

# UAB IT Research Computing Update

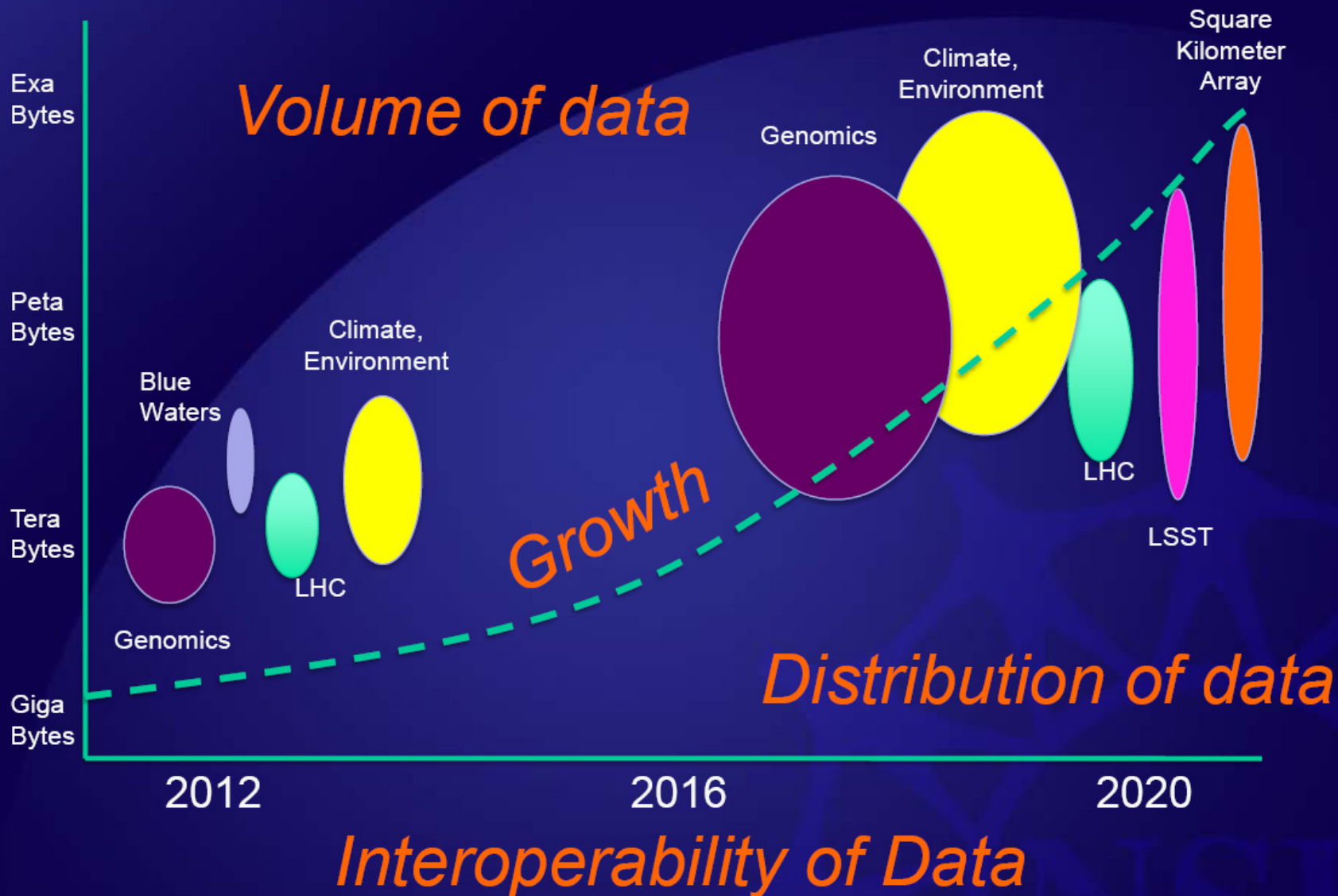
## UAB IT Research Computing Day

September 15, 2011

# Discussion Topics

- Research Applications
- Developer Tools
- Hardware
- Networks
- People
- Current Projects

# Data Challenges





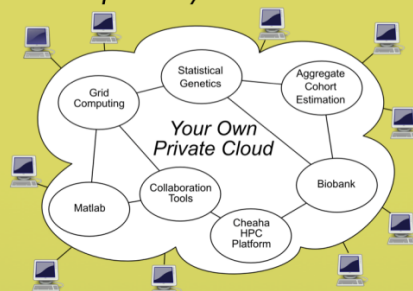
# Research Computing Support

brought to you by UAB IT



## UAB Research Cloud

platforms, computing and storage  
to power your research



The research cloud is designed to support private workspaces with access to the tools you need where ever you need them. Take your lab anywhere, virtually.

## UASRON your connection to the world



### UASRON Alabama



UASRON is a very high bandwidth lambda network that connects the UA System to National Lambda Rail and Internet2. UAS RON is a collaboration between the UA System, NASA, and Southern Light Rail (SLR)@GaTech.

University of Alabama System Owned and Operated, UASRON is a Dense Wave Division Multiplex (DWDM) Network, analogous to a 40-Lane, bi-directional, Data Highway connecting UAB, UAH, UA, ASA, to Atlanta and Nashville. Current technology allows each lane (channel) to carry a payload of 10Gigabits-per-second.

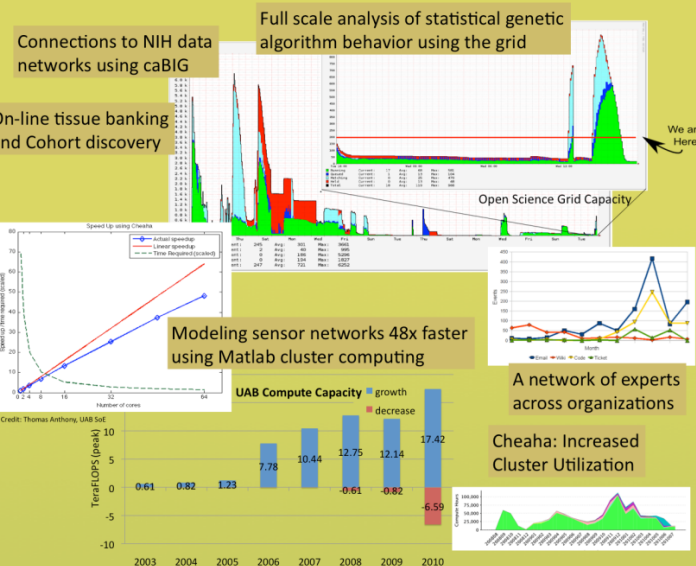
## Research Apps solutions built to order



MatLab, Biobank/caTissue, ACE, Gromacs, NAMD, Amber, FSL, GNU Scientific Toolkit, and much more

A large variety of High Performance Computing apps help make sense of piles of data and model complex systems. Designer apps help speed up tasks and fill in gaps whether its world wide computing, tissue banking, cohort discovery, or data archiving.

## Results Delivered



A growing resource fabric: combining funds from research Awards and sustained UAB IT investments to fund on-going expansion of compute, network, and storage capacity and maximize the impact and availability of acquisitions.

## Developer Accounts an open platform to build research



Engage your peers. Our platforms, languages, and collaboration tools make it easy to build solutions, document code, record changes, track bugs and collaborate across organizations. Mediawiki, WordPress, mailing lists, Trac, Git, your favorite languages, and slew of other tools: we live on the same platform and run our lab that way too.



## The Team

An engaged professional staff collaborate with investigators to maximize the impact of research

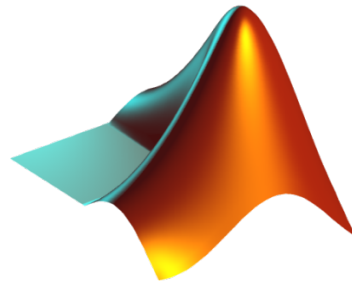
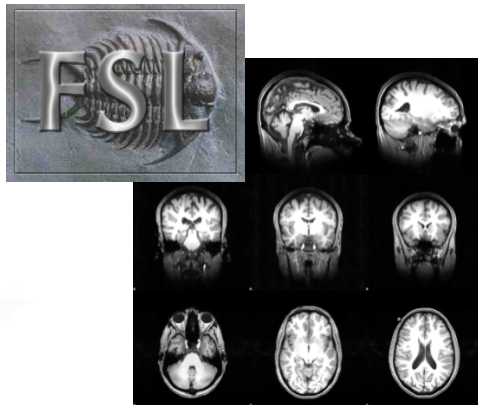




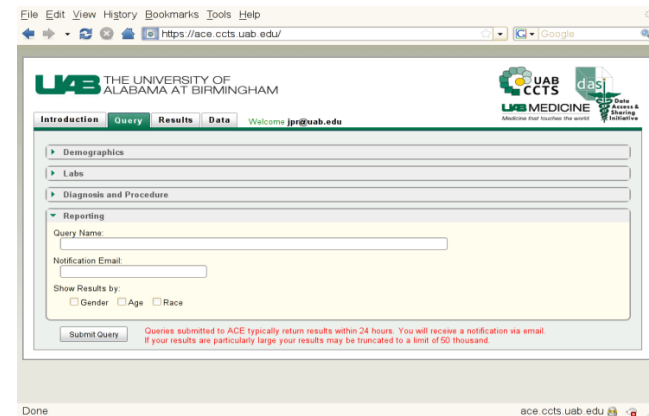
# Research Applications *solutions build to order*

Patient Cohort: ACE

Neuroimaging

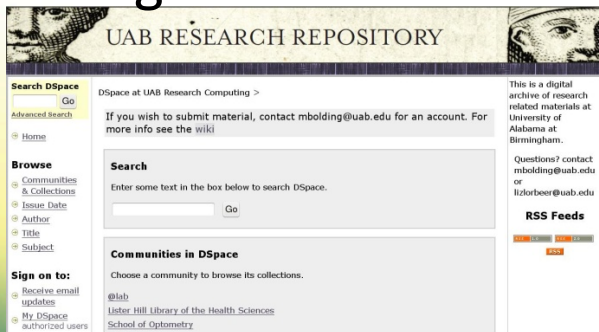


Data Analysis  
with MATLAB

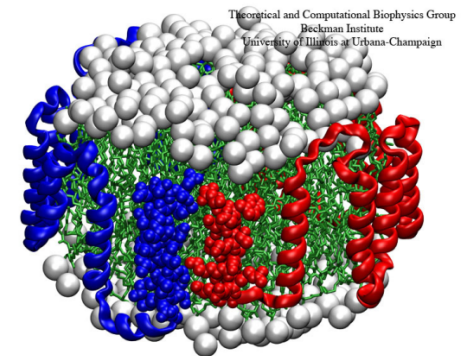


NextGen Sequencing

Digital Archives



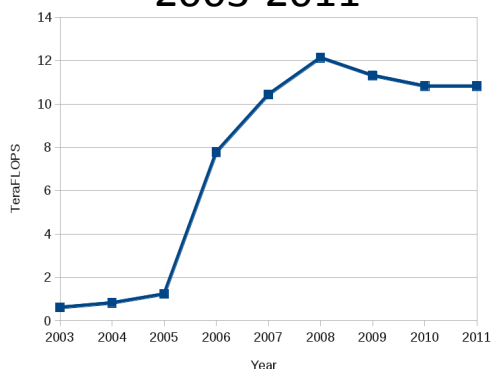
Molecular Dynamics



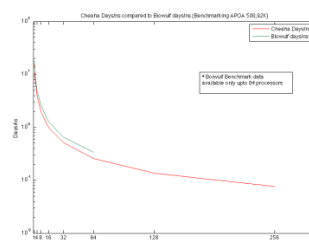
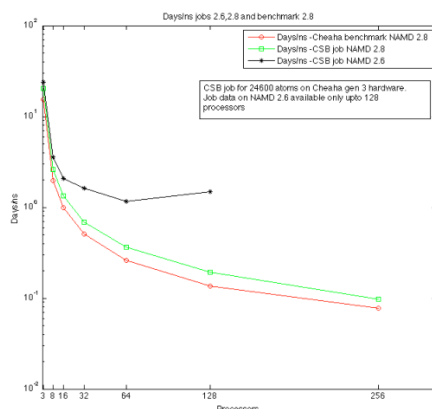
Fast apps help make sense of piles of data and model complex systems. Designer apps improve efficiency and fill the gaps whether its nextgen sequencing, cohort discovery, data archiving, or computing across the world.

# Results Delivered

## HPC TeraFLOPS 2003-2011



## Application Performance 8x Increase for NAMD



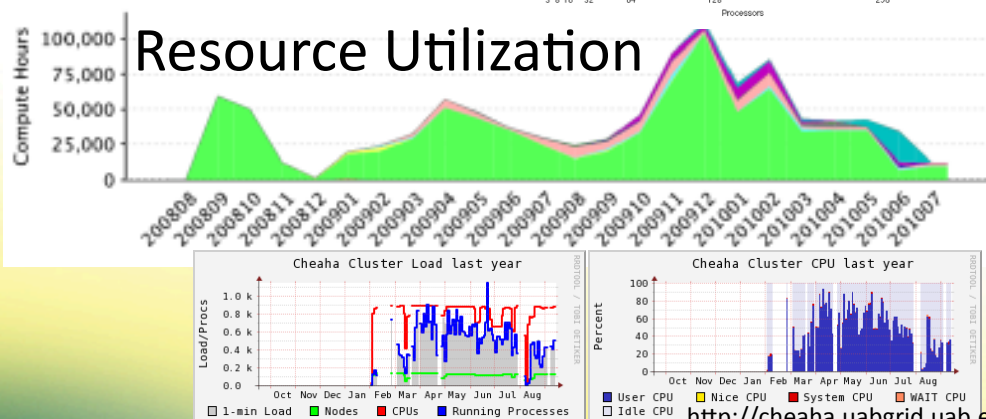
NAMD Comparison

System	Atoms	ns/day	CPU count
UAB BlueGene	246,000	0.80	256
Cheaha with SIG	235,000	0.88	32

## Compute Community Galaxy and MATLAB

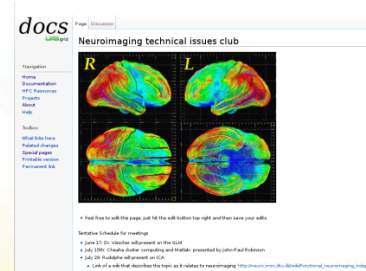


## Resource Utilization



<http://cheaha.uabgrid.uab.edu/ganglia>

## Collaboration Support



### Wiki & Project Spaces:

- Neuroimaging Club
- MATLAB docs
- Digital Archives
- NLP
- Cerner PowerInsight

A growing resource fabric combines dollars from research Awards and UAB IT investments to fund on-going expansion of compute, network, and storage capacity and maximize the impact and availability of acquisitions.

# Developer Tools

## an open platform to build research

docs UABgrid

Page [Discussion](#) [Read](#) [View source](#) [View history](#)

### Welcome

Welcome to **UABgrid**

UABgrid is a technology platform being built to support research at UAB.

UABgrid provides a framework for sharing data, accessing compute power, and collaborating with peers on campus and around the globe. Our goal is to construct a dynamic network of services that supports the development and execution of advanced research processes.

The UABgrid pilot was launched in September 2007 and has since focused on demonstrating the utility of the platform in three key areas:

- **High Performance Computing (HPC)** - **Cheaha**, a compute fabric to power even the largest analysis work loads
- **Data Sharing** - supporting the trusted exchange of information to spark new ideas
- **Collaboration** - putting the tools you need at your fingertips

The research platform is being built on the same technology foundation that has successfully served the needs of prominent national and international initiatives like **caBIG**, the **Open Science Grid**, **TeraGrid**, and the **LHC Computing Grid**.

Construction of the research platform is lead by UAB IT in cooperation with a growing list of collaborators that includes the Center for Clinical and Translational Science (CTS), the Comprehensive Cancer Center (CCC), the Department of Computer and Information Sciences, the Department of Mechanical Engineering, and Health System Information Services (HSIS). Platform development is conducted openly via the [UABgrid development wiki](#).

File Edit View Help

Commit message: Merge remote-tracking branch 'ozborn/add\_fastqc' into add\_fastqc\_ticket

Added the FastQC tool as discussed many months ago. This also contains a bugfix (from galaxy's perspective) in Fixed ticket [ticket:77]

Fixed Ticket [ticket:49]

Changes according to [Ticket:75]

Added C. elegans genome build o66

Increased run-time limit for lophat and bowtie to 16 hours and 6 hours respectively.

Fixed [ticket:74]: Displays env command output in history/web-browser.

Fixed [ticket:72]: Galaxy sets metadata on every job output. This configuration change will run set metadata step on the cluster instead

Merge branch 'upstream-tracker' into release:72045407d to merge with 'upstream-949e45fa03a'

[remotes/origin:upstream-tracker] Added git directory to .hgignore file

commit after updating (hg pull -u) to upstream galaxy-dist revision 949e45fa03a

Merge branch 'dev' of home:ozborn/project:galaxy/galaxy into ozborn/develop

Added Sacromyces Cerevisiae genome (build 3) for bwa, bowtie and sambtools

Applied patch from upstream [https://bitbucket.org/galaxy/galaxy-central/changeset/07de40a5a0b9 galaxy-central repository] to fix

Changed personalize.sh script to use universe\_wsgi.ini sanitized file

Changed memory limits v and h, mem for bowtie, bowtie\_color and yoplat in universe\_wsgi.ini sanitized

Added parallel environment option to drmaa url for lophat, bowtie and bowtie\_color

Untrack universe\_wsgi.ini file using 'git rm --cached'

Modified personalize.sh script to use sanitized version of universe\_wsgi.ini file

Added universe\_wsgi.ini sanitized file - same copy as deployed galaxy uabgrid configuration. Any changes to configuration should

Modified personalize.sh script with read command with prompt option than print and read commands. Also, added -e option to 'yes'

Updated drmaa configuration for cutcompare, cutdiff, outlinks, pileup, parser, snpEff, and lophat run-time and memory limits.

Added TMP and TMPDIR variables in age and drmaa job submission templates. We had tried adding this to environment/modules

Ignore sqllite database universe\_sqllite from git repository - added universe\_sqllite to .gitignore

Added a script that generates sed-commands file which modifies universe\_wsgi.ini file later.

Added personalization and universe\_wsgi.ini old files to git ignore list; personalization file contains sed commands that modify ma

Added Tree view build 62 to bowtie\_indexes.txt file

Following tool-data files were downloaded by galaxy after starting galaxy process for the first time after merge with upstream galax

[remotes/origin:upstream-tracker] Merge branch 'upstream-tracker' into release:72045407d

commit after updating (hg pull -u) to upstream galaxy-dist revision 72045407d

[branch:release:72045407d] Modified .gitignore file to exclude following files

Updating to run with the latest version of outlink and lophat. It turns out that they are not compatible with our current

Updated bwa to print out some debug information to resolve the problem where the /tmp space on the compute node fills up

Shantanu Pa 2011-09-13 16:01

John David C 2011-09-13 16:01

Shantanu Pa 2011-09-13 16:01

Shantanu Pa 2011-09-13 15:52

Shantanu Pa 2011-09-13 08:05

John David C 2011-09-08 10:02

Shantanu Pa 2011-09-06 11:14

Shantanu Pa 2011-09-02 16:02

Shantanu Pa 2011-09-02 09:01

Shantanu Pa 2011-08-31 14:02

Shantanu Pa 2011-08-31 13:52

Shantanu Pa 2011-08-31 13:52

Shantanu Pa 2011-08-26 17:25

John David C 2011-08-26 15:05

Shantanu Pa 2011-08-26 16:02

Shantanu Pa 2011-08-25 11:15

Shantanu Pa 2011-08-24 16:01

Shantanu Pa 2011-08-18 14:02

Shantanu Pa 2011-08-17 14:02

Shantanu Pa 2011-08-17 13:52

Shantanu Pa 2011-08-17 10:02

Shantanu Pa 2011-08-05 12:02

Shantanu Pa 2011-07-26 12:02

Shantanu Pa 2011-07-25 14:02

Shantanu Pa 2011-07-25 13:02

Shantanu Pa 2011-07-25 13:02

Shantanu Pa 2011-07-14 06:01

Shantanu Pa 2011-07-14 04:02

Shantanu Pa 2011-07-13 17:02

Shantanu Pa 2011-07-13 17:02

Shantanu Pa 2011-07-07 15:02

John Osborn 2011-07-07 12:02

John Osborn 2011-07-07 12:02



### Service Points:

- docs.uabgrid.uab.edu
- cheaha.uabgrid.uab.edu
- dev.uabgrid.uab.edu
- git.uabgrid.uab.edu
- blogs.uabgrid.uab.edu
- projects.uabgrid.uab.edu/<pname>

File Edit View History Bookmarks Tools Help

UABgrid dev

Search

Table of Contents

- Overview
- Activities
- Components
- Identity
- Processes and Files
- Communication
- Applications
- Accounting and Reporting
- Policies and Practices
- Documentation
- Support
- Mailing Lists
- UABgrid Security Announcements
- UABgrid Incident Reports
- Topic of Interest
- Resource Support pages
- General Help

Welcome to the UABgrid Development Project

Overview

The UABgrid Collaboration Environment is a vision for a user controlled collection of applications, tools, and services integrated through a uniform understanding of user identity and the communities (groups) to which the user belongs. This environment enables seamless access to shared resources and facilitates collaborations within departments, across campus and around the globe.

The UABgrid development project is a collaborative effort to build a programmatically controllable suite of services to support the construction and operation of research applications. UABgrid is the collection of these resources and the communities surrounding their development.

UABgrid is an open development effort that welcomes participation of those interested in the construction, use, or operation of this infrastructure. The project is supported by the services of the UABgrid Collaboration Environment as a demonstration of its features, as an assertion of confidence in its design, and as a method of ensuring that problems are discovered and resolved effectively. We eat our own dog food.

The computer geek in you may recognize that these themes parallel the development of distributed operating systems. Grid resources are the devices, components, and services at the core of future computer systems. The grid presents interfaces that allow developers to harness resources in ways appropriate for their domain. The platform strives for neutrality so developers can leverage select components to build the federated applications that best meet their needs. Applications can be as simple or complex as needed. Follow the plan that's right for you. Our control system for UABgrid will follow [Plan9](#).

Activities

Our initial focus has been expansion of HPC resources through grid computing. The world wide HPC community gave rise to the grid infrastructure through the [caBIG ToolKit](#) and other tools designed to harness resources distributed across many organizations. We are

dev.uabgrid.uab.edu

File Edit View History Bookmarks Tools Help

dasi

Find  next prev commit containing:  Exact  All fields

Search  Patch Tree

Diff - Old version - New version Lines of context: 3

Authors: Shantanu Pa 2011-09-13 16:01

Committer: Shantanu Pa 2011-09-13 16:01

(1) Active Tickets (4 tickets)

- List all active tickets by priority.
- Color each row based on priority.
- If a ticket has been accepted, a "\*" is appended after the owner's name

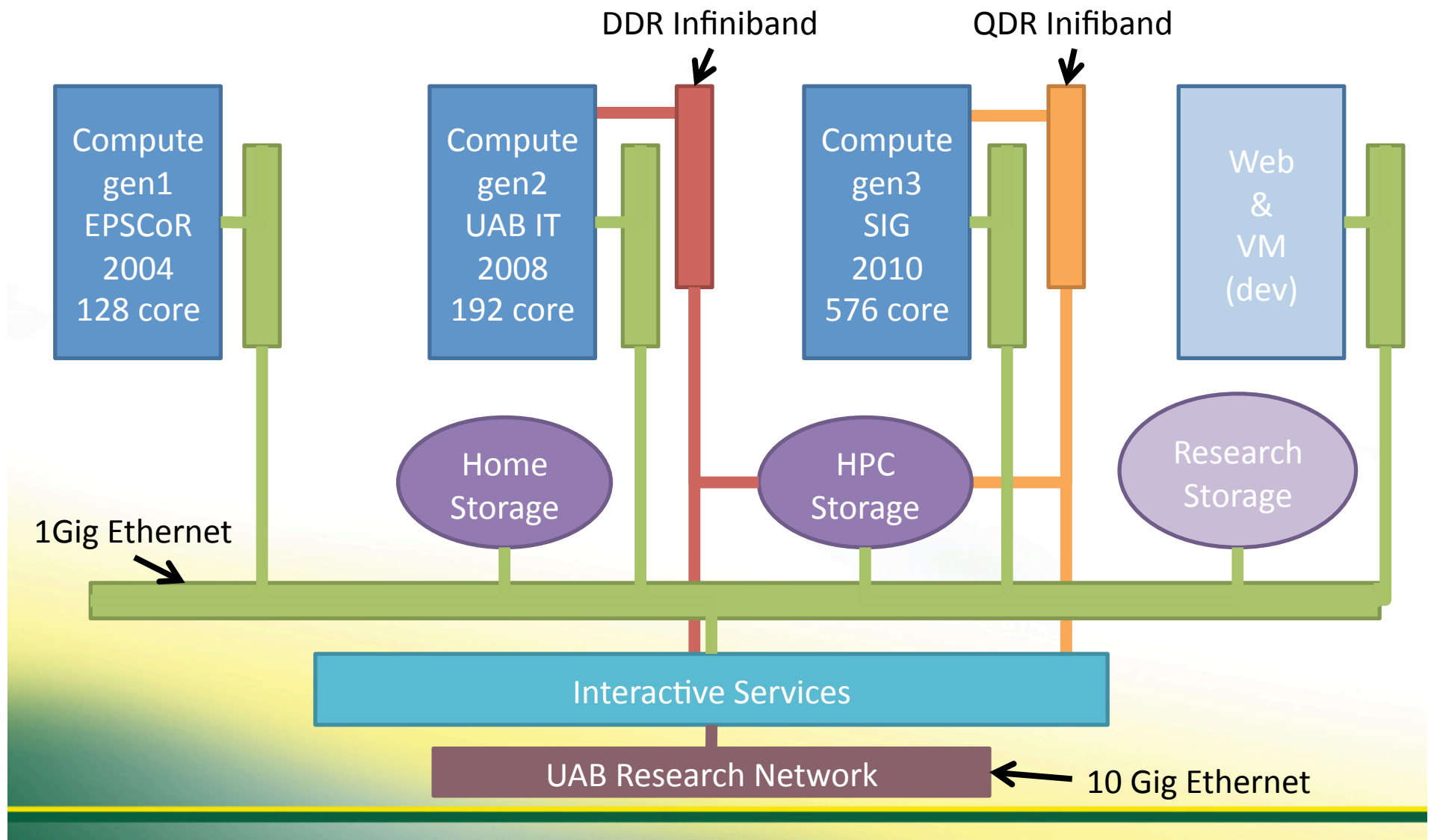
Edit report Copy report

Ticket	Summary	Component	Version	Priority
#187	UI sorting using SERVER_NAME	None	None	None
#215	Optimize DB indexing	None	None	None
#193	Enumeration Data Service	None	None	None
#102	Internal Server Error	None	None	None
#198	Race fields blank in English query for Query 26 in HITS grid stack	None	None	None
#207	delete query fails in SE	None	None	None
#236	Description of Lab Panel	None	None	None
#46	Controller/GUI: reloading query in GUI for re-submission	None	None	None
#140	Info button on query criteria	None	None	None
#141	Initial disclaimer and End User License Agreement for CollectID	None	None	None

Engage your peers. Our platforms, languages, and collaboration tools make it easy to build solutions, document code, record changes, track bugs and collaborate across organizations. Mediawiki, WordPress, mailing lists, Trac, Git, your favorite languages, and slew of other tools: we live on the same platform and run our lab that way too.



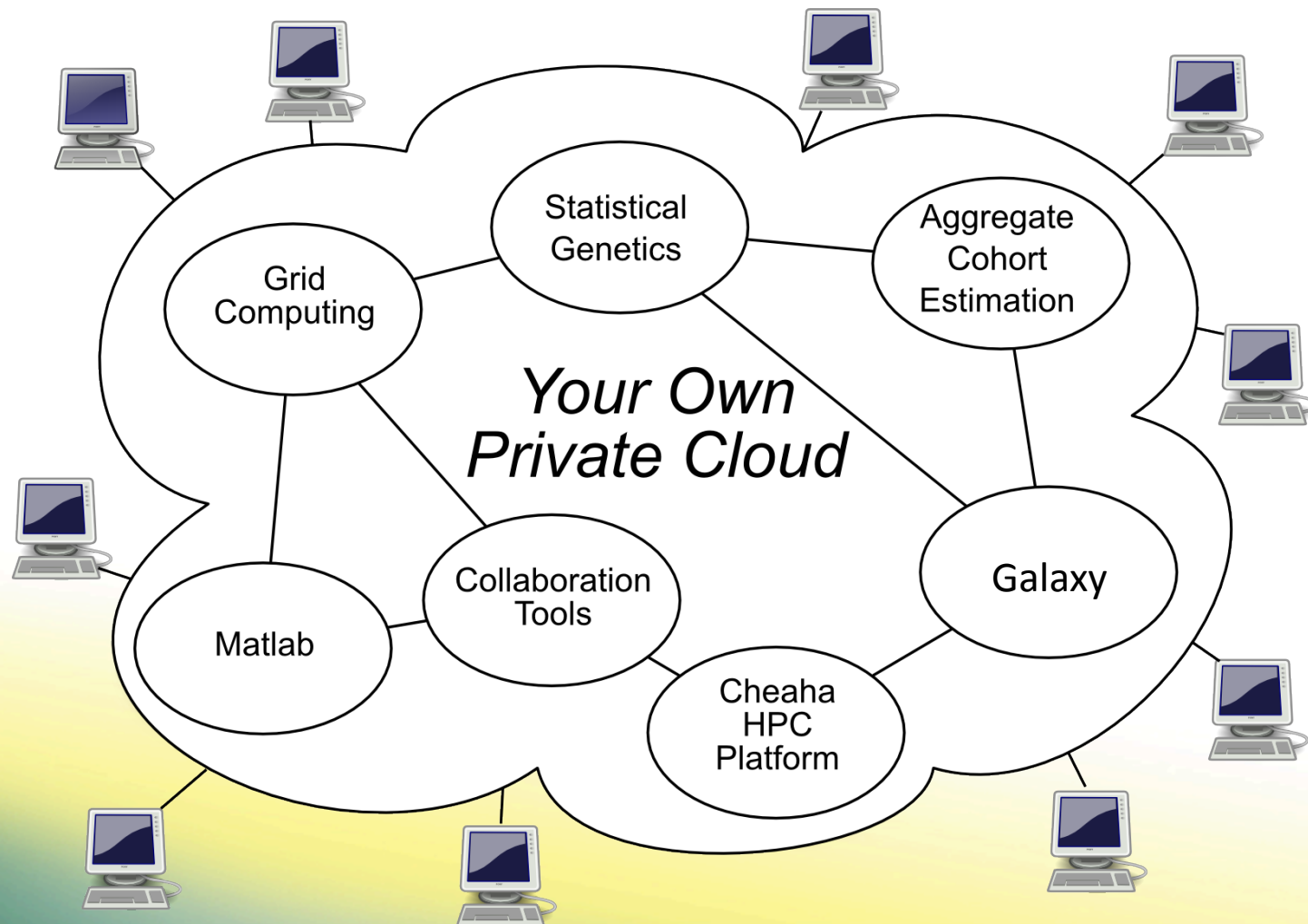
# Research Hardware Platform





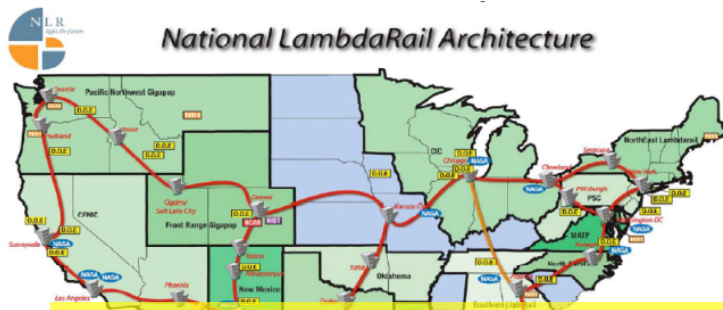
# UAB Research Cloud

*platforms, computing and storage  
to power your research*



# UAB Research Network and UASRON

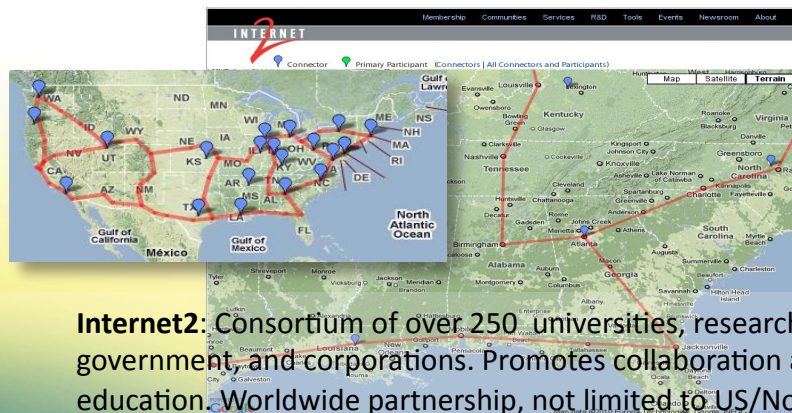
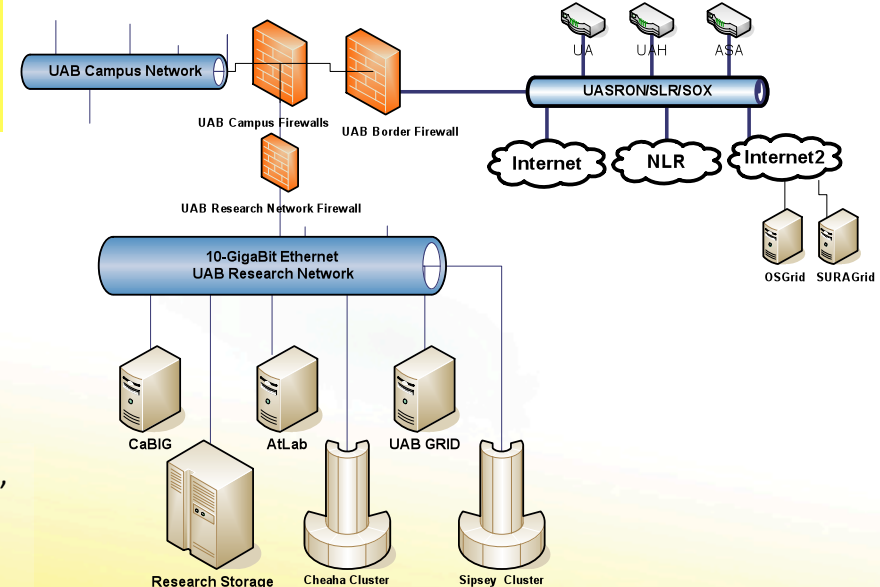
## *connections to the world*



Chan Soon-Shiong Institute for Advanced Health announces **NATIONAL HEALTH INTERNET** – \$100 million funding of National LambdaRail to bring the power of supercomputing and genomic analysis to the point of care

## UASRON: Leased Dark Fiber

A very high bandwidth lambda network that connects the UA System to National Lambda Rail and Internet2. UASRON is a collaboration between the UA System, NASA, and Southern Light Rail (SLR)@GaTech.



**Internet2:** Consortium of over 250 universities, research labs, government, and corporations. Promotes collaboration and education. Worldwide partnership, not limited to US/North America. More than 50-International Partners.

UAB Research Network

Sept, 2011

# University of Alabama System

## Regional Optical Network (RON)

- UASRON is a Privately Owned ,10-Gigabit-Ethernet, Dense Wavelength Division Multiplex (DWDM) network.
- Utilizes Leased Dark Fiber from: Level3, EarthLink, Charter Communications, and XO in Nashville.
- Connects UA System to the National LambdaRail (NLR) and Internet2 in both Atlanta and Nashville.
- Owned and Operated by the UA System.
- Utilized by UA, UAB, UAH ,NASA, Alabama Super Computer Authority, and Southern Light Rail (SLR) @GaTech.

# New NLR Partner for FY2012

- INSTITUTE FOR ADVANCED HEALTH ANNOUNCES HIGH PERFORMANCE SECURE “NATIONAL HEALTH INTRANET”
- Chan Soon-Shiong Institute for Advanced Health Provides \$100 million Funding of National LambdaRail to Bring the Power of Supercomputing and Genomic Analysis to the Point of Care



# UASRON Description

- Dense Wave Division Multiplex Network,
- DWDM is Analogous to a 40-Lane, bi-directional, Data Highway connecting UAB, UAH, UA, ASA, to Atlanta and Nashville.
- Current technology allows each lane (channel) to carry a payload of 10Gigabits-per-second.

# History

- Approved by UA Board of Trustees in 2005
- Completed initial Build-out in 2007.
- Funded by a Partnership between the UA System and NASA.
- Twenty-Year IRU Fiber Lease guarantees long term ownership , and un-restricted partnerships.
- Principal Vendors: Adva Optical, Southern Light Rail ( SLR), Level3, Earthlink, Charter Communications, Telx.

# Wavelengths ( Lambdas) in Operation

- UAB Hub provides 10 Gigabit-Ethernet to 4Nodes:  
University of Alabama ( Tuscaloosa), ASA ( Huntsville) , and SLR/SOX(Atlanta).  
Sox (Nashville) was added in Spring-2011  
SOX (Ga. Tech.)provides access to: Internet2, National Lambda Rail (NLR), and Commodity Internet.
- Other UASRON Lambdas include:
  - 3ea ASA/AREN 10G Ethernet, connecting Huntsville to Birmingham, and Huntsville to SOX-Atlanta , and Level3-Nashville.
  - 2ea UA to SOX-Atlanta
  - 1ea UAH to SOX-Atlanta
  - 2ea NASA to NASA-Atlanta
  - 3ea ORNL Local Channels in Nashville

# Future Wavelength Plans for UASRON

- SOX-Nashville: New SOX Node to be established in Nashville, Requires Local Channels for NLR, Internet2.
- UAB to SOX-Nashville for Redundancy.
- SOX-Nashville to SOX-Atlanta.



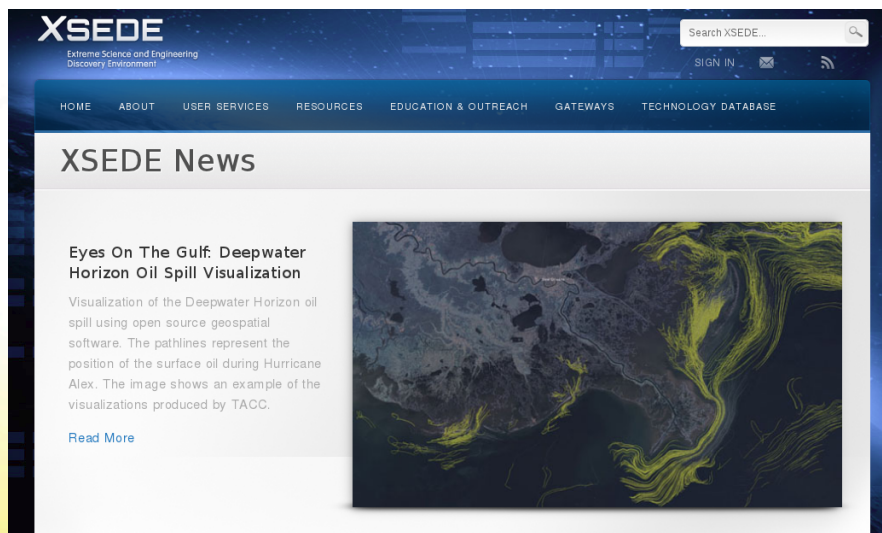
# Internet2

- Started in 1996
- Consortium of over 250 Universities, Research Labs, Government, and Corporations.
- Also supports K-20 Education Community.
- Network and Community Services for
  - Advanced network research
  - Revolutionary Internet Technologies
  - Promotes Collaboration and Education
  - Worldwide Partnership, not limited to US/North America. More than 50-International Partners.

# National Research Resources

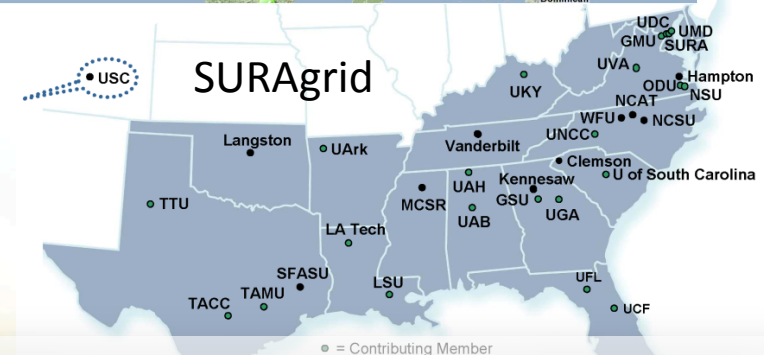
## XSEDE, Open Science Grid, SURAggrid

The Open Science Grid (OSG) advances science through open distributed computing. The OSG is a multi-disciplinary partnership to federate local, regional, community and national cyberinfrastructures to meet the needs of research and academic communities at all scales.



XSEDE is a five-year, \$121-million project supported by the National Science Foundation. It replaces and expands on the NSF TeraGrid project.

### Open Science Grid



SURAggrid is a consortium of organizations collaborating and combining resources to meet diverse users' needs: locally contributed resources, project-specific tools and environments, highly specialized or HPC access, and gateways to national and international cyberinfrastructure.

# Collaboration Across UAB

## AN ASSESSMENT OF THE IMPACT OF THE NCI CANCER BIOMEDICAL INFORMATICS GRID (caBIG®)

Report of the Board of Scientific Advisors  
Ad Hoc Working Group

March 2011

*“Perhaps the greatest impact of the caBIG program on cancer research has been to gather several communities around a virtual table to help create and manage community-driven standards for data exchange and application interoperability.”*

## Applying Open Source Lessons to Research Services Development

### The Challenge

**Advanced science demands teams of informatics and technology professionals to build services**

Talent is spread thinly across pockets of the enterprise and is difficult to harness

### The Solution

**Open source development outlines business practices that unlock distributed talent**

Engaging and coordinating individuals across organizational boundaries is key to unlocking talent needed to solve vexing scientific problems

### Open Source Patterns



#### Predictable Cycles

Predictable refinement cycles support agility, contributor engagement, and tight customer feedback loops



#### Regular Releases

Regular release schedules allow teams to mesh schedules and construct aggregate systems of services - distributions



#### Project Watersheds

Project watersheds enable projects to flow into each other and encourage software reuse

### Open Source Practices

- Provide a community building infrastructure - create active and self-supporting communities
- Open interfaces to support engagement
- Scratch Your Own Itch - collaborate to mutual benefit
- Release Early and Release Often - get into sync and make sure you are on the same page
- Transparency - enable review of project structure so others can build on your activities
- Openness Creates Ownership - avoid lock-in by enabling developers to customize solution and trust the implementation

### Further Reading

Commercial Open Source Business Model  
<http://dinkhele.com/2009/03/02/the-commercial-open-source-business-model/>

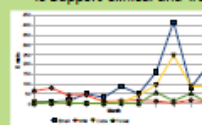
Characteristics of Open Source Projects  
<http://portal.acm.org/citation.cfm?id=973581>

Open Source Watershed  
<http://tarnier.org/docs/shawcraft-thesis.pdf>

Free Software Project Management HOWTO  
<http://fdp.org/HOWTO/Software-Proj-Mgmt-HOWTO/index.html>

### Successful Teams

**Build A Patient Cohort Discovery Service To Support Clinical and Translational Science (CTSA)**



Project Activity  
11th Month with 1st Stable Mark



Distributed Talent  
Research, Central IT, Hospital IT

The UAB team leveraged collaboration tools standard with open source projects to coordinate talent across the enterprise, providing transparency to support project decision makers

We defined a development schedule centered on three one-month cycles. Three cycles (9 months) equals a stable release mark. The cycles provide a consistent framework to measure progress and organize effort

The code is accessible to all team members with an open test infrastructure encouraging developer engagement

### Successful Tools

**Support Spontaneity:** The tools for this generation are expected to scale from napkins to networked collaboration environments. Group messaging (email lists) is the communication backbone; Mediawiki/wiki captures group knowledge; Wordpress blogs capture personal insight; Subversion and Git record the history of development; and Trac helps manage issues

**Supply Coherent Services:** We are transitioning our pilot campus collaboration and grid infrastructure into application services directly controlled by research groups to support their own projects



Presented at EDUCAUSE Southeast Regional Meeting 2010  
Thursday, June 3, 2010, Atlanta, Georgia  
John-Paul Robinson (jpr@uab.edu), Lead System Programmer  
UAB IT Research Computing  
University of Alabama at Birmingham

# IT Research Computing Team



**Bob Cloud**  
Executive Director  
Infrastructure Services  
UAB IT



**David Shealy**  
Director  
Research Computing  
Physics / UAB IT



**Puri Bangalore**  
HPC Consultant to  
Research Computing  
CIS / UAB IT



**John-Paul Robinson**  
System Architect  
Research Computing  
UAB IT



**Mike Hanby**  
Information Systems  
Specialist  
Engineering / UAB IT



**Poornima Pochana**  
Programmer/Analyst  
Research Computing  
UAB IT



**Shantanu Pavgi**  
Programmer/Analyst  
Research Computing  
UAB IT

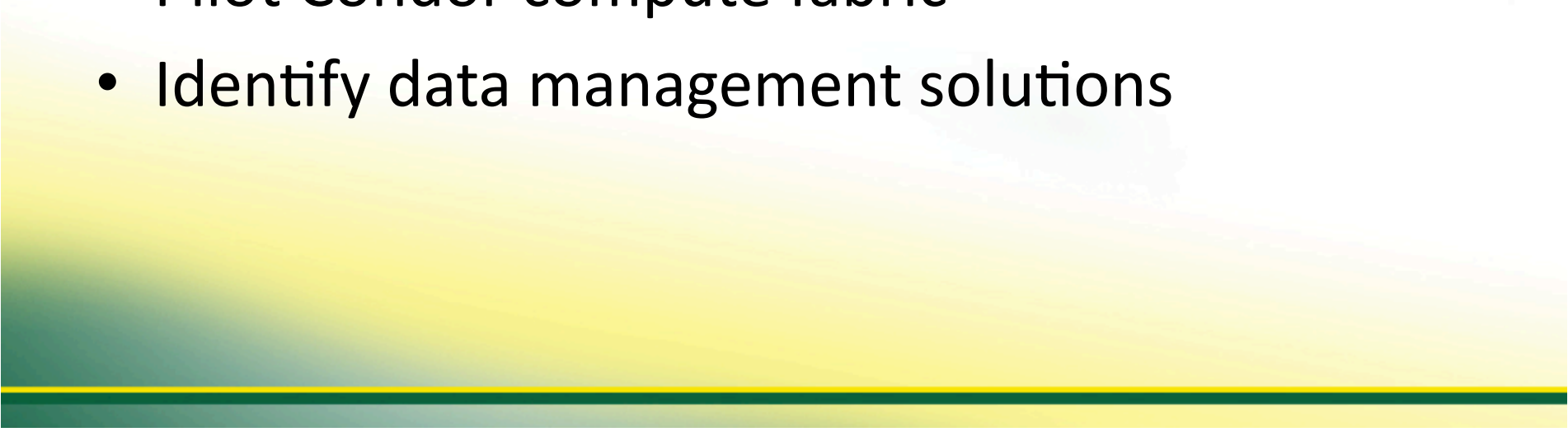


**Bill Bradley**  
System Programmer  
Infrastructure Services  
UAB IT

Engaged professional staff collaborate with investigators to maximize the research impact



# Current Projects

- Documentation!
  - Enhance platform features
  - Grow MATLAB use and support
  - Build out NGS tools and services
  - Pilot Condor compute fabric
  - Identify data management solutions
- 

# **Proposed Plans**

## **UAB Research Computing Data Storage**

- Users: PI's of grants, users of HPC environment (Cheaha), and Group Access
- Target initial maximum allocation of 1TB
- Access Technologies: BlazerID authenticated; Network attached; WebDAV, Drop Box, & SFTP
- Data Retention: indefinite retention online; archive to tape or other media is available for a fee upon request.