



invent, develop, deploy, and deliver
**PERVASIVE
TECHNOLOGY
INSTITUTE**

Cyberinfrastructure Begins at Home

An introduction to IU's Cyberinfrastructure Delivery, Development, and Research Strategies

David Hancock Manager
High Performance Systems

Robert Henschel Manager
Scientific Applications & Performance Tuning

Kristy Kallback-Rose Manager
Research Storage

SPXXL – January 2015



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

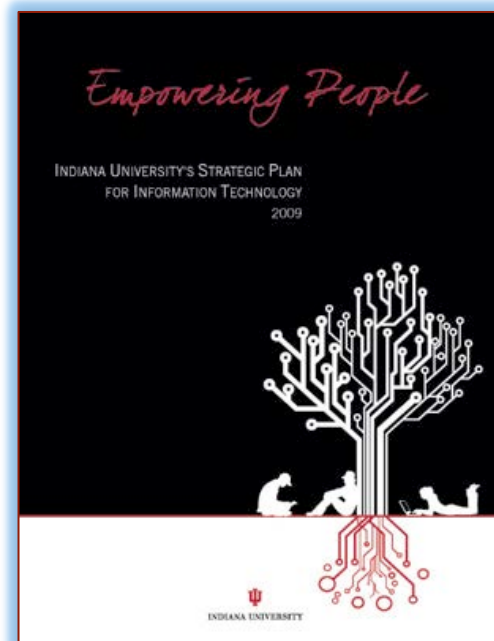
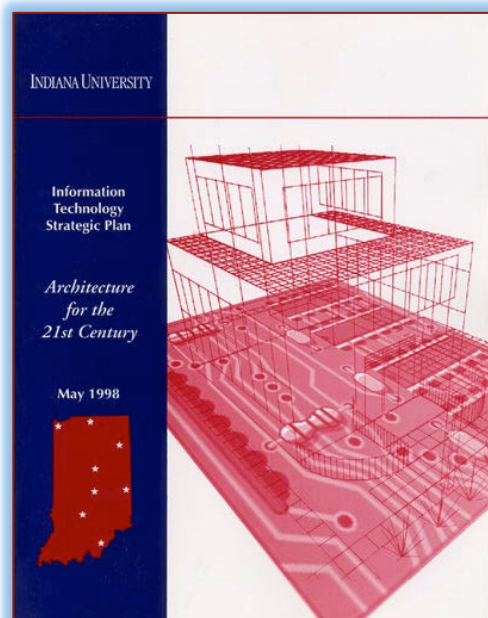
Outline

- IU Goals
- Indiana University background
- How we got here
- Pervasive Technology Institute
- Research Technologies Division of UITS
- Relevant Projects / Groups



IU goals

- To be one of the great public universities of the 21st Century (Michael A. McRobbie, 18th President of IU)
- To be a leader, “in absolute terms for uses and applications of IT” (Myles Brand, 16th President of IU)



IU – Campuses and Medical School Centers



IU Campuses



IU School of Medicine campuses and clinics



IU – Two Core Research/Education Campuses, Six Regional Campuses

Campus	Academic appointees	Staff (non-academic)	Undergraduate Students	Graduate & Professional Students
IUB	3,012	5,554	36,862	9,955
IUPUI	3,914	4,375	22,409	8,079
IU Northwest	399	245	5,904	483
IU South Bend	515	313	7,512	561
IU East	282	167	4,326	130
IP Fort Wayne	838	736	12,500	800
IU Kokomo	210	138	3,981	197
IU Southeast	504	234	6,148	585
Totals	9,674	11,762	99,642	20,790

IU community: 141,868 people total



University Information Technology Services

- Centralized IT across all campuses
- 700+ professional staff
- 500+ part time staff
- Regional CIOs report to VP of IT and CIO (Wheeler)



Key IU Metrics



Indiana University Health

IU Budget Category	2014/2015 Budget
Unrestricted	\$2,283,067,642
Restricted	\$629,755,034
Auxiliary	\$417,540,459
Total	\$3,330,363,135

IU Health Patient Metrics – 2014	
Admissions	136,731
Outpatient visits	2,638,074
Staffed Beds	3,098



Pervasive Technology Institute

The mission of the Indiana University Pervasive Technology Institute (PTI) 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.

- PTI is a collaboration among the School of Informatics and Computing, the Maurer School of Law, the Office of the Vice President for Information Technology, University Information Technology Services, and the College of Arts and Sciences.
- Consists of two types of centers:
 - **Research Centers:** D2I; CACR; DSC; CREST. PTI Research Centers' fundamental goal: research excellence, sustained.
 - **Cyberinfrastructure and Service Centers:** The PTI Service & Cyberinfrastructure Centers (RT & NCGAS) develop and deliver leading-edge services.



RT Mission

The mission of the Research Technologies division of UITs 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.

- We are a mission- and value-driven organization. We are not a technology-driven organization.
- We identify needs, identify possibilities, and discover new ways to meet those needs, realize those possibilities, and create new ones. In so doing we create, deploy, and support technology.

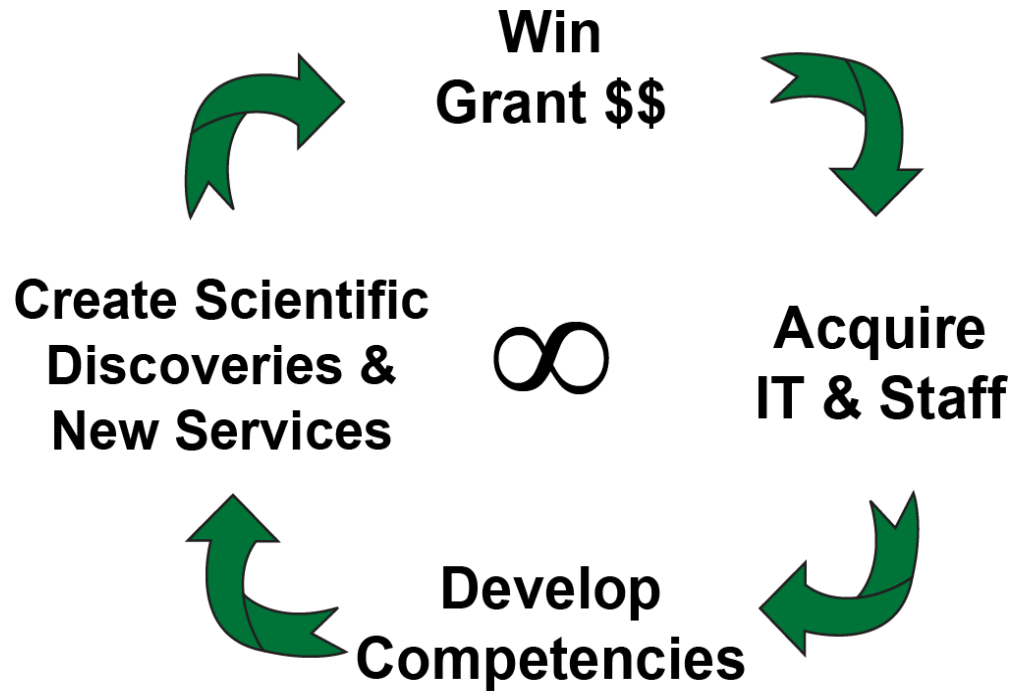




Permission to use kindly granted by
Ridin-Hy Ranch, PO Box 369, Warrensburg NY 12885



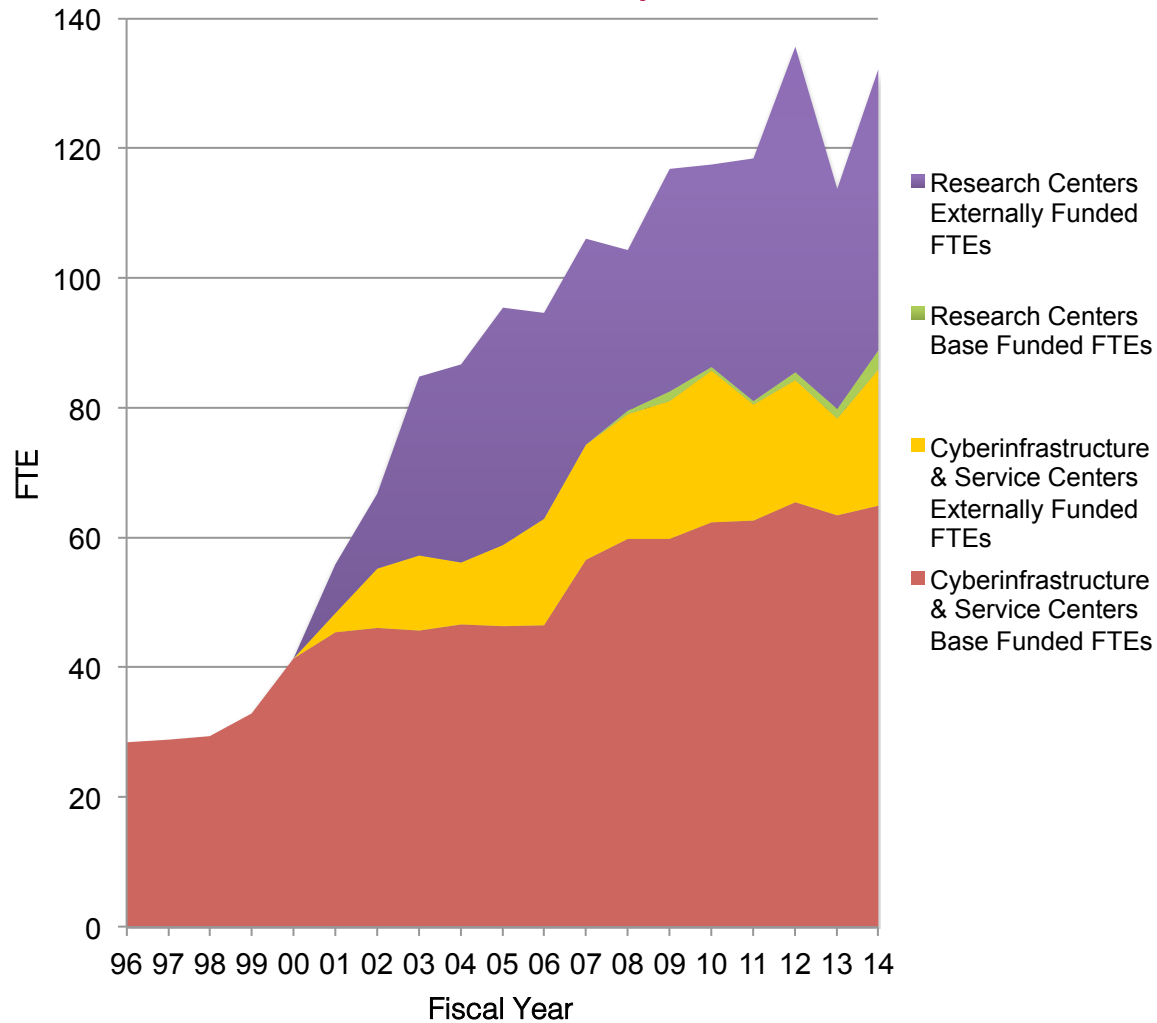
Positive Feedback Loop



Based on graphic by Bradley C. Wheeler. Used under Creative Commons 3.0 unported attribution license



PTI FTE Count by Fiscal Year



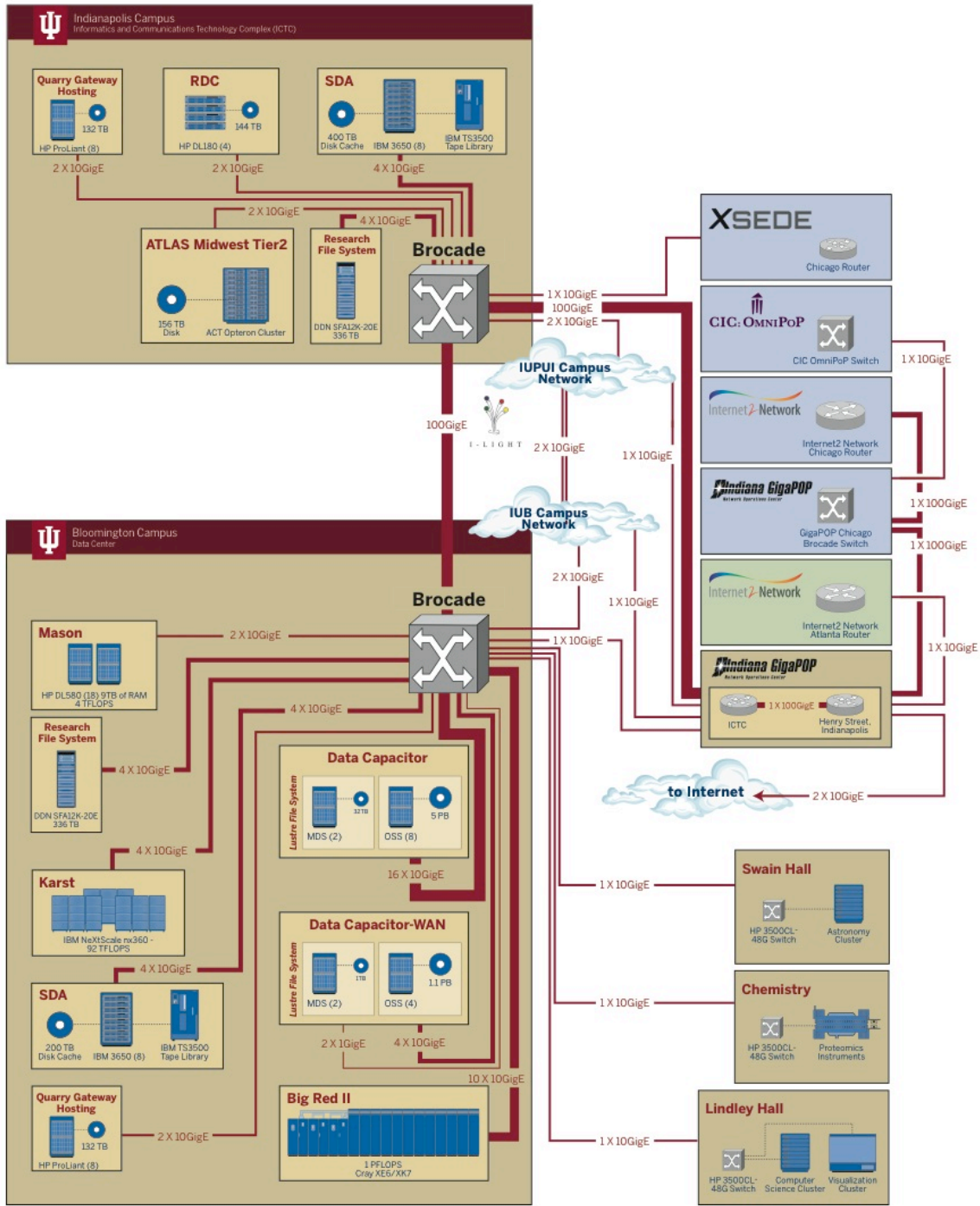
PTI Research Center Funding 2008-2014

PTI Center	Total Funding Awarded	# Awards
Digital Science Center	\$38,529,948	72
Data to Insight Center	\$35,495,380	45
<i>Research Technologies</i>	<i>\$31,389,903</i>	<i>49</i>
CREST	\$17,215,634	34
CACR	\$9,550,686	22
NCGAS	\$3,919,934	10
PTI Total	\$136,101,485	232



IU Cyberinfrastructure



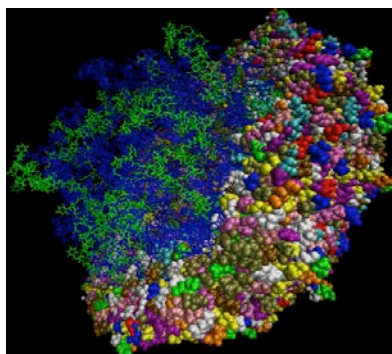


Global NOC Supported Projects



RT Services Spotlight – Computation

- **Big Red II** – IU's main system for high-performance parallel computing
 - Owned solely by IU
 - 1 petaFLOPS
 - CPUs/GPUs
 - 3.5 PB of scratch

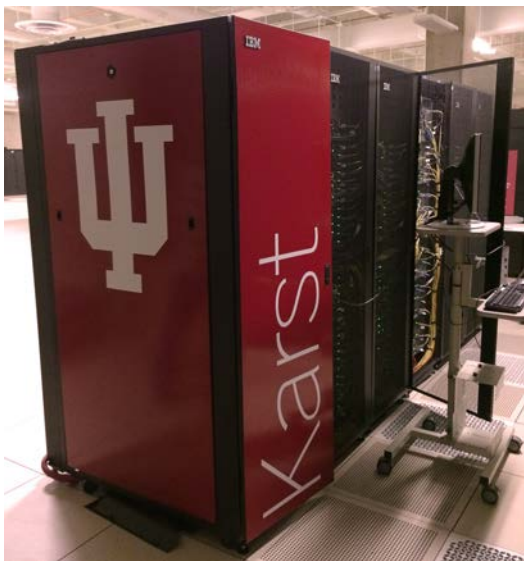


- Available to all Faculty, Staff, and Graduate Students
- Support/consulting available



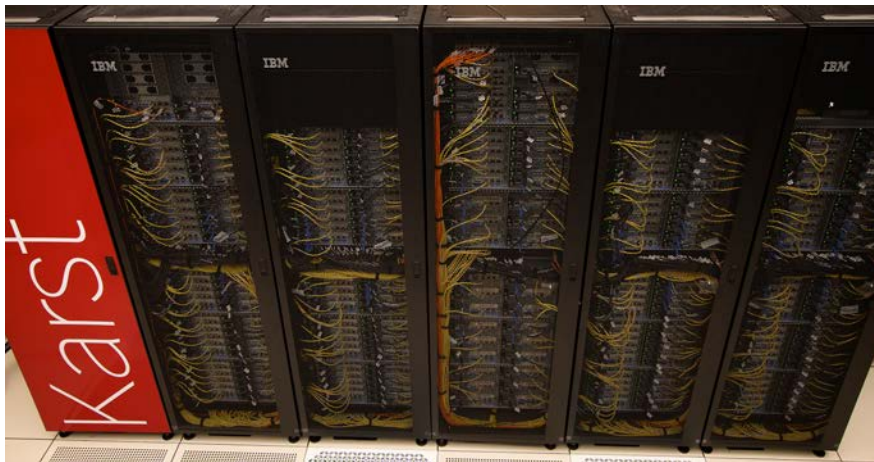
RT Services Spotlight – Computation

- **Karst** – standard cluster available for expansion
 - General purpose Intel Linux cluster
 - Condo nodes may be purchased for special needs or greater response
- **Mason** – large memory computer cluster
 - Supports data-intensive, high-performance computing tasks
 - Specifically for genome researchers
 - Uses include genome assembly software, large-scale phylogenetic software, and other genome applications

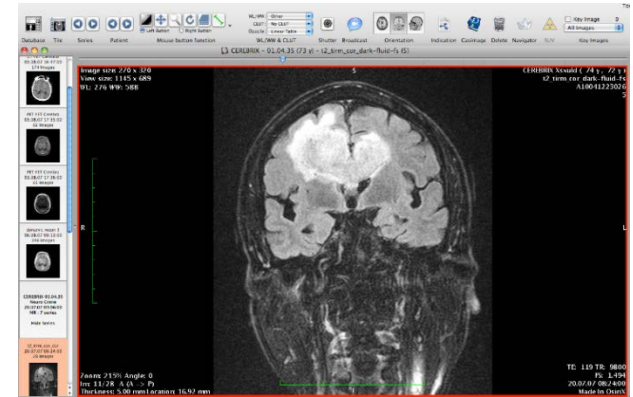


Karst Installation

- Delivery Schedule Accelerated
- Serial number assignment in question
- Networking/Node discovery issues
- Not delivered with configuration requested
- Benchmarking/Acceptance problems due to personnel assignments
- Follow-Up Orders a hot mess



Campus Bridging and Research Infrastructure



- Make the transition from lab and instrument to central IT easier, better, simpler
- Logistics and field support for Ice Sheet research with KU's Center for Remote Sensing of Ice Sheets
 - NSF and NASA-led initiatives in Greenland, Chile, Antarctica
- Clinical Translational Science Institute Hub and Bio-IT Core Services
- Medical Imaging and Tissue Management applications services
- XSEDE Campus Bridging initiatives, lowering barriers to doing research with XSEDE



What is Jetstream?



Award #1445604

- NSF's first cloud dedicates top support for science and engineering research across all areas of activity supported by the NSF
- Jetstream will be a user-friendly cloud environment designed to give researchers and research students access to interactive computing and data analysis resources "on demand."
- User-selectable library of VMs
- VMs can be imported and customized.
- It will enable countless discoveries across disciplines such as biology, atmospheric science, economics, network science, observational astronomy, and social sciences.
- Two especially important biology platforms will be supported - iPlant and Galaxy.



The setup



Award #1445604

- National Science Foundation (NSF) estimates that 299,000 researchers, educators, and learners received direct support during the year ending September 2013, yet merely 1.5% completed a computation, data analysis, or visualization task on XD program resources and less than 3% had an account on the XSEDE portal.
- Jetstream will be a new, interactive cloud-based system that, through its scale and flexibility, will dramatically enhance the diversity and size of the US researcher community benefiting from XD resources.



Timeline & Partners



Award #1445604

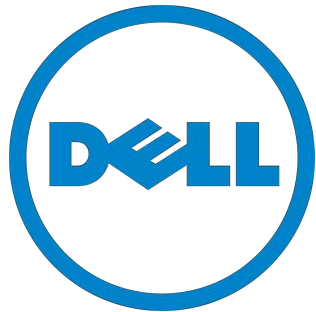
- Program Execution Plan (PEP) – Finalized by end of Feb
 - Test gear to IU – 1Q2015
 - Production gear shipped to IU and TACC – mid-2015; test gear then shipped to Univ. of Arizona
 - Friendly user mode – ~SC15
 - Workshop at SC to do final adjustments
 - Full Production – January 2016
-
- Indiana University, Texas Advanced Computing Center, University of Arizona, and University of Chicago
 - Cornell, John Hopkins, NSIDC, Odum Institute at UNC, Penn State, University of Arkansas at Pine Bluff, University of Hawaii, University of Texas (San Antonio)



Key Technologies



Award #1445604



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY

RT Services Spotlight – Data Storage

- Data Capacitor II, DC-WAN
 - Data storage on disk, not backed up (scratch & projects)
 - Temporary storage of research data – purged regularly
 - 3.5 PB DCII / 339 TB DC-WAN
- Wrangler coming soon (dual-site 20 PB environment with TACC)



- Lustre-based file systems
- Serves all IU campuses
- Mounted on Big Red II, Karst, and Mason research computing systems



RT Services Spotlight – Data Storage

- Research File System (RFS) / Computational Home Dirs
 - Centralized disk storage designed to support researchers
 - Backed up regularly – 2 copies (IUB and IUPUI)
 - Open to IU community – undergrads/non-IU must have sponsor



- Can utilize project areas in file system
- HIPAA-aligned
- GPFS based



RFS Update – Access Methods

- November 2014, Home Directories
 - Transition from NetApp 3170 to GPFS via NFS (CTDB not cNFS, because we are using CTDB for Samba)
- February 2015, Desktop Access
 - Individual space, project space
 - Replacement for OpenAFS solution for same purpose
- March 2015, Home Directories
 - Transition from NFS to GPFS Native Client



GPFS Features in Use

- Synchronous Replication (2 site)
 - Works well, lost enclosures hard at one site due to PDU snafu, the file system did not hiccup
 - Strong HA feature, limiting downtime



GPFS Futures

- GPFS-HPSS-Integration (GHI)
 - POC beginning Feb 2015
 - Plan to use for DR backups
 - Currently homebrewed script solution to HPSS
 - Leverage Lower tape cost
 - Want to be able to leverage GPFS ILM to stub out lesser used files, will require user training...how to show stubbed file in Samba, e.g.
- Funding model may move to larger labs paying for space
 - Dedicated hardware solution (GSS?) or just disk expansion...TBD



RT Services Spotlight – Data Storage

- Scholarly Data Archive (SDA)
 - Distributed tape storage for large-scale archival/near-line storage
 - Mirrored– 2 copies (IUB and IUPUI)
 - Open to IU community – undergrads/non-IU must have sponsor
 - Supports collaborative activities



- 43 PB of tape storage capacity
- Supports SFTP, HSI, HPSS API
- HIPAA-aligned



Education and outreach



We Live the Myth of Sisyphus



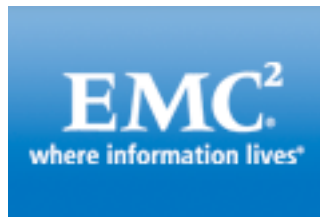
"The struggle itself...is enough to fill a [person's] heart. One must imagine Sisyphus happy."

–Albert Camus

Sisyphus (1548-1549) by Titian, Prado Museum, Madrid, Spain
http://en.wikipedia.org/wiki/File:Punishment_sisyph.jpg
This work is in the public domain in the United States, and those countries with a copyright term of life of the author plus 100 years or fewer.



Thanks to our supporters



Contact Information

David Hancock Manager

High Performance Systems

dyhancoc@iu.edu

Robert Henschel Manager

Scientific Application & Performance Tuning

henschel@iu.edu

Kristy Kallback-Rose Manager

Research Storage

kallbac@iu.edu

IU Pervasive Technology Institute

IU Research Technologies

<http://pti.iu.edu/>

<http://rt.uits.iu.edu/>

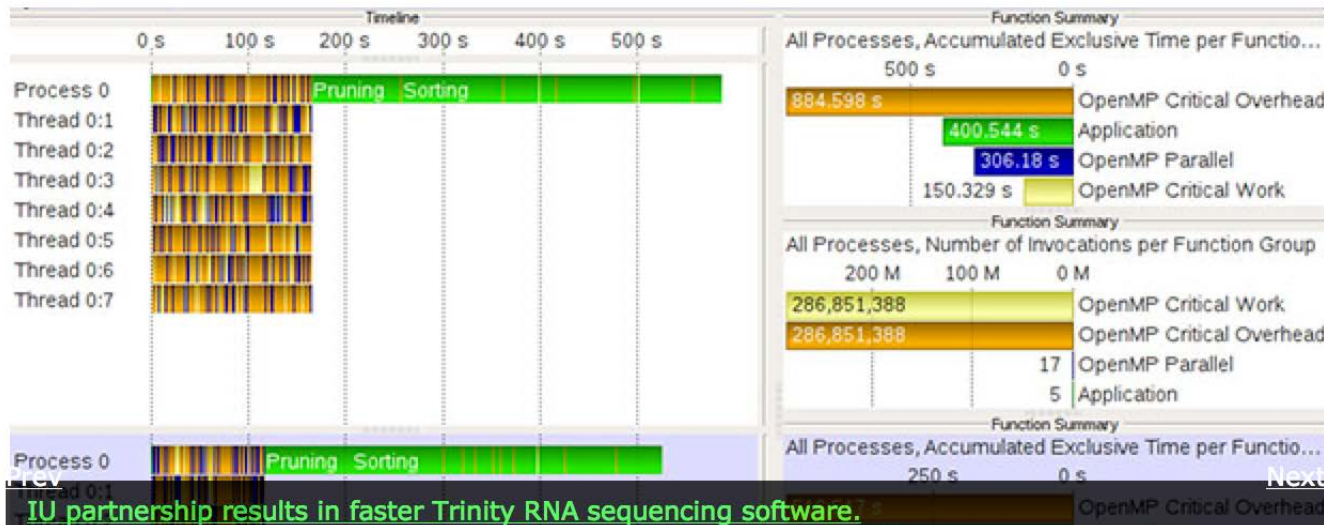


RT Services Spotlight – SciAPT

Scientific Applications and Performance Tuning

To deliver and support software tools that promote effective and efficient use of IU's advanced cyberinfrastructure – which, in turn, improves research and enables discoveries.

- Short and Long-term consulting
- Performance analysis and tuning
- HPC workshops for researchers



Broad Collaboration
4X speed up

