National Center for Genome Analysis Support) aids effective use of IU's advanced cyberinfrastructure. This collaboration embedded into the structure of PTI – tying together research success in computer science and informatics with rapid delivery of innovative services to the general IU research community – is based on the following dynamic:

- PTI Research Centers focus on innovation and new discoveries in computer science, computational science, and informatics. Each of these centers is a world-class research group led by internationally recognized faculty and each is highly successful in obtaining federal grant funding. Research Technologies provides operational and technical support for many of the activities of the PTI Research Centers, enabling them to work at scales beyond the capabilities of many research groups.
- PTI Cyberinfrastructure & Service Centers focus on developing and delivering leading edge services that benefit the university and state, and advance the scholarly community nationally – sometimes based on innovation by RT and NCGAS staff, often aided by intellectual interactions with PTI Research Centers. RT and NCGAS, the two current PTI Service & Cyberinfrastructure Centers, serve a “translation” function – implementing new technologies developed by the PTI Research Centers in ways that make these new innovations robust and usable by the general IU research community. This gives the IU research community access to innovative computing and IT tools earlier and faster than their peers and intellectual competitors at other institutions.
- As a PTI Service & Cyberinfrastructure Center, Research Technologies delivers excellent, widely used services that are identified with both UITS and PTI.
- All Centers affiliated with PTI are engaged in education, outreach, and economic development, furthering accomplishment of IU's missions in the state, nation, and world.
- The result is a virtuous mix of innovation, application, service, and economic development that aid IU and advance accomplishment of IU's broad missions of research and creative activity, education, and service. In all cases, there are two commonalities: federal funding to IU through PTI and RT benefits the IU research community as well as the national research community; and the existence of commodity resources does not eliminate the need for system and software innovations developed, deployed, and delivered locally.

Commercial cloud computing resources and federally funded cyberinfrastructure resources are components of the overall cyberinfrastructure used by the IU community. However, IU has more than 50 years of research distinction based on its strengths in information technology in support of research and creative activity, and this distinctiveness cannot be maintained solely on the use of commercially (and commonly) available cloud facilities and federally funded supercomputers (the use of which is rationed, generally limited in scope to fields funded by the National Science Foundation (NSF) or Department of Energy (DOE), and allocated by entities outside IU).

IU's investment in advanced cyberinfrastructure has been the result of purposeful execution of strategic goals set by Indiana University and endorsed by the Trustees of Indiana University⁶. The key goals set in the 1998 Indiana University Information Plan relative to research cyberinfrastructure include the following:

⁶ McRobbie, Michael A. 1998. Information Technology Strategic Plan – Architecture for the 21st Century. <http://hdl.handle.net/2022/6823>

RECOMMENDATION 5: In support of research, UITs should provide broad support for basic collaboration technologies and begin implementing more advanced technologies. UITs should provide advanced data storage and management services to researchers. The University should continue its commitment to high performance computing and computation, so as to contribute to and benefit from initiatives to develop a national computational grid.

ACTION 27: UITs should launch an aggressive program to systematically evaluate and deploy across the University state-of-the-art tools and infrastructure that can support collaboration within the University, nationally and globally.

ACTION 28: UITs should explore and deploy advanced and experimental collaborative technologies within the University's production information technology environment, first as prototypes and then if successful, more broadly.

ACTION 29: In order to maintain its position of leadership in the constantly changing field of high performance computing, the University should plan to continuously upgrade and replace its high-performance computing facilities to keep them at a level that satisfies the increasing demand for computational power.

ACTION 30: The University needs to provide facilities and support for computationally and data-intensive research, for non-traditional areas such as the arts and humanities, as well as for the more traditional areas of scientific computation.

ACTION 31: The University should plan to evolve its high performance computing and communications infrastructure so it has the features to be compatible with and can participate in the emerging national computational grid.

ACTION 32: The University should evaluate and acquire high-capacity storage systems, capable of managing very large data volumes from research instruments, remote sensors, and other data-gathering facilities.

ACTION 33: The University through UITs should provide support for a wider range of research software including database systems, text-based and text-markup tools, scientific text processing systems, and software for statistical analysis. UITs should investigate the possibilities for enterprise-wide agreements for software acquisitions similar to the Microsoft Enterprise License Agreement.

ACTION 34: UITs should participate with faculty on major research initiatives involving information technology, where it is appropriate and of institutional advantage. Further, UITs should provide proactive encouragement and supportive services that create opportunities where faculty from diverse disciplines might come together on collaborative projects involving information technology.

These goals were the foundation of and impetus for years of successful work advancing research at IU and set the stage for a new set of goals related to research and cyberinfrastructure contained in the 2008 IU Information Technology Strategic plan – Empowering People⁷. Goals related to cyberinfrastructure in the 2008 plan include:

- *Recommendation AI. Indiana University's national and international leadership should be sustained through continued maintenance and advancement of an IT infrastructure that is supported by sound fiscal planning.*
 - *Action 4: Cyberinfrastructure. IU should continue to advance its local cyberinfrastructure, participation in national cyberinfrastructure, and its efforts to win federal funding of cyberinfrastructure programs that enhance IU's research capabilities.*

⁷ ep.iu.edu

- *Action 5: Philosophy of abundance. IU should pursue strategies that approximate a philosophy of abundance, within reason, towards unmetered availability of basic IT services, support, and infrastructure for creative activity, storage, computation, communication, and other activities fundamental to the work of the university via any appropriate sourcing strategy.*
- *Action 6: Leveraging partnerships. IU should continue its highly successful program of relationships with hardware, software, and services vendors, and seek additional partnerships and creative exchanges that provide mutual benefits.*
- *Action 7: Consolidated services. IU should maintain and refresh its IT infrastructure by consolidating enterprise-scale (multi-campus) services for software systems, server and data hosting, networks, backup, messaging, support services, and training, while also enabling innovative departmental-scale technology services provided at the edge.*
- *Recommendation A2. Indiana University should ensure that its wired and wireless campus networks continually evolve just ahead of the needs of IU's faculty, staff, and students. The network must provide secure, reliable, effective, and appropriate access to support the missions of the university.*
 - *Action 9: Network partnerships. IU should continue to pursue opportunities for strategic partnerships that can provide services for advanced networks to further the missions of the university.*
- *Recommendation A4. Indiana University should continue to practice responsible stewardship of all financial resources devoted to information technology across the university by providing transparency and accountability in support of wise decision-making.*
 - *Action 16: External funding. OVPIT should continue to lead and expand its efforts to effectively partner with academic units, campuses, administrative units, or individual investigators for external funding opportunities.*
 - *Action 25: Research into IT. IU should support and pursue research into information technology itself. IT professionals and faculty should seek partnership opportunities for scholarly publication and invention disclosure that document meritorious research and discovery. (PTI leads; RT supporting)*
- *Recommendation B8. Indiana University should implement a variety of approaches to IT education, skills acquisition, support, and communication that enable any willing learner to efficiently acquire desired IT skills.*
 - *Action 27: Human-centered support. IU should continue to pioneer and provision effective means of user support through advanced tools for self-service and connection to IU experts to help faculty, staff, and students effectively use IT. IU should continue its work as a support infrastructure provider for national research projects and services.*
- *Recommendation B9. Indiana University should provision appropriate "data utilities" for administrative data/information, research data, teaching and learning resources, and multimedia scholarly life. These utilities should provide convenient, timely, and secure access to university data/information by the IU community and authorized collaborators beyond IU.*
 - *Action 33: Research data utility. IU should provision a data utility service for research data that affords abundant near- and long-term storage, ease of use, and preservation capabilities.*
- *Recommendation B11. Indiana University should work within its missions as a public institution to deepen its technology-supported engagement with institutions and communities beyond IU that advance public health, education, research, economic development, and culture in the State of Indiana.*
 - *Action 51: Technology transfer. IU should develop its IT capabilities to support and enhance the flow of innovation from researchers and innovators to the practical use of the public and private sectors of the state of Indiana and beyond.*
- *Recommendation C15. While Indiana University should advance IT-enabled research across all disciplines, it should also focus on a few highly promising opportunities for which it has a skills, knowledge, and reputational advantage to push the frontiers of IT-enabled research and scholarship.*

- *Action 70: IT-enabled research. IU should purposefully select areas of great and timely promise for strategic development of IT-enabled research, scholarship, and/or creative activity.*

In this report, we discuss IU's current advanced cyberinfrastructure, examine patterns of use of that cyberinfrastructure in fiscal year 2011/2012 (1 July 2011 to 30 June 2012), and describe the results of surveys and interviews regarding future needs relative to computational and data-intensive cyberinfrastructure.

4. Indiana University's cyberinfrastructure - 2012

The current components of IU's computational and data-oriented cyberinfrastructure are summarized below.

4.1. Physical facilities

IU's cyberinfrastructure leverages the university's unusual arrangement of two major research campuses separated by 50 miles and connected by university-owned optical networks. This creates tremendous resilience in case of natural or man-made disaster, as well as an outstanding testbed for development of grid and distributed computing innovations. Table 1 summarizes IU's data center facilities. IU has at present a net of 500 kW of power available to support new and expanded research cyberinfrastructure.

	Machine room total ft ²	Avail. ft ²	Power total	Net power avail.	Cooling capacity total (tons)	Cooling capacity avail. (tons)
ICTC	8,300	1,400	600 kW	70 kW	290	150
IUB Data Center	30,000	15,000	2.5 MW	500 kW	2750	550

Table 1. Summary of physical facilities at Indiana University.

4.2. Overall structure and support of IU's advanced research cyberinfrastructure

4.2.1. High performance computing (HPC) systems

IU has the following production high-performance computing systems.

- *Big Red.* Big Red is an IBM e1350 distributed shared memory cluster with 4096 processor cores, 6 terabytes (TB) total memory capacity, and a peak theoretical processing capability of 40.96 teraflops.
- *Quarry.* Quarry consists of two components:
 - A computational system based on an IBM e1350 distributed shared memory cluster with 2960 processor cores, 4.9 TB total memory capacity, and a peak theoretical capability of 26.11 teraflops. The compute nodes consist of 140 HS21 Blade servers and 230 dx360 iDataPlex nodes, each with two quad-core Intel Xeon processors and 8-16 GB of memory. The cluster includes 42 TB of local spinning disk and is attached to the Data Capacitor for high performance storage.
 - *Quarry Gateway Hosting system.* The Quarry Gateway Web Services Hosting resource at Indiana University consists of multiple Intel-based HP systems geographically distributed for failover in Indianapolis and Bloomington. Currently there are four HP DL160 front-end systems at each site all configured with dual quad-core Intel E5603 processors, 24 gigabytes (GB) of random access memory (RAM), and a 10 gigabit Ethernet adapter. The front-end systems host the kernel-based virtual machines (KVMs). Virtual machine (VM) block storage is provided by two HP DL180 servers at each site configured with a quad-core Intel X5606 processor, 12 GB of RAM, a 10 gigabit Ethernet adapter, and a RAID controller attached to an HP storage array. A standard VM consists of one virtual CPU, 4 GB of memory, and 10 GB of persistent local storage. Service owners are granted root access to their virtual

machine. The Data Capacitor wide area network (WAN) file system can also be mounted for larger project and scratch space. The host operating system is CentOS 6. The supported virtual machine operating systems are Red Hat Enterprise Linux, CentOS, Ubuntu, and Debian Linux.

- *Mason*. Mason is an HP distributed shared memory cluster with 512 processor cores, 8 TB total memory capacity, and a peak theoretical capability of 3 teraflops. The compute nodes consist of 16 DL580 G7 servers, each with four eight-core Intel Xeon L7555 processors, 512 GB of memory, and a PCIe 10Gb Ethernet adapter for high-bandwidth data transfer. The cluster includes 16 TB of local spinning disk.
- *The IU portion of the US ATLAS Midwest Tier 2 Center* (<http://mwt2.usatlasfacility.org/>). The IU portion of the MWT2 facility is a heterogeneous cluster of 20 Dell 1950 servers, 56 Dell R410 servers, and 80 white-box servers, connected by a 1.0 Gbps network. This heterogeneous cluster has a total of 1312 processor cores, 4.0 TB total memory capacity, and a peak theoretical capability of 13.6 teraflops. The Dell and HP compute nodes include a mix of 4-core Quad Core Xeon E5440 Processors and 6-core Intel Xeon CPU X5660 processors, with between 2 and 4 GB of memory per core. The white-box servers include a mix of Dual and Quad-Core AMD Opteron processors.
- *Research Database Complex*. The Research Database Complex (RDC) is dedicated to research-related Oracle databases and data-intensive applications that require an Oracle database. The RDC also provides an environment for database-driven web applications with a research focus. The RDC consists of 4 HP DL160 servers, each with dual Intel E5620 processors, two 72 GB SAS disks, and 72 GB of memory. The web environment is a Dell 2950 with a Quad-core Intel Xeon processor and 8 GB of memory. The RDC has 72 TB of SAN-attached storage for database hosting.
- *FutureGrid systems*. Indiana University received a major grant award from the National Science Foundation to implement an experimental, high-performance grid test-bed called FutureGrid.
 - The *IBM iDataPlex* system (India) is an IBM e1350 distributed shared memory cluster with 1024 processor cores and 3 TB total memory capacity. The compute nodes consist of 128 dx360 M2 iDataPlex servers, each with two quad-core Intel Xeon processors, 24 GB of memory, and a PCIe Mellanox ConnectX 4x DDR InfiniBand adapter for high bandwidth, low-latency MPI applications.
 - The *Cray XT5m* (Xray) is a distributed shared memory cluster with 672 processor cores and 1.3 TB total memory capacity. The compute blades consist of 21 XT5 blades, each with eight quad-core AMD Shanghai processors, 64 GB of memory, and an integrated Cray SeaStar adapter for high bandwidth, low-latency MPI applications.
 - The *large-memory HP cluster* (Bravo) is an HP distributed shared memory cluster with 192 processor cores and 3 TB total memory capacity. The compute nodes consist of 16 HP DL180 servers, each with two quad-core Intel Xeon E5620 processors, 192 GB of memory, 12 TB of local attached storage, and a PCIe 4x QDR InfiniBand adapter for high bandwidth, low-latency MPI applications.
 - International Computer Concepts (ICC) supplied the InfiniBand-connected 16-node *Delta* machine, comprising two Intel X5660 CPUs with 6 cores and 12 threads on each node together with two NVIDIA C2070 Tesla GPUs, 192 GB of memory, and 12 TB of local disk storage.
 - *Echo* is also an ICC cluster with 16 InfiniBand-connected nodes with 384 GB memory and two Intel E5-2640 CPUs with 6 cores and 12 threads each. This cluster supports the ScaleMP

distributed shared memory architecture with up to 5 TB of memory per job. Each node of Echo has 12 TB of local disk storage.

Name	Architecture	Teraflops	Total RAM (TB)	Local disk (TB)
Quarry	IBM e1350 Intel Xeon (HS21 blades)	26.1	4.9	42
Mason	HP DL580 G7 Intel Xeon servers	3.8	8.0	16
RDC	HP DL160 database servers, Dell 2950 Web server	N/A **	0.3	72
India	IBM e1350 dx360 M2 iDataPlex	12.8	6.0	384
Xray	Cray XT5m	6.7	1.3	5.5
Bravo	HP Cluster (Large memory, Large disk)	1.2	3.0	192
Delta	ICC Cluster (Large memory, Large disk, GPU)	10.8	3.0	192
Echo	ICC Cluster (Very Large memory per node, ScaleMP between nodes, Large disk)	1.8	6.0	192
Totals		64	33.1	1196

Table 2. Summary of computational resources at Indiana University.

**** Not used for calculations.**

4.2.2. Data storage systems

In addition to the locally attached storage listed above, IU has three major disk-based file systems and one archival storage system that serve local and remote users:

- The Research File System.* The IU Research File System currently has a capacity of 80 TB (60 TB usable) and allows for group collaboration via file sharing. Users have a highly flexible system for granting access to files, and the underlying OpenAFS technology used for the system can enable users at multiple institutions to share files. Researchers can request dedicated project space for each project requiring dedicated storage and collaboration. Project space quotas start at 100 GB and can be increased upon request. Users can access files from their desktops using the Common Internet File System (CIFS), via the web, and via secure file transfer protocol (SFTP). In the first half of 2013, the RFS system will be upgraded to a new DataDirect Networks (DDN) solution using IBM's General Parallel File System (GPFS) that will provide, on two IU campuses, 420 TB (336 TB usable) of storage that will be synchronously mirrored for disaster recovery and availability. The new environment will also provide home directory space for IU's HPC systems. It will use GPFS as the underlying file system with the same interfaces researchers use today.
- The Data Capacitor (DC) and The Data Capacitor Wide Area Network (DC-WAN)* file systems are high speed/high bandwidth Lustre storage systems for research computing that serve all IU campuses and other sites throughout the country. While the DC serves local, national, and international users of HPC resources local to IU, DC-WAN does this while also serving HPC resources at other institutions primarily by wide area network (remote) Lustre file system mounts. DC has a total formatted capacity of 1.1 petabytes (PB) and DC-WAN has a total formatted capacity of 339 TB. Both DC and DC-WAN consist of six Dell servers running the Lustre file system configured with four servers used for object storage in failover pairs equipped with 10-gigabit Ethernet cards, and two used for Lustre metadata in an active-passive configuration that use Gigabit Ethernet. DC functions primarily to provide ephemeral scratch space for users while providing some mid-term storage for projects. For DC-WAN the reverse is true as the bulk of DC-WAN users have more project-centric data requirements. The most important feature of DC-WAN that distinguishes it from DC is that it provides the capability of mapping remote users to local users, spanning administrative domains and allowing machines with heterogeneous namespaces to communicate seamlessly. DC-WAN currently serves a legacy role within XSEDE for 40 XSEDE project allocations that require DC-WAN's capabilities for file storage and long distance accessibility, and have not found any of the current XSEDE services suitable to meet

their needs. IU provides wide area file system connections for more than 10 collaborators and facilities.

- *IU's Scholarly Data Archive (SDA)*. SDA uses High Performance Storage System (HPSS) software to make available to IU researchers a total storage capacity exceeding 15 PB. Data are written to a fast, front-end disk cache and migrated over time to IBM TS3500 tape libraries on the Indianapolis and Bloomington campuses. Data written to IU's HPSS system are copied simultaneously to both locations, providing highly reliable disaster protection. Users can access data over the network from central research systems or from personal workstations, using SFTP, pftp_client, HSI/Htar, CIFS, and HTTP. The default allowance is 50 TB of mirrored data, with additional space provided upon request. SDA stores and provides access to data for the IUScholarWorks Repository⁸, a document and data archiving system created using DSpace software.

Name	File system	Disk PB unformatted	Disk (PB) usable (formatted)	Tape (PB)
Research File System	OpenAFS / GPFS	0.42	0.37	NA
DC	Lustre	1.5	1.1	NA
DC-WAN	Lustre	0.48	0.34	NA
Scholarly Data Archive	HPSS	0.80	0.60	15
Totals		3.20	2.41	15

Table 3. Summary of data storage resources available at Indiana University.

⁸ <http://scholarworks.iu.edu>

5. Recent patterns of IU cyberinfrastructure use

5.1. Use of IU HPC systems

Figure 1 shows growth in total annual use of CPU hours on Big Red and Quarry from 2006 to the present. At present there is significant contention for access to processors on both Big Red and Quarry, and particularly strong contention for processor nodes in Quarry. IU's current computational systems are not sufficient to support current demand for simulations and data analyses.

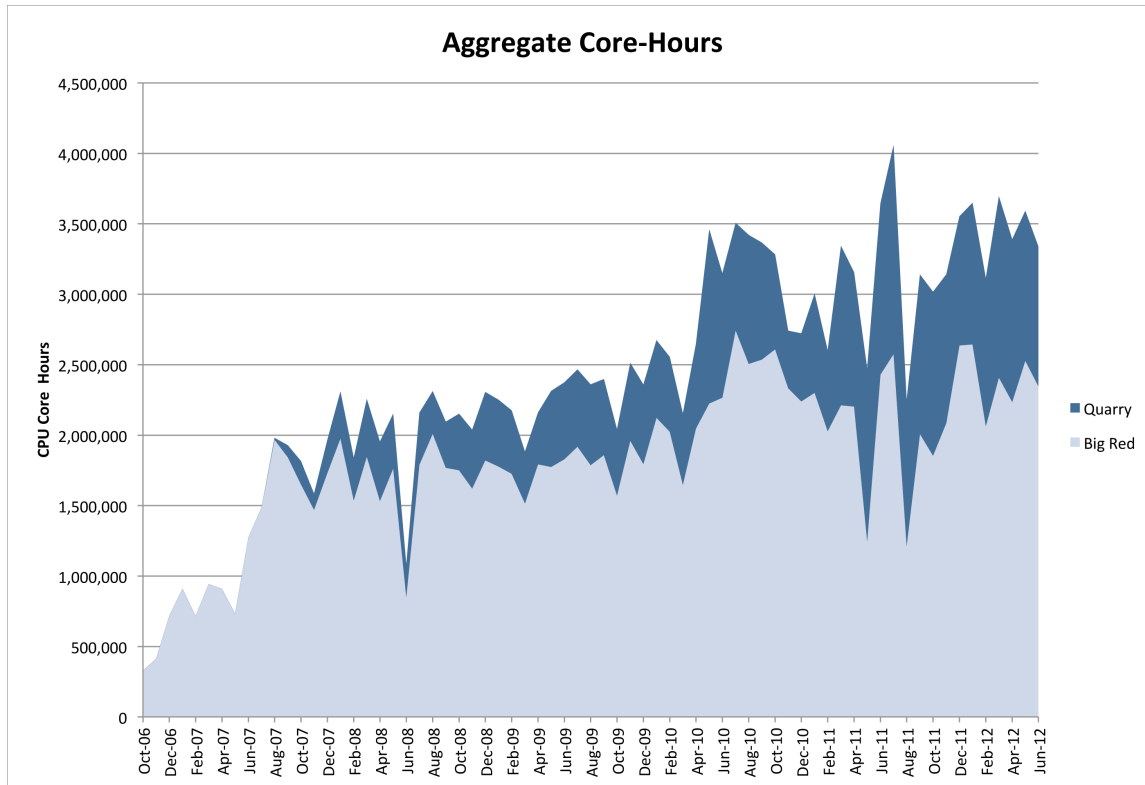


Figure 1. Growth in total CPU hours utilized on Big Red and Quarry over time.

Detailed usage metrics from IU's cyberinfrastructure for calendar 2011 are as follows:

System	# Users	Usage metrics
Quarry - High Throughput Usage	6,134	12.7M core hours
Big Red - Tightly Coupled	5,880	29.4M core hours

Table 4. Key usage metrics for IU cyberinfrastructure systems for calendar 2011.

The pattern of Quarry and Big Red usage – in terms of the degree of parallelism in jobs used on the system in 2010 and 2011 – is shown in Figure 2 and Figure 3. In calendar 2011 the large majority of Quarry usage was restricted to a single node of Quarry (fewer than eight cores). In calendar 2011 the large majority of Big Red usage involved 64 cores or more. As a result, jobs of 64 or more cores used the majority of the teraflops-hours delivered by Big Red and Quarry combined.

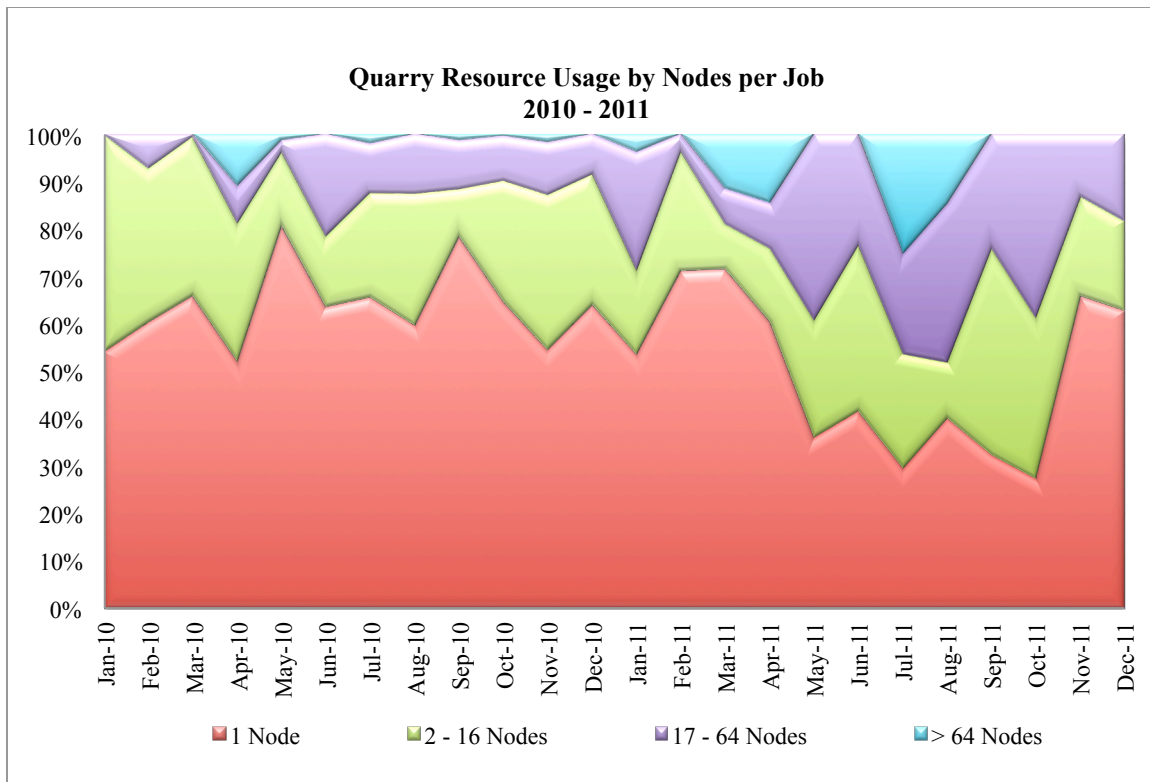


Figure 2. Quarry usage for calendar year 2010 and 2011.

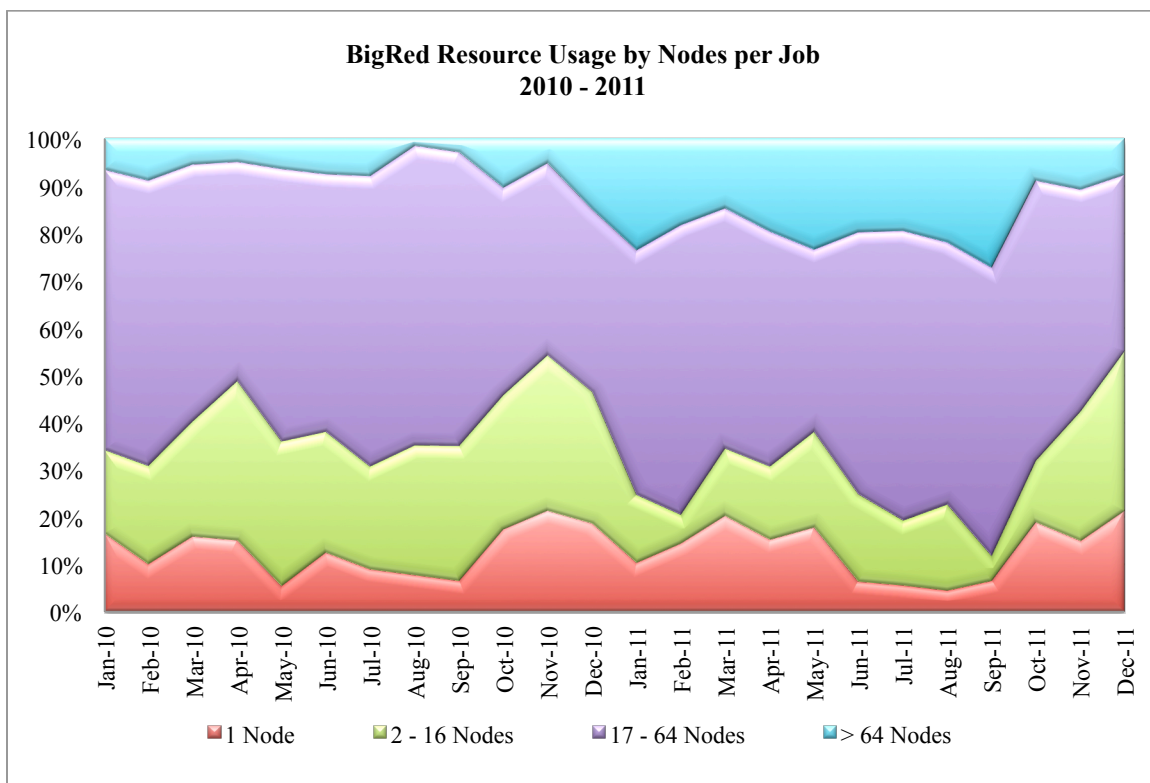


Figure 3. Big Red usage for calendar year 2010 and 2011.

It is instructive to look at the diversity of departments that use Big Red and Quarry.

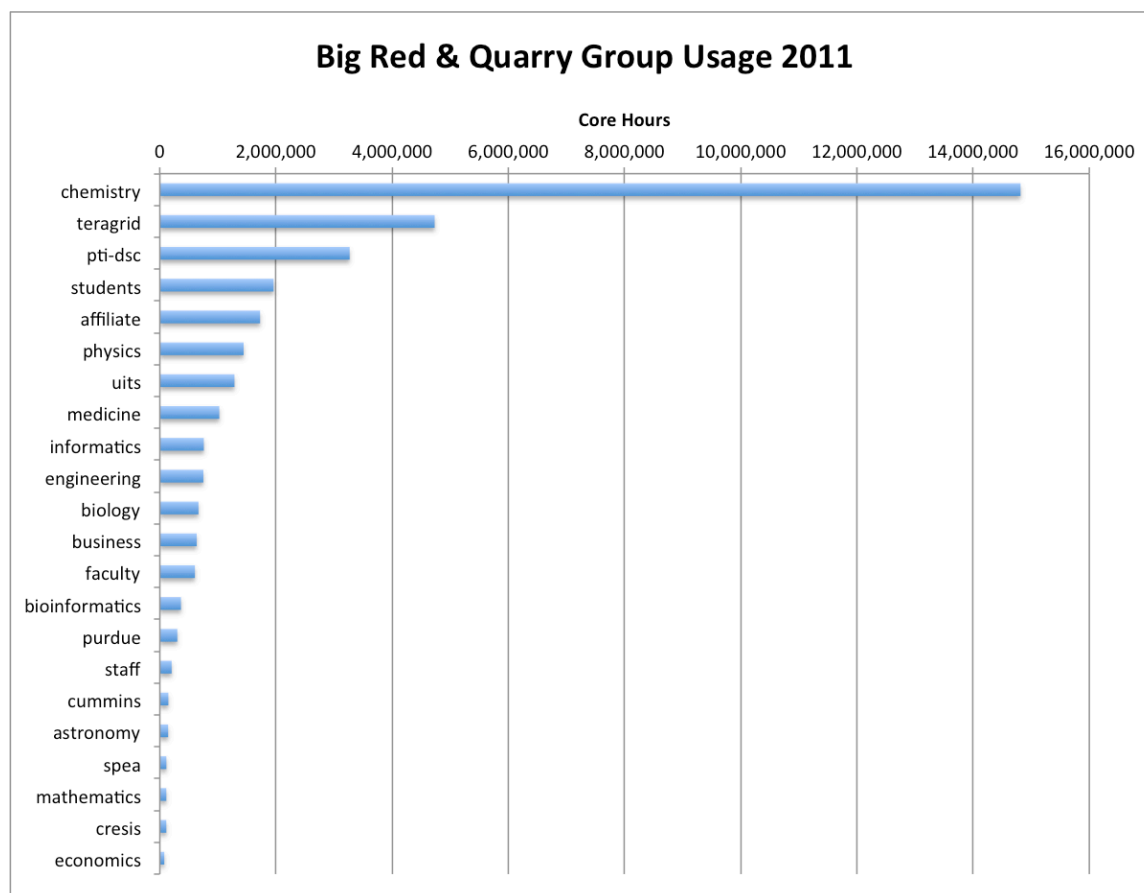


Figure 4. Usage of Big Red and Quarry by group for CY2011 for top groups.

There are also many extremely capable and sophisticated researchers who use highly scalable parallel applications. The discipline and number of CPU hours used on Big Red of the top 25 individual users of 2011 are shown in Table 5. These researchers are all limited in their current research activities by the processing capability of IU's parallel computing systems.

User Rank (largest user of Big Red first)	Department	Teraflops- Hours	% of total system utilization
1	IUB-CHEM	35,265	12.88%
2	IUB-CHEM	32,519	11.87%
3	IUB-CHEM	23,900	8.73%
4	TERAGRID	16,891	6.17%
5	UA-VPIT	10,314	3.77%
6	IUPUI-BIOM	10,096	3.69%
7	IUPUI-CHEM	9,888	3.61%
8	TERAGRID	8,292	3.03%
9	TERAGRID	7,380	2.69%
10	TERAGRID	6,073	2.22%
11	IUPUI-ENGT	5,618	2.05%
12	IUB-CHEM	4,912	1.79%
13	IUPUI-CHEM	4,682	1.71%
14	IUB-CHEM	4,170	1.52%
15	IUB-PHYS	4,070	1.49%
16	IUB-BUS	4,028	1.47%
17	IUPUI-BIOM	3,384	1.24%
18	IUB-CHEM	3,337	1.22%
19	IUB-CEEM	3,264	1.19%
20	IUB-PHYS	3,010	1.10%
21	IUPUI-PHTX	2,747	1.00%
22	IUPUI-ENGT	2,468	0.90%
23	IUB-CHEM	2,432	0.89%
24	IUB-CHEM	2,372	0.87%
25	IUB-CHEM	2,149	0.78%

Table 5. Characteristics (campus, department, teraflops-hours used, % of total system hours) of the 25 individual top users of Big Red during CY2011.

Legend: BL-BUS = IU Bloomington, Kelley School of Business; IUB-CEEM = Center for Exploration of Energy and Matter; IUB-Chem = Chemistry at IU Bloomington; IUB-Physics = IU Bloomington, Physics dept.; IUPUI-BIOM = IUPUI, Biomedical research (bioinformatics related); IUPUI-ENGT = IUPUI, Engineering & Technology; IUPUI-PHTX = Pharmacology and Toxicology (School of Medicine); TERAGRID = Non-IU users, accessing IU systems via allocations through the NSF-funded TeraGrid project and associated grant awards to IU; UA-VPIT = all OVPIT affiliated accounts, IT and informatics research.

5.2. Use of IU research storage systems

5.2.1. “Scratch” and short-term file system usage

In 2005, IU’s proposal to the NSF for a facility to be called Data Capacitor was forward thinking in its approach to the problems of storing and manipulating very large data sets. IU was ahead of “big data” by years. The Data Capacitor fills two important niches – providing users of IU’s computational resources with fast scratch space and high-speed project space that is allocated for fixed a set of time in order to support a specific research project or collection of projects.

Our high-speed file system resources are no longer able to adequately serve our life science faculty and graduate students. Yunlong Liu (IU School of Medicine) recently asked for 100 TB of project space on DC-WAN, which is more than 25% of the resource’s total capacity. One of Haixu Tang’s graduate students had more than 90 TB of regularly used data in her scratch directory. We see students consistently able to generate over 2 TB an hour running simulations. At that rate, a single student could fill our current Data Capacitor in a week.

The Data Capacitor is at or beyond its practical capacity as a high-speed file system for computation and “big data” analyses. We have – so far on rare occasions – been forced to terminate simulations running on Big Red because otherwise we would run out of physical storage space on the Data Capacitor, creating a catastrophic failure of that system. (This is because simulations were creating data on the Data Capacitor faster than it was possible to move data off the Data Capacitor to tape storage). Below is a graph showing file system usage, as monthly high water marks, since the beginning of calendar year 2012.

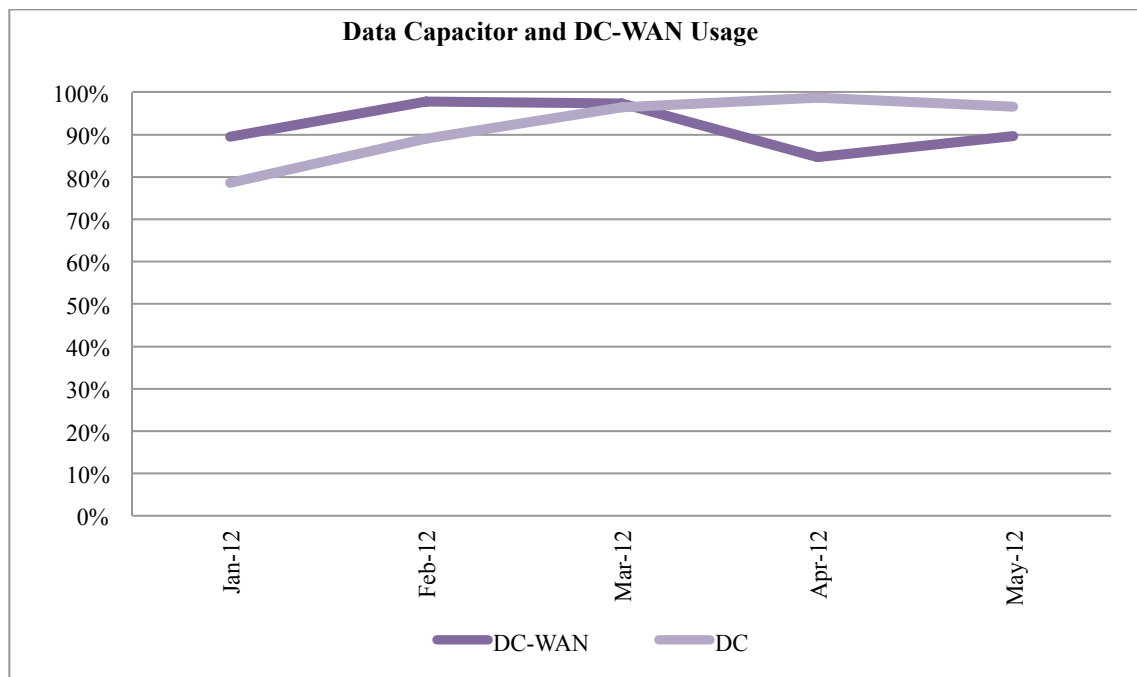


Figure 5. Data Capacitor and Data Capacitor-Wide Area Network (DC-WAN) usage in the first half of calendar year 2012. Data points represent the % of capacity of disk space used at peak during each month.

5.2.2. Home directories

Home directories are crucial in supporting IU’s computational resources, particularly with a major upgrade. The current home directory file systems support meager quotas of 10-100 GB per user. This resource provides secure, reliable, and modestly sized file systems that port medium-bandwidth, high Input/Output Operations Per Second (IOPS) workloads. Program compilation, metadata-intensive, and other interactive use is not well suited for large parallel file systems such as the Data Capacitor. Additionally, home directory file systems are of the size that centralized backups and failover for disaster recovery can be supported.

Central research home directories have seen annual growth rates of 143% per year over the last six years. There is an increased need for research file systems shared between the desktop and central resources as departments are being encouraged to consolidate. This is especially the case in medical research where HIPAA-aligned storage is required.

6. Satisfaction with support for research computation and data storage cyberinfrastructure

UITs conducts a large-scale annual survey of satisfaction of the IU community with UITs services. The details of this survey, the methodology, and results, are available online.⁹ We summarize here the results of survey questions regarding computation and data storage services offered by the Research Technologies division. The results shown are for surveys of faculty, research staff, and graduate students.

Alone, or in partnership with other campus units, UITs provides facilities and services in support of research. If you use such facilities and services, please indicate your overall satisfaction by selecting the appropriate response.

	Average	Satisfaction	Usage
Central research and high performance computers (Big Red, Quarry, and RDC clusters) [F, Staff, G]	4.135	92.75%	14.60%
Scholarly Data Archive (formerly referred to as MDSS / HPSS) (MDSS/HPSS) [F, Staff, G]	4.035	94%	12.05%
Support for software applications using IU and national high performance computer resources (including TeraGrid, Open Science Grid, and XSEDE) [F, Staff, G]	4.135	96.35%	7.65%

Overall, how satisfied are you with the UITs research technology services available at IUB/IUPUI?

Average	Satisfaction	Usage
4.205	96.3%	32.3%

Currently, Research Technologies is funded to provide 81 FTE in systems administration and support in various forms to the IU community. (Approximately 30 of these are funded by external contracts and grants). Extrapolating from the 2012 user survey data, RT serves a total of at least 4,300 individuals on the IUB and IUPUI campuses, or a ratio of 1 FTE staff per 53 users. Considering just users of Quarry and Big Red, and rounding off to an even 6,000 distinct users (a reasonable approximation), RT's staffing levels are 1 FTE staff per 74 users.

The best available numbers for comparison are from XSEDE¹⁰. There are a total of 6,636 individuals with accounts on XSEDE. The majority of the computational capability in XSEDE is at NICS (National Institute for Computational Sciences¹¹) or TACC (Texas Advanced Computing Center¹²). Adding staff at NICS and TACC to the central XSEDE staff, the ratio of staff to users is approximately 1 FTE staff per 26 users.

IU supports a much larger diversity of applications and disciplines of its users, and based on preliminary XSEDE survey results, it does so with a higher degree of satisfaction overall than XSEDE delivers.

7. Key benefits to IU as a result of IU's advanced cyberinfrastructure

There are a variety of ways to approach the value of IU's advanced cyberinfrastructure. From 2006 to present, based on award data from the Office of Research Administration, IU brought in \$4.23B in grants. Forty-one Big Red users listed as "Primary Project Directors" and 99 lab directors or advisors using or directing use of Big Red brought in \$253M of this total, including \$65.4M in facilities and administration.

⁹ <http://www.indiana.edu/~uitssur/>

¹⁰ <https://www.xsede.org/>

¹¹ <http://www.nics.tennessee.edu/>

¹² <http://www.tacc.utexas.edu/>

Another very simple metric of large-scale impact is to tally IU's grant awards of \$10,000,000 or more and analyze the role of IU's cyberinfrastructure in each. The following is a comprehensive list of all awards from federal funding agencies where a single award, or a closely linked set of consecutive awards, exceeded \$10,000,000. As it stands, three out of four such initiatives were critically dependent on IU's advanced cyberinfrastructure.

- **TeraGrid and XSEDE awards and subcontracts.** The set of awards and subcontracts related to IU's involvement in TeraGrid and XSEDE now exceeds \$10M in value. These awards were highly dependent on IU's advanced cyberinfrastructure (particularly Big Red and the Data Capacitor) and benefitted research led by the School of Informatics and Computing (SoIC) and domain scientists.
- **FutureGrid.** IU would not have received this \$10.1M award if the existing cyberinfrastructure were not in place as a foundation for FutureGrid. We were also able to demonstrate the capability of UITS staff in standing up new systems quickly to become part of FutureGrid.
- **Indiana Clinical and Translational Studies Institute (ICTSI).** This \$25M award to IU and Purdue University provides the foundation for all clinical research done by the IU health research community. UITS and IU's advanced cyberinfrastructure played a significant role in winning this award. The cyberinfrastructure developed for ICTSI has been critical to its success and will be instrumental to ICTSI as it seeks a second round of funding.
- **IU Cyclotron.** The Office of the Vice President for Research has confirmed that the set of awards that helped create and make use of the IU Cyclotron Facility totaled more than \$10M. IU's cyberinfrastructure played essentially no role in these grant awards.

A recent report¹³ outlines the usage of IU cyberinfrastructure resources and some of the key benefits from that cyberinfrastructure. A few scientifically critical projects that are made possible as a direct result of IU's cyberinfrastructure include the following.

- The **One Degree Imager (ODI)** is a 1-gigapixel camera being built by the WIYN consortium (a 501(c)(3) organization supported by the University of Wisconsin, Indiana University, Yale University, and National Optical Astronomy Observatory) that will be installed in an existing 3.5m ground-based telescope located at Kitt Peak Mountain, AZ. When it begins operation, ODI will be one of the best ground-based telescopes in existence. UITS and PTI have collaborated with the WIYN consortium to develop new tools for managing and analyzing data from the ODI. UITS and PTI are also developing a software pipeline to process raw data and create "science usable" images. If RT and PTI were unable to use our advanced cyberinfrastructure to assist WIYN with some very challenging data management issues, the ODI project might have been abandoned, and as a result the WIYN consortium would have likely collapsed altogether. Instead the WIYN consortium has been revitalized, and now IU astronomers will be able to use data from ODI through a software pipeline that was developed at IU and powered by IU's cyberinfrastructure.
- The Data to Insight Center (D2I), directed by Professor Beth Plale, has led the development of the **HathiTrust Research Center**, which supports research on digitized texts that are part of the HathiTrust repository (over 10 million digitized volumes and counting – see <http://www.hathitrust.org/about> for more information).

D2I is also one of the leading partners in an NSF-funded project called **Sustainable Environments – Actionable Data (SEAD)** that is focused on collecting and providing access to a variety of data related to sustainability science. These data will be of significant use to a number of IU scholars in sociology, political science, economics, and SPEA.

¹³ Link, Matthew R. 2012. Usage of UITS advanced research cyberinfrastructure for 2011. <http://hdl.handle.net/2022/14352>

Professor Doctor Martin Swany has joined SoIC and D2I. He is an expert in movement and management of large data sets and is currently developing new software tools that will aid researchers in the College of Arts and Sciences who need to access very large data sets. There are a number of grants currently pending that were made possible through the collaboration of Swany, the GlobalNOC, and RT which make use of IU's advanced cyberinfrastructure and RT's skilled staff.

- PTI has added a new center – the **Center for Research in Extreme Scale Technologies (CREST)**, led by SoIC Professor Andrew Lumsdaine and SoIC Professor and OVPIT Distinguished Scientist Thomas Sterling. (Professor Sterling is widely known as the “father of Beowulf clusters.”) Technology being developed by CREST will be particularly valuable in programming the largest supercomputers in the world to address challenges in genomics, physics, and astronomy. CREST has already received three critical grant awards, which put CREST and IU in the forefront of exascale computing environment development.
- UITS and PTI, in collaboration with the Department of Biology, received \$1.5M in funding from the NSF to create the **National Center for Genome Analysis Support (NCGAS)**. NCGAS will aid researchers nationally and locally with deployment and use of new applications in genome assembly and metagenomics. This builds upon years of innovations in genomics done at IU, using IU's cyberinfrastructure – including the first assembly of the *Daphnia* genome.
- **Communicating science and the value of cyberinfrastructure.** Clinical professor Albert William of the IUPUI School of Informatics has partnered with visualization experts in RT to develop a series of award-winning stereoscopic movies that highlight the impact of science and the importance of cyberinfrastructure to modern research methods. This work is critical in helping to foster interest in science, technology, engineering, and mathematics (STEM) careers in the upcoming generation, as well as to educate the general populace about the value of cyberinfrastructure investments. These movies were funded by the NSF TeraGrid and have been seen by tens of thousands of children and adults at exhibits, conferences, and online at <http://3d.iu.edu/>.
- **Fetal Alcohol Spectrum Disorders (FASD) scanning.** Researchers in the IU School of Medicine have been instrumental in the NIH-funded international Collaborative Initiative on Fetal Alcohol Spectrum Disorders, now in its 10th year. IU is funded to lead the informatics core and the 3D facial imaging core. The availability of 3D scanning equipment and programming experts in RT enabled the deployment of scanning systems to multiple sites around the world, allowing the team to study longitudinal changes of dysmorphology across different ethnic groups and levels of exposure. This same set of expertise and technology is now being used to support scanning natural history collections at IUB.

More than 14% of the Bloomington campus research community and more than 10% of the IUPUI community currently use IU's advanced cyberinfrastructure. Comparable figures are hard to come by, but our peers and competitor institutions informally report usage figures in single digit percentages. Use of IU's advanced research cyberinfrastructure is so widespread that it is no longer possible to track effectively the number of scholarly publications produced by IU researchers that make use of these resources – the number of such publications is too high.

We can effectively track some of the key financial metrics relative to IU's advanced cyberinfrastructure. Aggregate research awards to IU in support of information technology and informatics research, including PTI and its predecessor, Pervasive Technology Labs, total \$173M since 1999. Considering very narrowly grant awards that depended heavily on the IU Data Center (e.g. FutureGrid) or grants led by Research Technologies, a total of \$31M in NSF and NIH funding has been awarded competitively to IU; of this \$6.5M was facilities & administration funds.

The foundation around which all of this success has been built is IU's strategic and ongoing investment in advanced cyberinfrastructure systems, deployment, and support by world-class staff with both technical and discipline-facing expertise.

8. Role of local vs. national and commercial facilities

Advanced computing is prone to trends like any other human endeavor. Today a common cry is "just do it in the cloud." Another common refrain is "XSEDE can solve all our problems." Neither view is correct as a solution to supporting IU research, scholarly, and artistic needs, and both will remain incorrect at least for some years to come, as explained below.

8.1. Cloud computing as compared to high throughput, loosely coupled clusters

As regards cloud computing the three critical problems are performance, cost, and security. Amazon, Microsoft, and other commercial cloud providers deliver services that would generally support many of the workflows and tasks performed on Quarry. An Amazon EC2 instance equivalent to the size of one node on Quarry would cost \$0.64 per hour. The node cost is \$3,800, and at 80% utilization the cost of buying time on Amazon EC2 exceeds the cost of purchasing a node in 310 days. This cost does not include the cost of data transfer, which can be very high. As part of the review process for the SEAD grant award mentioned above we demonstrated a workflow that SEAD would support, using Quarry and the Data Capacitor. We were prepared for the first question we received after the demo: "Why not just do this in the cloud?" The answer: that demo, which represented a workflow that would be done dozens of times a day, would cost \$57,000 in data transfer fees. The high data costs, and the fact that we typically use any given node for 4-6 years at 80%+ utilization, means that the "lease vs. buy" calculation strongly favors "buy."

There are legal concerns as well. Most commercial providers will not tell you where their data centers are, so it is impossible to know what the prevailing intellectual property laws are for the territory surrounding a commercial cloud provider's data center. It is unlikely that prevailing local laws are as stringent as US law. Furthermore, in the case of Google, the mere act of storing data or information on a Google service confers to Google an eternal, no cost, worldwide license to the material stored. There is considerable risk as regards intellectual property and many of the commercial cloud providers. We note that the Internet2 NET+ service¹⁴ is an exception to this.

Last, and definitely not least, is the fact that commercial cloud providers either do not support large parallel jobs or support them with very low efficiency.

These concerns do not mean that commercial clouds are bad – it just means that for the moment there are limitations. It might well be that services where computing environments and tools that are purchased together – such as Microsoft Trident – can be obtained most effectively from a commercial provider. In addition, the Internet2 NET+ initiative may someday provide tremendous economies of scale for some of the sorts of computing done on Quarry. If and when that becomes the case, IU should certainly use such resources and support them well, so that IU's ability to use nationally shared facilities is a competitive advantage for IU researchers. In similar ways, RT and its predecessor service units have in the past supported migration of work from central supercomputers to personal workstations and – when appropriate – departmental clusters. Research Technologies is leading the way with an unusual form of outsourcing – outsourcing to Penguin Computing using systems that are located inside IU facilities. This may prove an interesting and economic model for delivery of commodity resources.

In sum, cloud computing remains an interesting and important trend. IU should experiment with its use when appropriate, particularly through continuing to lead the Penguin Computing pilot and through participating in the Internet2 NET+ service. For the next few years, however, ownership and management

¹⁴ <http://www.internet2.edu/netplus/>

of loosely coupled (Ethernet interconnect) clusters such as Quarry will remain important to supporting IU research, and will constitute the most cost effective and scientifically productive way to meet a class of computing needs that represents the largest group of IU advanced cyberinfrastructure users.

8.2. *Tightly coupled, massively parallel computing systems – local, national, and experimental computing systems*

A large-scale, tightly coupled, parallel computing resource and experimental computing systems provide advantages to IU researchers that are not available at most other institutions of higher education. For more than 15 years, the availability of such systems at IU has been a critical asset in enabling new discoveries by IU scientists in domain sciences and computer science. Collaborations between CREST, biomedical researchers, and bioinformaticians are just one current example of how IU is setting trends in ways that simultaneously benefit domain science research and computer science research. With CREST in particular we have the opportunity to position IU as a leader in new and important ways that serve the IU community. IU has always been a leader in the development and application of innovative information technology. Maintaining that advantage is a critical part of IU's efforts to differentiate itself and be one of the leading universities of the 21st century.

The critical reasons that a local resource is a crucial element of IU's advanced cyberinfrastructure, when national resources such as XSEDE are available, include the following:

- As regards production, tightly coupled, large-scale parallel systems:
 - The software environments, scheduling policies, and availability are based on reaction to local computer user needs, not policies set by a group that searches for one reasonable solution for all national parallel computing users in a facility like XSEDE. This approach, while needed for XSEDE, often results in a least common denominator approach to software and policy.
 - A local system provides extremely high bandwidth and short network paths to local data sources, data repositories, and visualization facilities. For example, we have several dedicated 10 Gbps connections from instruments on the IU Bloomington campus to the Data Capacitor, and 40 Gbps aggregate bandwidth between the Data Capacitor and Big Red.
 - Even a production quality HPC system can be challenging to use. The ability to have an in-person meeting with expert consultants, systems programmers, and systems administrators greatly facilitates efficiency in use of HPC systems, and these conversations often lead to new innovations in software and hardware configurations, and enable scientific innovations not otherwise possible.
- As regards local management of experimental systems:
 - FutureGrid provides a very flexible system for several types of experimentation, but is focused on clouds. It does not provide, for example, an FPGA-based system.
 - Certain types of experiments, including experiments involving electrical efficiency and energy use, require the ability to touch and physically manipulate the experimental systems.
 - FutureGrid does a very good job of installing operating environments within a VM and on top of a Linux environment. Experiments that require the ability to install a development version of an OS on “bare metal” and take sole control of a computing system are not well supported by FutureGrid. Sterling and Lumsdaine require these types of experiments.

XSEDE provides important resources to the national community. IU researchers consume computational capacity on XSEDE (and before it the TeraGrid) at a rate of more than \$1M in resources per year. XSEDE cannot serve as a replacement for local tightly coupled computing resources, however. Key

reasons that local resources are needed in addition to XSEDE in order to serve IU's mission and maintain IU scholarly and artistic leadership are as follows:

- IU's tightly coupled parallel resources serve types of scholarly and artistic endeavor outside of the disciplines served by and approved for use on XSEDE. Artistic rendering is the obvious example, but other areas including humanities text studies are supported by IU's tightly coupled systems and are not eligible to run on XSEDE.
- XSEDE is almost exclusively oriented toward short-term, large-scale batch processing. IU's local resources serve a variety of usage modalities, including interactive computing and very long running jobs.
- Requests for use of XSEDE exceed its aggregate capacity by a factor of two, and the current trajectory suggests that the gap between requests and resources will continue to grow over time.
- The XSEDE review process avoids double jeopardy in reviews by assuming that scientists requesting resources with a current funding award from the NSF or NIH have had their science peer reviewed and approved. Because of this, and because XSEDE capacity is not able to meet demand, requests from younger faculty that come in without a grant in hand get reviewed with particular severity, often by people who may not be current experts in the scientific areas being studied by younger researchers.
- With a local parallel system IU is able to quickly implement computer science innovations developed by the School of Informatics and Computing, delivering the practical benefits of these innovations to IU domain researchers.

A combination of a local tightly coupled resource and active encouragement of IU researchers to seek allocations on XSEDE has been and will remain a very advantageous strategy. IU scientists can start a research project right away on an IU resource, supported by IU information technologists. Such early studies help scientists be more competitive for federal funding. IU researchers benefit greatly from experience running codes (and demonstrating parallel efficiency) on IU systems. This experience and the ability to demonstrate successful work with parallel codes running on IU systems is a strong benefit to researchers applying for computing time on XSEDE. Indeed, the availability of a tightly coupled parallel computing resource acts to increase the amount of resources obtained and used by IU researchers on XSEDE.

9. Analysis of current and future researcher needs in computation and storage

This report comes as IU's main supercomputer – Big Red – is aging – at or beyond the age at which a supercomputer is typically retired.

9.1. Survey

Working with the Center for Survey Research we assembled a wide variety of questions and topics to gauge current and future needs for computational and storage resources. This survey was distributed to faculty members throughout the university, both current users of RT services and potential new customers. In conjunction with this effort we conducted several in-depth in person interviews with key faculty and researchers and their research teams on both the IUB and IUPUI campuses. This final report synthesizes the results from each approach by incorporating detailed analysis of the current utilization of our systems, the responses from the broad based user survey, and the results from the 16 in-person interviews we conducted should provide more than enough insight into the current and future computational needs of the researchers at Indiana University.

Over 280 respondents replied to the Research Technologies HPC survey, of 1000 invited to participate. The detailed findings are presented in the Appendix of this report. We present here the key findings from the responses we collected:

- A large fraction of respondents (49%) use in their research computational resources beyond what they personally own.
- Of those respondents who use some form of shared computing for their research, 90% of them indicated that they use departmental resources (e.g. a department sponsored or research group sponsored cluster).
- Of those respondents who use some form of shared computing for their research, 30% indicated that they have used Research Technologies supported system in some way.
- Of those respondents who use some form of shared computing for their research, 91% indicated that they have not used any computing resources outside IU in their research pursuits.
- Of those respondents who use some form of shared computing for their research, there was a mixture of software requirements. 92% of the respondents use some form of third party software, and 40% use software that they develop within their research group.
- Of those respondents who use some form of shared computing for their research, 3% indicated that they had a preference in processor architecture for the new machine. 100% of the respondents with a preference said they would prefer an Intel/AMD based solution.

9.2. Interviews

In addition to the survey mentioned above, we conducted several in-person interviews with leading researchers at IU. The results of these interviews are summarized below.

9.2.1. Common themes from interviews regarding computational systems

In our interviews we identified two clear groups of computational system users that mirror the patterns identified in job submission and teraflops-hours usage. One group, *capacity researchers*, can use a large loosely coupled cluster like Quarry to run a large number of capacity jobs – typically jobs that use one computing core or at most one node. This group of users is the largest in terms of number of users.

The second group, *capability researchers*, is the largest in terms of resource usage. They run large parallel computing jobs and are taking full advantage of Big Red's fast, high-bandwidth, low-latency interconnect for node-to-node communication. Capability researchers often undertake world-class scientific research programs, push ever-larger calculations, and will scale their jobs to even higher node counts in the future. Many such researchers feel limited by the capability of Big Red relative to current generation systems. Several capability researchers are using IU cyberinfrastructure as a stepping stone to national resources, and in addition, a subset of their codes are capable of using accelerators to enhance their computations.

Common themes emerging in our interviews were as follows:

- **Capacity computing and third-party ISV codes:** The majority of the groups surveyed would like a decrease in wait time for capacity jobs. This desire is in conjunction with a requirement to run standard third-party ISV (Independent Software Vendor) codes.
- **Capability computing and high performance file system support:** Researchers utilizing scaling codes and science use cases that require large core counts require high performance file system availability and support in order to run multi-node jobs effectively. There is concern that our current high performance file system environment is not adequate for this type of use due to recent capacity issues with the Data Capacitor.

- **Node memory requirements:** A large number of the interview participants expressed a need for a small number of large-memory nodes and widely held the view that the current two gigabytes of memory per core ratio should be at least maintained.
- **Application and system environment support:** Multiple researchers expressed that expert staff support was one of Research Technologies' strongest core competencies and a competitive advantage for IU. However, several expressed a desire for more expert support personnel and there were several specific requests for training to enable researchers to more effectively use, or expand their use of, IU's supercomputers.

9.2.2. *Specific and specialized needs*

In addition to the general themes, we identified several particular needs for new computing systems that stand out because of the strategic nature of the research that could be supported:

- The Center for Research in Extreme Scale Technology (CREST) has very specific needs to support the development of a next generation exascale runtime system. CREST requires a cluster that they can control both on the software and hardware side and have also outlined a strong need for field-programmable gate array (FPGA) technology. Thomas Sterling revolutionized supercomputing once; he and Andrew Lumsdaine believe that together they can do it again – by combining Lumsdaine's work in graph processing and Sterling's work in exascale runtime and operating environments. Papers and panel discussions at ISC12 support the focus on graph processing and revealed how far CREST is ahead of the field. However, CREST – and other people at IU experimenting with new application approaches – need some sort of dedicated experimental system, most likely one including FPGAs.
- Continued needs for large memory nodes. The availability of Mason with 0.5 TB of RAM per node has resulted in additional requests to support biological research and other disciplines: more nodes with 0.5 TB RAM and nodes with 1.0 TB of RAM.
- Prof. Steven Gottlieb outlined the need for a general-purpose compute resource of at least one petaflop, or an equivalently powerful machine using accelerators, to support his advanced physics research.
- Many applications heavily used and / or requested by IU researchers and external collaborators are available in versions that effectively exploit GPUs. These codes include molecular dynamics, docking, and engineering codes. RT's research collaborator with Dr. John Duer of Cummins supports the need for accelerator-based nodes – a use that would be of particular value in economic development within the state of Indiana.
- The proposed Indiana Network Science Institute (INSIEME) has begun outlining its cyberinfrastructure needs. Large amounts of data storage and intense bursts of computation are likely based on the early work planning of INSIEME and on the pattern of past usage. Examples from the past include using Big Red to do daily predictions of the spread of the H1N1 virus. Some labs that will certainly become part of INSIEME have traditionally had local cyberinfrastructure based on Sun servers deployed within departmental labs. A robust expansion of central systems and queuing policies that meet the needs of INSIEME researchers could significantly aid in the centralization of IT resources and achievement of economies of scale.








9.2.3. *Results from interviews regarding user needs for file systems*

IU has long been a leader in data-centric advanced computing. We were doing “big data” a decade before that term was coined. We are at capacity to meet current needs and do not have the capacity to meet projected needs for disk-based storage and manipulation of funded research projects, such as:




- HathiTrust Research Center, which currently needs 17 TB of project space that will grow to 500 TB over the next five years.
- The Center for Neuroimaging, which currently needs 200 TB of project space for ongoing research in the national Alzheimer's Disease Neuroimaging Initiative (ADNI).
- Storage of genomics data by the Center for Genomics and Bioinformatics is currently at 375 TB. In three years that could double to 750 TB.
- The Center for Medical Genomics in the IU School of Medicine requires 150 TB per year, resulting in 450 TB in three years.
- The Indiana CTSI whole genome sequencing project requires 200 TB per year, resulting in 600 TB in three years.
- Edenberg Lab sequencing projects will generate 50 TB per year, resulting in 150 TB in three years.
- Tatiana Foroud's Parkinson's and other rare diseases research requires 50 TB.
- Edenberg's alcohol genetics research projects will generate 40 TB.
- The National Center for Genome Analysis Support currently requires 45 TB per year, resulting in 135 TB in three years.
- The IU School of Medicine Pathology Core, to be launched in late 2012, will require 100 TB.

10. Appendix – results of 2012 survey of computational and data usage patterns and needs

Which IU supercomputers/advanced computing systems do you use? (Select all that apply.)

#	Answer	Bar	Responses	%
1	Big Red		12	8.22%
2	Quarry		23	15.75%
3	Mason		5	3.42%
4	FutureGrid - India		2	1.37%
5	FutureGrid - Xray		1	0.68%
6	FutureGrid - Other		2	1.37%
7	Departmental or School computational resources		132	90.41%
	Total		177	100.00%

Of the supercomputers/advanced computing systems listed below, which one is most useful to you?

#	Answer	Bar	Responses	%
1	Big Red		4	2.92%
2	Quarry		17	12.41%
3	Mason		0	0.00%
4	FutureGrid - India		0	0.00%
5	FutureGrid - Xray		0	0.00%
6	FutureGrid - Other		1	0.73%
7	Departmental or School computational resources		115	83.94%
	Total		137	100.00%

**Which federally funded supercomputing resources, provided by any or all of the following agencies do you use?
(Select all that apply.)**

#	Answer	Bar	Responses	%
1	XSEDE - Kraken		3	2.13%
3	XSEDE - Blacklight		0	0.00%
6	DOE InCite		2	1.42%
4	XSEDE - Trestles		1	0.71%
7	Other National Organization		8	5.67%
8	Have not used non-IU resource		129	91.49%
2	XSEDE - Ranger		4	2.84%
5	Open Science Grid		2	1.42%
	Total		149	100.00%

Of the national supercomputing resources you have used or know of, which ONE has been or would be the MOST effective in enabling your research?

Text Entry

Ranger

NCBI ToolKit

Kraken

Argonne National Laboratory's Intrepid system

Please tell us what value you get from local versus national supercomputer facilities.

Text Entry

Testing

local: national=30:70





I get almost no value from UITS services. I do have a small GPU cluster that has been useful for development work and for my graduate student. It is hard to pick a single national supercomputing resource for the question above. Kraken, Longhorn, Hopper, Dirac, Forge, Keeneland and others all play a useful role in my work. It would be useful to have a local GPU cluster or a computer on which one could regularly run on >1000 cores. Basically, I have given up on using BigRed or Quarry.

National-scale systems are larger, including allowing reasonable turnaround on medium-sized jobs for debugging. Local resources are good because our departmental cluster is not busy (but out-of-date and slow) and gives basically instant turnaround.

Of the two main supercomputers at IU (Big Red and Quarry), which one do you use MOST often?

#	Answer	Bar	Responses	%
1	Big Red	<div></div>	6	4.32%
2	Quarry	<div></div>	22	15.83%
3	Don't use Big Red or Quarry	<div></div>	111	79.86%
	Total		139	100.00%

Considering the jobs you send to \${q://QID10/ChoiceGroup/SelectedChoices}, how much wall clock time does a typical job require? (Consider only the time from job start to job finish, not the time spent waiting in the queue.)

#	Answer	Bar	Responses	%
1	Less than 5 hours		15	60.00%
2	5 to less than 10 hours		4	16.00%
3	10 to less than 34 hours		4	16.00%
4	34 or more hours		2	8.00%
	Total		25	100.00%

On average, how many jobs do you submit to $\{q://QID10/ChoiceGroup/SelectedChoices\}$ in a week?

#	Answer	Min Value	Max Value	Average Value	Standard Deviation
1	Average number of jobs submitted to $\{q://QID10/ChoiceGroup/SelectedChoices\}$ in a week	0	10000	486.82	2179.85

How many nodes does a typical job you submit to $\{q://QID10/ChoiceGroup/SelectedChoices\}$ require?

#	Answer	Min Value	Max Value	Average Value	Standard Deviation
1	Nodes Required	0	200	22.53	46.04

If UITS were to purchase a replacement system for Big Red, would you have a preference in the processor architecture?

#	Answer	Bar	Responses	%
1	Yes	<div></div>	6	23.08%
2	No	<div></div>	20	76.92%
	Total		26	100.00%

Which processor architecture would you prefer?

#	Answer	Bar	Responses	%
1	Intel/AMD based	<div></div>	5	83.33%
5	Other		0	0.00%
2	POWER7 based		0	0.00%
3	BlueGene (PowerPC)	<div></div>	1	16.67%
4	SPARC based		0	0.00%
	Total		6	100.00%

If UITs were to purchase a replacement system for Big Red, would you have a preference in the internal networking fabric?

#	Answer	Bar	Responses	%
1	Yes	<div></div>	3	11.54%
2	No	<div></div>	23	88.46%
	Total		26	100.00%

Which networking fabric would you prefer?

#	Answer	Bar	Responses	%
1	10Gb Ethernet	<div></div>	1	33.33%
2	Infiniband	<div></div>	2	66.67%
3	Cray Interconnect		0	0.00%
4	NUMalink		0	0.00%
5	Other		0	0.00%
	Total		3	100.00%

Are there any computational research projects that you would like to undertake but are currently unable to, due to a shortcoming in UITS offerings?

#	Answer	Bar	Responses	%
1	Yes	<div></div>	12	8.76%
2	No	<div></div>	125	91.24%
	Total		137	100.00%

Please describe a computational research project that you would like to undertake, and explain the specific shortcomings that need to be remedied, before you can start.

Text Entry
<p>We have been using Gaussian03 (G03), a computational chemistry software, in our project on both BigRed and Quarry. Unfortunately, the parallel use of G03 is only limited to single node (4 cpus on BigRed and 8 cpus on Quarry). This limit prevents us from performing quantum calculations for large-sized molecular systems. As far as we know, parallel G03 runs are made possible through a software called TCP/Linda, therefore the problem may be related to incompatibility between TCP/Linda and the BigRed/Quarry platforms. Regarding Gaussian on IU computing resources, another shortcoming that need to be remedied is to upgrade G03 to G09, which provides a number of new functions, without which calculations using the most up-to-date computational methods cannot be done.</p>
<p>I would like my student to be able to more than a thousand jobs each taking on the order of 200 core hours. If that is successful, I would really like to be able to run jobs taking >1000 cores on CPUs. Alternatively, I would like to be able to run on 10-64 GPUs in parallel.</p>
<p>I would like access to more file sharing, survey, and qualitative data analysis software (DropBox, something like SurveyMonkey that is accessible for non-IU collaborators, and NVivo).</p>
<p>distributed rendering for 3D software, IE a render farm for advanced rendering of images from software such as Autodesk Maya</p>
<p>3d image</p>
<p>Drop Box, questionnaire program like surveymonkey, additional data for icloud</p>
<p>Need bigger systems</p>
<p>Finite element analysis</p>
<p>More locally available hard disk space (beyond 10 GB). It's unclear how to ask for such space, and it seems like most people who need more just do their computing locally -- which is what I have decided to do, albeit with inferior processors.</p>
<p>I do functional neuroimaging analysis. I tried to use Quarry, but there were two problems that prevented me from using it effectively. First, the bandwidth between the Psychology building and Quarry is too slow for my large imaging data sets, so by the time I transfer them there and back, I might as well have run it on my local workstation, because it is not any faster. Second, the ability to mount drives and map user IDs between my linux server and clamp is poor at best. I tried kernel upgrades with Lustre but it never could map properly to my 192.168* subnet space. I then tried fuse to mount over sshfs, but that was cumbersome and required mounts separately on each Quarry node. So I gave up and am now just using my own servers. If you can improve bandwidth and network mounting (with user id mapping), that would really help.</p>
<p>View More</p>

Which of the high performance file systems do you use on either Big Red or Quarry? Select all that apply.

#	Answer	Bar	Responses	%
1	Data Capacitor	<div></div>	7	30.43%
2	Data Capacitor WAN	<div></div>	2	8.70%
3	Neither	<div></div>	16	69.57%
	Total		25	100.00%

How many output files do you produce in a typical job on Big Red or Quarry?

#	Answer	Min Value	Max Value	Average Value	Standard Deviation
1	Output files	1	500	47.05	113.74





What is the size of an output file for a typical job on Big Red or Quarry?

#	Answer	Bar	Responses	%
1	less than 10MB	<div></div>	5	26.32%
2	10MB to under 100MB	<div></div>	6	31.58%
3	100MB to 1GB	<div></div>	5	26.32%
4	more than 1GB	<div></div>	3	15.79%
	Total		19	100.00%





Do you use IU's Scholarly Data Archive (formerly known as Massive Data Storage System) to backup or store data?

#	Answer	Bar	Responses	%
1	Yes	<div></div>	34	24.64%
2	No	<div></div>	104	75.36%
	Total		138	100.00%

What is the total volume of your stored data in the Scholarly Data Archive?

#	Answer	Bar	Responses	%
1	less than 10GB		3	9.09%
2	10GB to under 100GB		10	30.30%
3	100GB to under 1TB		13	39.39%
4	1TB to 10TB		7	21.21%
5	more than 10TB		0	0.00%
	Total		33	100.00%



Please indicate the selection below that best describes how often you write data to the Scholarly Data Archive.

#	Answer	Bar	Responses	%
1	Daily		7	21.88%
2	1 or more times per week		3	9.38%
3	1 or more times per month		6	18.75%
4	3-4 times per year		10	31.25%
5	1-2 times per year or less		6	18.75%
	Total		32	100.00%

So that IU as a whole can make most effective use of its resources, UITs is considering an approach of asking the very few top users of IU cyberinfrastructure resources – perhaps the top 25 to 50 users – to write a very short annual request for use of resources (supercomputer time, storage space). This will help UITs plan better and also help us identify uses that might well be a good match for federally funded resources such as XSEDE, Open Science Grid, and InCite, so that UITs staff can help IU researchers obtain resources on those federally funded resources, thus maximizing the amount of IU-funded resource available to IU researchers as a whole.

Such a process would be very brief – a maximum of two pages of narrative describing the request, accompanied by an NSF - or NIH - format brief biographical sketch. UITs would engage faculty reviewers and all requests would receive one of three evaluations: possible candidate for an allocation on a federally-funded CI resource; not a possible candidate for a federally-funded CI resource but meritorious; a request that is frivolous. Only truly frivolous requests would not be fulfilled. Faculty who submit requests that were viewed as possible candidates for allocation on a federally-funded CI resource would be encouraged and assisted to apply for an allocation on a federal resource by UITs staff, and work would proceed on IU resources while requests for federal resources were pending. The review process would take less than four weeks from deadline for submission until a response was provided to faculty members, and would be conducted twice annually: once in spring in order to plan for the coming academic year, and once early in fall to aid in meeting needs of new incoming faculty.

Given that this process could help IU researchers increase use of existing federal resources and obtain millions of dollars worth of computing resources per year, do you think that a process such as described above – involving likely 50 or fewer faculty members - would be unnecessarily burdensome?

#	Answer	Bar	Responses	%
1	This would not be unnecessarily burdensome		99	74.44%
2	This would be unnecessarily burdensome		34	25.56%
	Total		133	100.00%

Next, we have a few questions about the types of software you use.

What type of software do you primarily use in your research?

#	Answer	Bar	Responses	%
1	Software written by my research group	<div></div>	11	8.33%
2	Software provided by a third party	<div></div>	79	59.85%
3	Both	<div></div>	42	31.82%
	Total		132	100.00%

Please list the third party applications you use most often on IU supercomputers.

Text Entry
I don't use the supercomputers as far as I can tell.
Guassian, CHARMM, NAMD
I have no idea what you are asking
Will need to ask my research group; all related to genome sequencing
Matlab, Fortran
afni, matlab, fsl
NA
I don't really do research here at the university as I am lecture level and primarily concerned with teaching. However since I teach technology I totally support using it. In my class and for other parts of my job I use the Microsoft Office Suite, Adobe Creative Suite, WinSCP, Audacity, and Windows Movie Maker regularly.
CERN statistical data package
None on supercomputers.
View More

Is the third party application you use most often a serial or parallel application?

#	Answer	Bar	Responses	%
1	Parallel	<div></div>	8	34.78%
2	Serial	<div></div>	11	47.83%
4	Other (Please Specify)	<div></div>	4	17.39%
	Total		23	100.00%

Is the third party application you use most often offered under a commercial or open-source license?

#	Answer	Bar	Responses	%
1	Commercial	<div></div>	41	69.49%
2	Open-source	<div></div>	18	30.51%
	Total		59	100.00%

Please help us understand your software development needs.



Please indicate the style you use for software development. (Select all that apply.)

#	Answer	Bar	Responses	%
1	Parallel	<div></div>	18	56.25%
2	Serial	<div></div>	18	56.25%
3	Other (Please specify)	<div></div>	3	9.38%
	Total		39	100.00%

Of the self-written codes you use, in what language are they primarily written?

#	Answer	Bar	Responses	%
1	C		2	4.88%
2	C++		8	19.51%
3	FORTRAN		8	19.51%
4	Java		3	7.32%
5	Matlab		8	19.51%
6	Perl		5	12.20%
7	Python		3	7.32%
8	Shell		0	0.00%
9	Other (Please specify)		4	9.76%
	Total		41	100.00%

What parallel method(s) do you use? (Select all that apply.)

#	Answer	Bar	Responses	%
1	Threads		10	58.82%
2	Message Passing (MPI)		6	35.29%
3	Hybrid threads/MPI		4	23.53%
4	Other (Please specify)		1	5.88%
	Total		21	100.00%

Please indicate your level of satisfaction for each of the following UITS service areas.

#	Question	Extremely Dissatisfied	Somewhat Dissatisfied	Neither Satisfied nor Dissatisfied	Somewhat Satisfied	Extremely Satisfied	Don' t Know/ Not Applicable	Responses	Mean
1	Available Software	1	5	12	46	41	16	121	4.40
2	User Support	2	7	7	42	47	15	120	4.42
3	Amount of Time Jobs Wait in Queue	3	3	17	23	15	56	117	4.81
4	Amount of Memory per Node	1	4	14	18	12	67	116	5.04
5	Space Available for Data Storage	2	10	13	28	23	40	116	4.55
6	Documentation of Available Systems	2	3	21	16	13	61	116	4.88
7	System Interconnect	2	7	11	11	12	73	116	5.09
8	Number of Nodes	3	1	11	11	8	82	116	5.29
9	Cores per Node	1	-	13	12	7	83	116	5.35

In an effort to increase outreach to undergraduates UITS has begun programs involving research projects for undergraduates and classroom education programs for undergraduate classes. Would you be interested in participating, or having your undergraduate class participate in a program like this? (Select all that apply.)

#	Answer	Bar	Responses	%
1	Yes, I'm interested in mentoring a student	<div></div>	8	6.45%
2	Yes, I'm interested in teaching a class	<div></div>	5	4.03%
3	Yes, I'm interested in having my class participate	<div></div>	12	9.68%
4	No, I'm not interested at this time	<div></div>	110	88.71%
	Total		135	100.00%