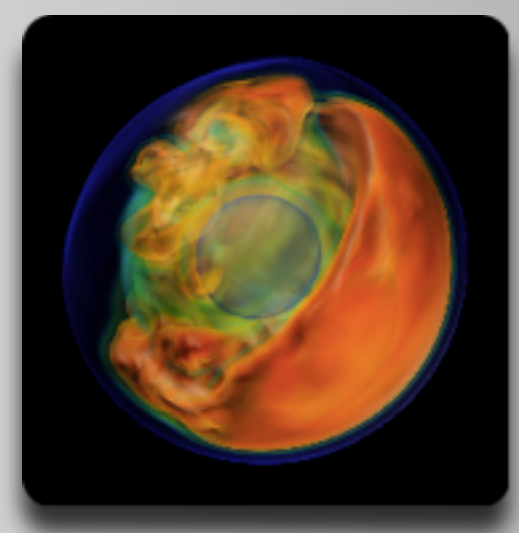
The University of Tennessee Center for Remote Data Analysis and Visualization (RDAV)

Sean Ahern, PI – University of Tennessee
Jian Huang, co-I – University of Tennessee
Wes Bethel, co-I – Lawrence Berkeley National Laboratory
Scott Klasky, co-I – Oak Ridge National Laboratory
Dave Semeraro, co-I – National Center for Supercomputing Applications (NCSA)
George Ostrouchov, Senior staff – Oak Ridge National Laboratory
Miron Livny, Senior staff – University of Wisconsin

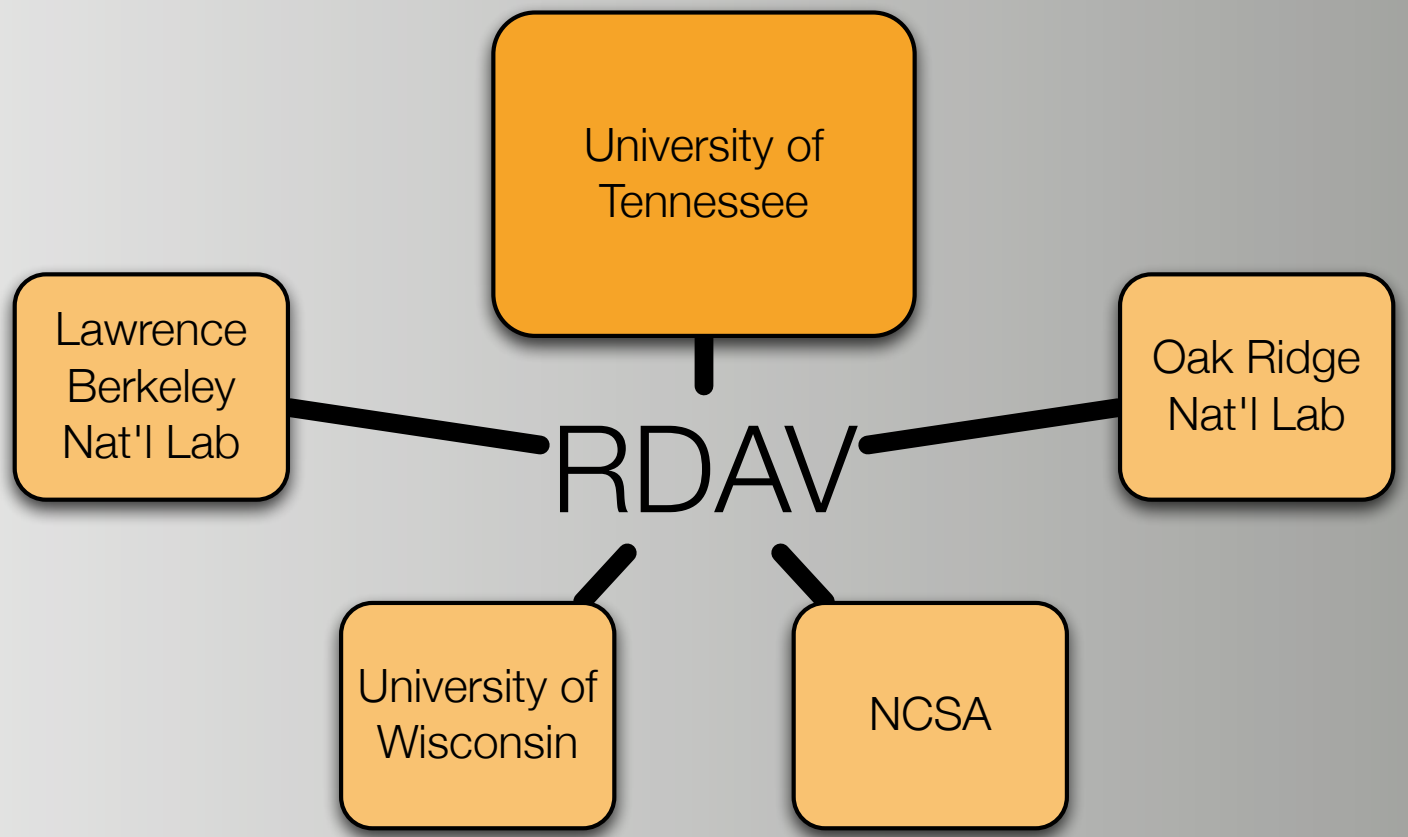
Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

Providing analysis services for TeraGrid XD users

- Provide remote and shared resources for the purpose of exploring/analyzing/ visualizing large scale data.
- Provide the ability to easily take advantage of remote and shared computing/ data storage infrastructure.
- Provide unique architecture for data analysis and visualization
- Leverage large amount of existing experience in deploying similar capabilities.
- Allocated through TRAC

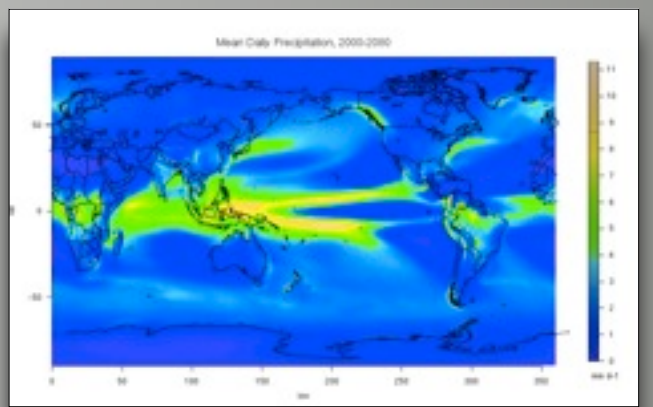


RDAV is a partnership between 5 institutions



We elicited representative user requirements

- Partnered with representative users, explored representative use cases
  - HPC simulation codes (ENZO astrophysics simulation)
  - Observational sources (distributed sensor networks)
  - Experimental data source (Spallation Neutron Source)
  - Non-traditional analysis (Document clustering)



- September 2008 User Survey
  - Surveyed 320 principal investigators of current NSF awards for projects involving data analysis or visualization with budgets of $500,000 or greater. 41 responses over the next 5 days:
    - Dataset sizes: from a few petabytes down to the 20- to 100-GB range (most respondents fell into the latter range). The high-energy physics community has the largest data requirements, both experimental and simulated. Other significant communities include earth sciences, astrophysics, and material sciences; these respondents expected significant data growth in the next few years.
    - Sharing/moving: Most move and share data, both post-processing and real-time analysis. The two most frequent data types were regularly gridded data and graph structure data.
    - Remote visualization: Many respondents expressed this need

Diverse use cases dictate unique architecture

- Many HPC users can use distributed memory analysis
  - data parallel, time parallel
- However, many general and statistical analysis algorithms favor large shared memory
  - Document clustering/searching
  - Generalized graph structures
  - Bioinformatics, genomics
  - ...
- *Large shared memory* is the only reasonable way to address all of these needs
- SGI UltraViolet architecture provides:
  - Large memory single-system image through NUMA
  - A “better” cluster architecture, accelerating distributed memory MPI

A large SMP is central to RDAV's hardware

- SGI UltraViolet system
  - 1,024 cores (Intel Nehalem EX)
  - 4 TB Global Shared Memory
  - 8-16 NVIDIA Fermi Tesla GPUS – "S" config
- ~1 PB shared filesystem
- ~30 GB/s bandwidth

Systems housed at National Institute for Computational Sciences (NICS)

- NICS is a collaboration between UT and ORNL
- Awarded the NSF Track 2B ($65M)
- Phased deployment of Cray XT systems
- Home of Kraken, currently #3 on Top 500

Early UltraViolet Test Systems

- SGI has delivered a UV test unit to ORNL
- System configuration:
  - 96 Nehalem EX cores
  - 96 GB memory
  - 2x 10 Gigabit Ethernet
  - 2x QDR Infiniband
  - 1x NVIDIA graphics card
- Beginning friendly user access in the next two weeks

Early UltraViolet Test Systems

- SGI has delivered a UV test unit to ORNL
- System configuration:
  - 96 Nehalem EX cores
  - 96 GB memory
  - 2x 10 Gigabit Ethernet
  - 2x QDR Infiniband
  - 1x NVIDIA graphics card
- Beginning friendly user access in the next two weeks

- Second test system will be delivered in early May.
- System configuration
  - 256 Nehalem EX cores
  - 512 GB memory
  - 2x 10 Gigabit Ethernet
  - 2x QDR Infiniband
  - 2x NVIDIA GPUs

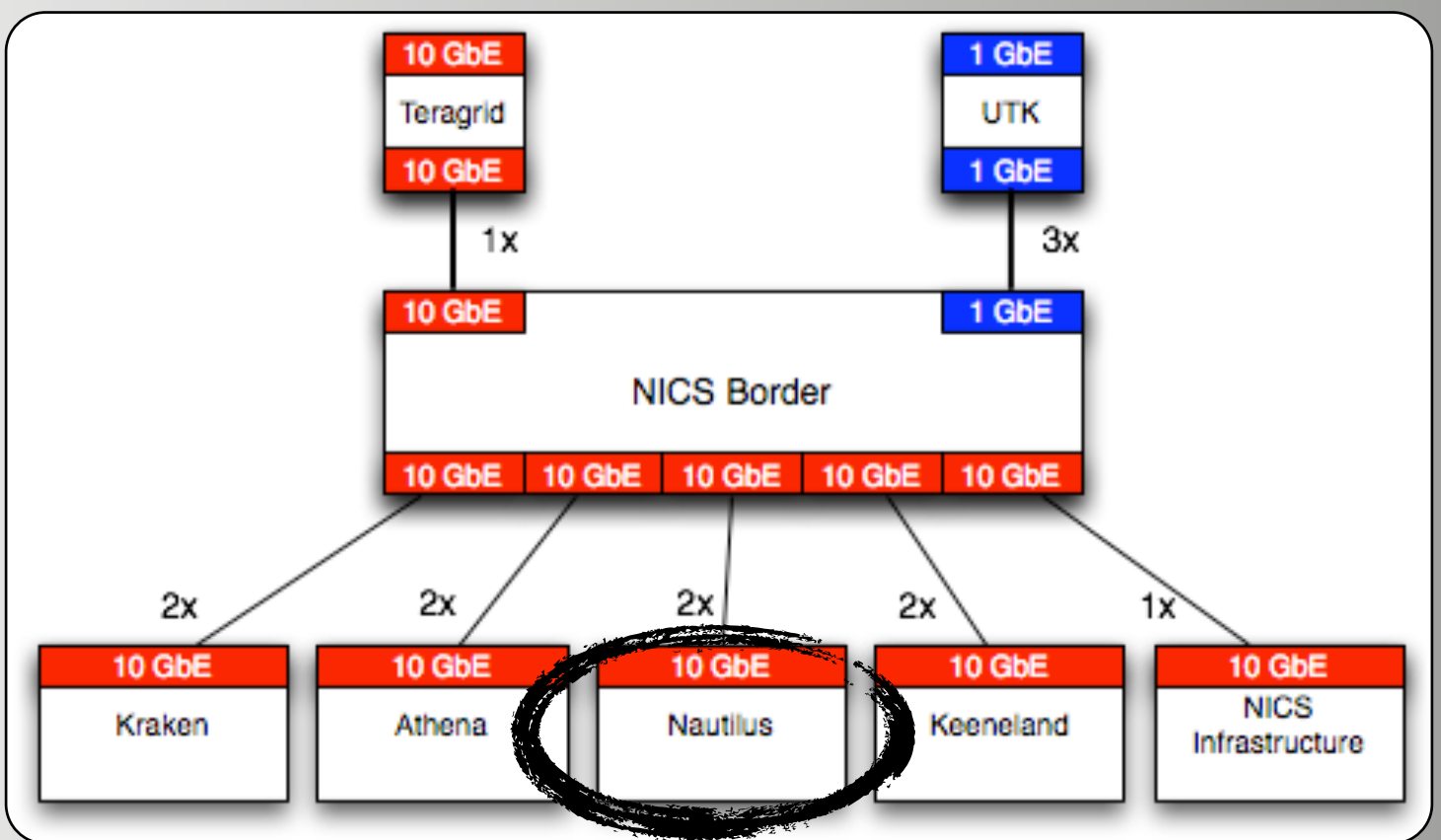
Final UV System – Nautilus

- The final UV system is expected to ship in early June. This will be a forklift upgrade (i.e. the early test system will be physically removed).
- System configuration of Nautilus:
  - 64 blades
    - 128 sockets/1024 cores
    - 4 TB memory
    - SLES 11
  - 4x 10 Gigabit Ethernet
  - 24x QDR 4x Infiniband
  - 8-16 NVIDIA Fermi Tesla
  - ~1 PB parallel file system
- Must be accepted and in production by early July

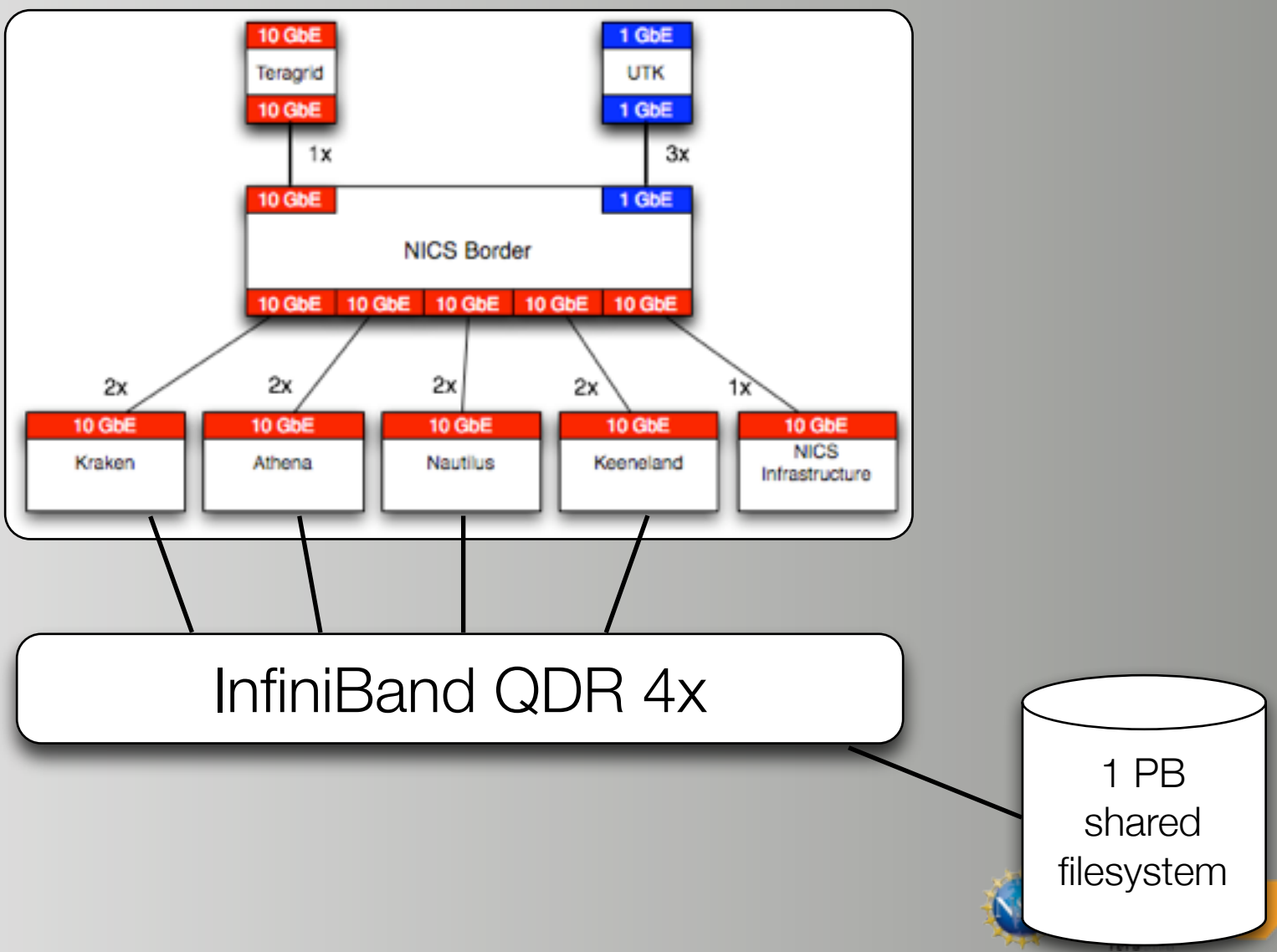
NICS Support for RDAV

- System Administration
  - Installation and management of Nautilus
  - Integration with NICS infrastructure including file systems, archival storage, networking, security, accounts and accounting, etc.
  - Integration with TeraGrid/XD
- User Support
  - Front line support of users including tickets
  - User surveys
  - User Advisory Committee

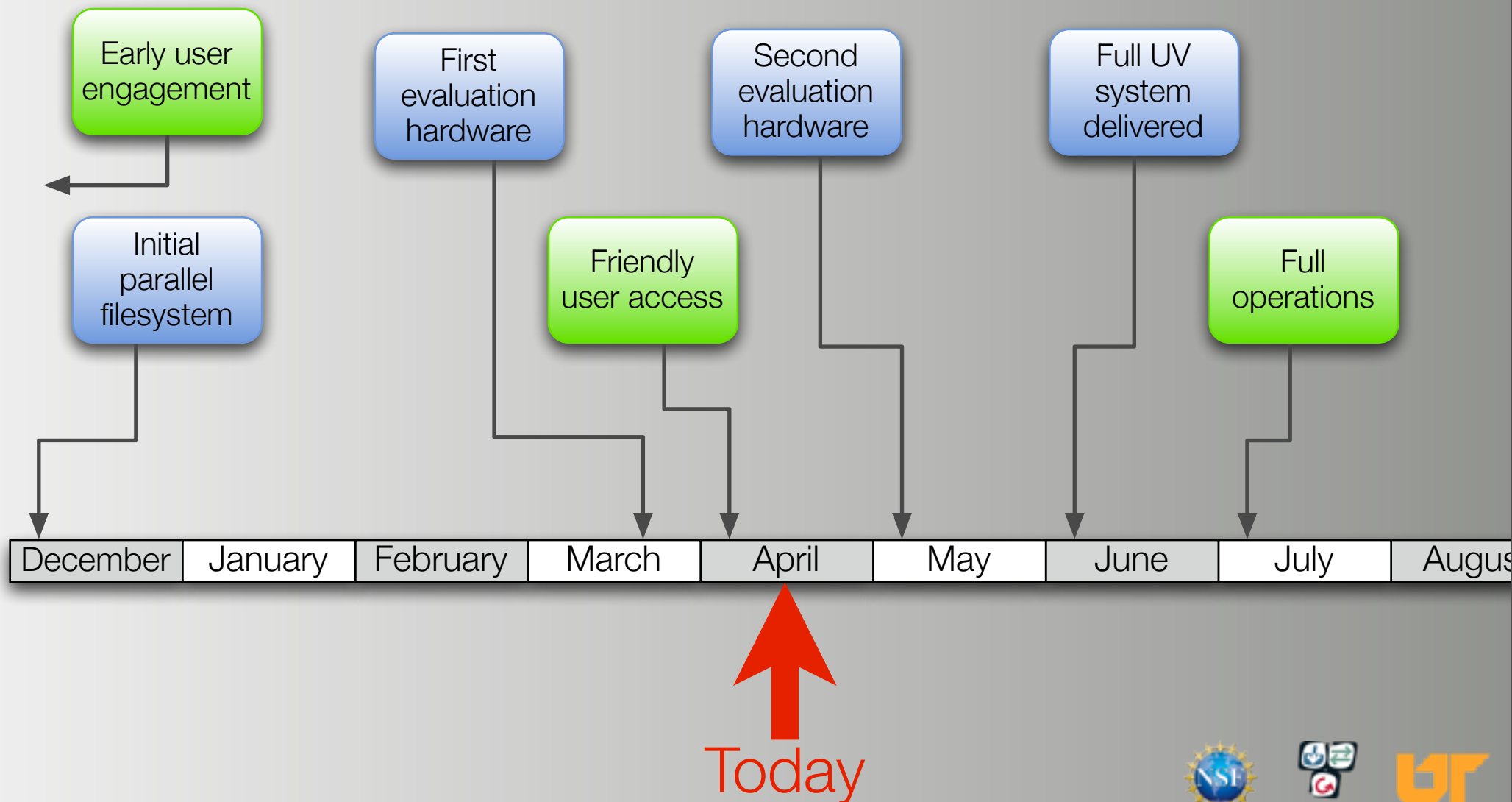
How Nautilus Fits Into NICS



How Nautilus Fits Into NICS



Our deployment timeline



RDAV provides User Services

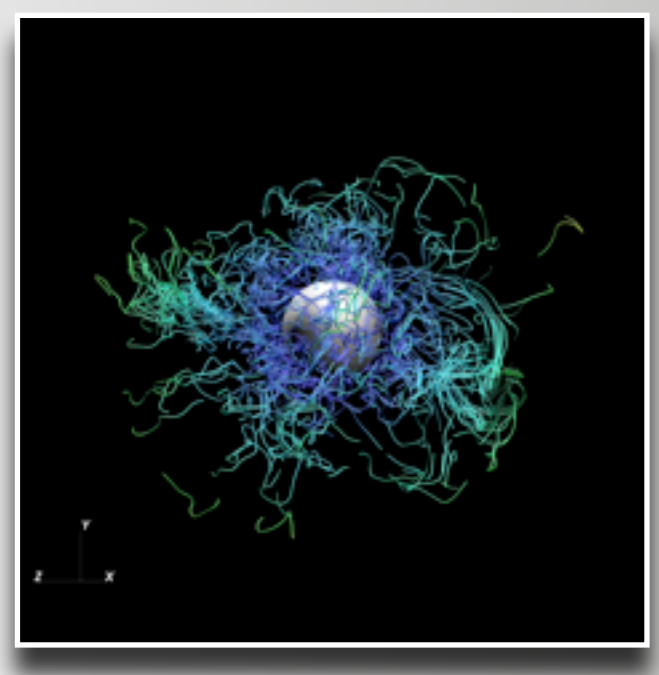
- Routine user services staff – Connected to NICS
  - Routine user services
  - Ticket triage and routing to specialist
- Specialized staff for Advanced Support for TeraGrid Applications (ASTA)
  - Specific for remote visualization, data analysis, workflow services, portal
  - Educates on effective use of existing tools or on custom development
  - Provides individualized assistance for center-wide software
- Dedicated staff for education, outreach, and training
- Dedicated staff for tool discovery, certification, integration

RDAV provides a range of software services

- Analysis applications: to be dictated by user needs and technology needed to solve user problems.
  "Whatever it takes!"
- Remote visualization and image generation
  - Provide interactive and batch image generation tools. (gnuplot, ImageMagick, etc.)
  - Remote parallel visualization (VisIt, ParaView, etc.)
  - Tools for custom application development
- Data analysis and statistical analysis
  - Octave, Parallel R, Matlab, etc.
- Workflow systems
  - DAGMan system automates batch actions on behalf of users
  - Infrequent current use, however, value is increasing and many users wish to explore.
- Dashboard delivery
  - Leverage DoE funding for eSimMon dashboard system.
- Portal system
  - Builds upon standard Liferay platform
  - Provides SAS services for analysis and visualization

RDAV will provide scalable visualization tools

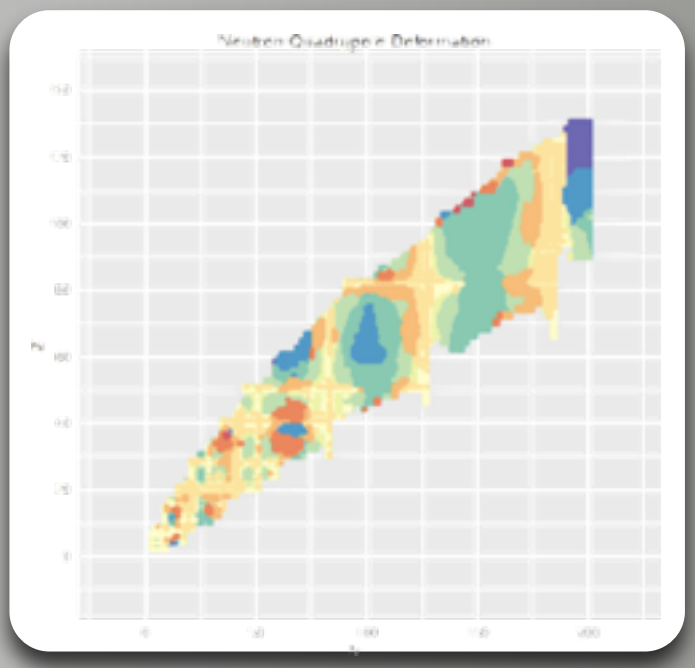
- TeraGrid standard visualization software stack
- Scalable tools (VisIt, ParaView)
- Low-end tools (gnuplot, ImageMagick)



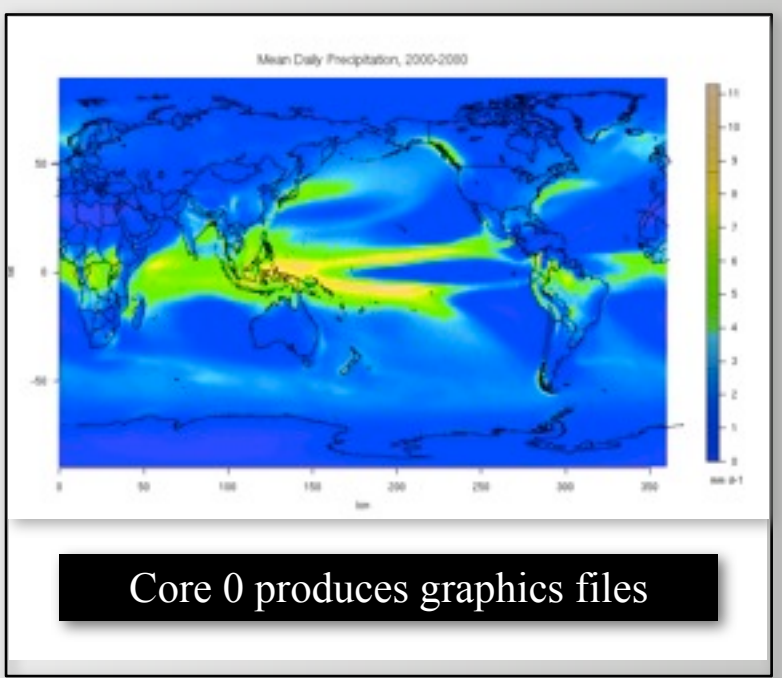
- Remote “vis” delivery software:
  - Some high-end packages already have a client-server architecture, so those can be used in "remote vis mode" today with no special 3rd party software.
  - Some packages have GUIs and displays that we'll want to provide remote access to. We can do so using either protocol accelerators (NX, VirtualGL) or remote desktop software (NX, VNC). There is some overlap in capability/functionality between these two approaches.

RDAV will provide general analysis software

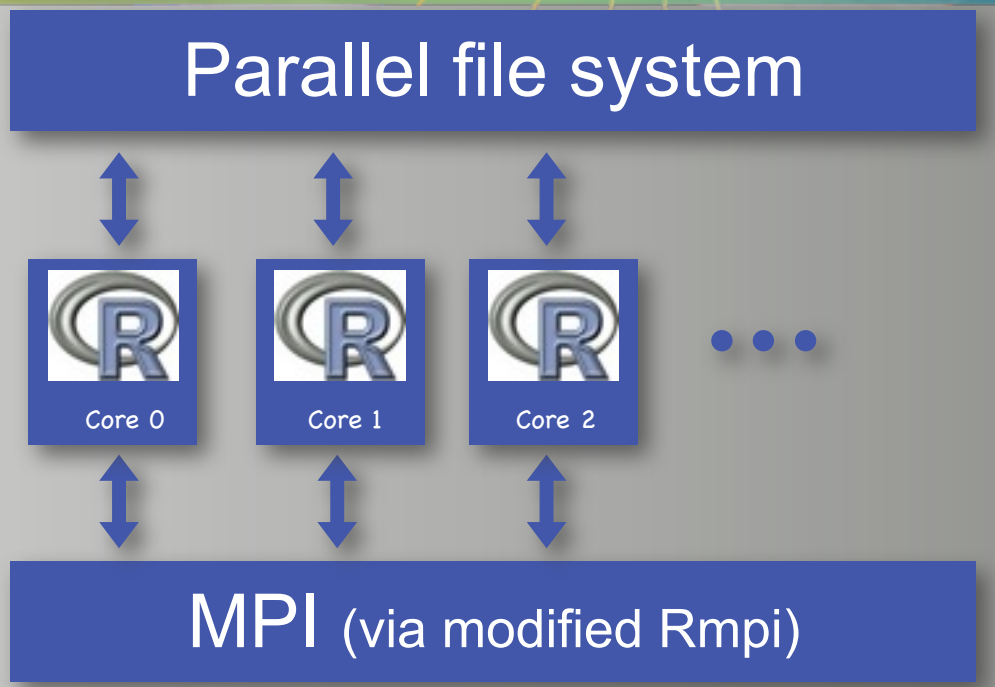
- MATLAB and Octave
- R, R-MPI, R+GPU
- Python
- IDL



RDAV will provide parallel statistical analysis tools



Core 0 produces graphics files



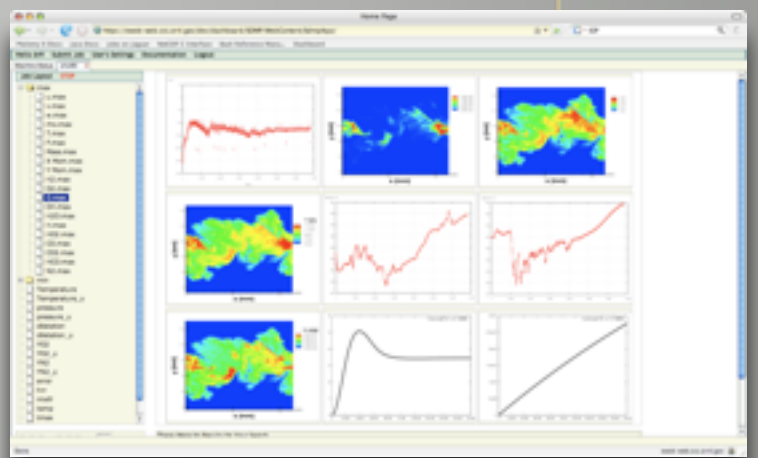
- Data parallel R leverages success of scalable visualization
- Widely-used open source statistical analysis

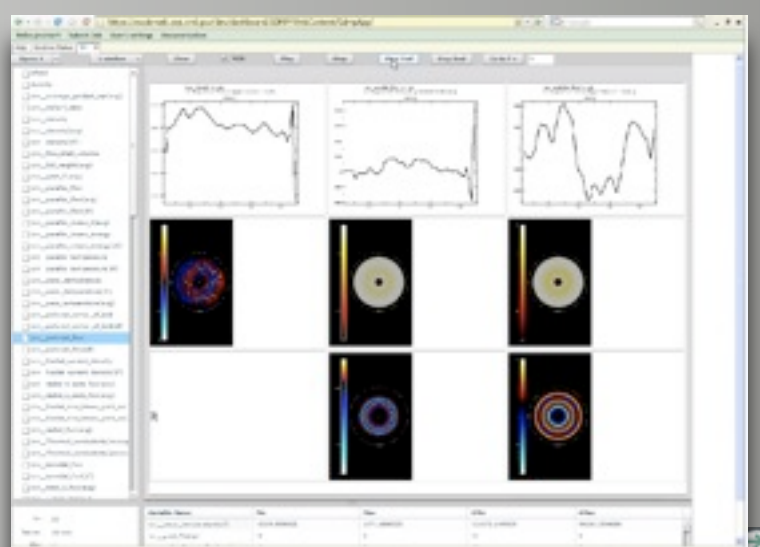
RDAV will provide portals for science gateways

- Portals give developers a chance to create collaborative environments that allow researchers or any community to share knowledge, analyze data, and solve problems
  - Common area to combine users, resources, services
    - Portlets for chat, email, forums, group invites by email, etc.
    - Plug-ins to provide rich services for research communities.
    - Custom portlets to access backend services
- Accessible through any browser.
- Using the Liferay framework, adopted by TeraGrid as a whole.

RDAV will provide a dashboard for monitoring

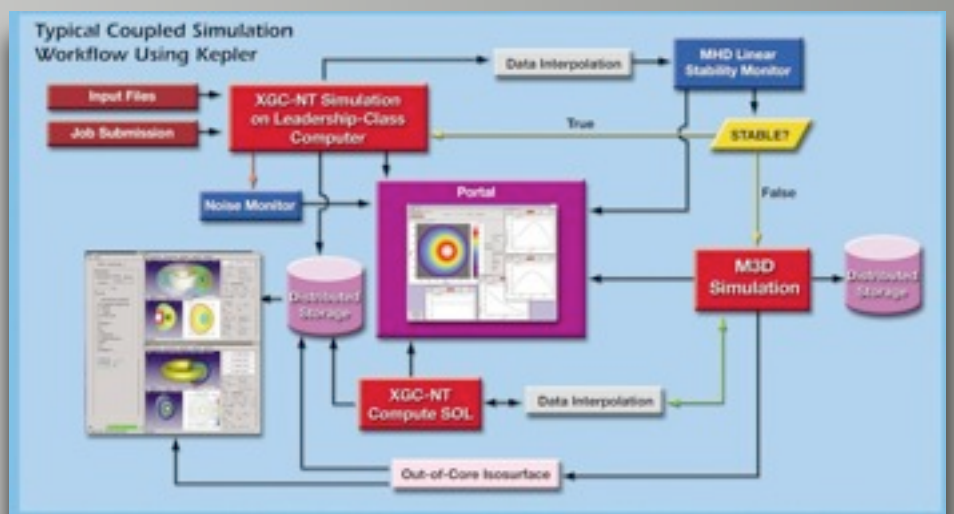
- Leveraging DOE investment in eSimMon
- Dynamic Front end:
  - Machine monitoring: standard web technology + AJAX
  - Simulation monitoring: Flash
- Back end: shell scripts, python scripts and PHP
  - Machine queues command
  - Users' personal information
  - Services to display and manipulate data before display
- Storage: MySQL (queue-info, min-max data, users' notes…)
- Wrapped in LifeRay





RDAV will provide workflow management tools

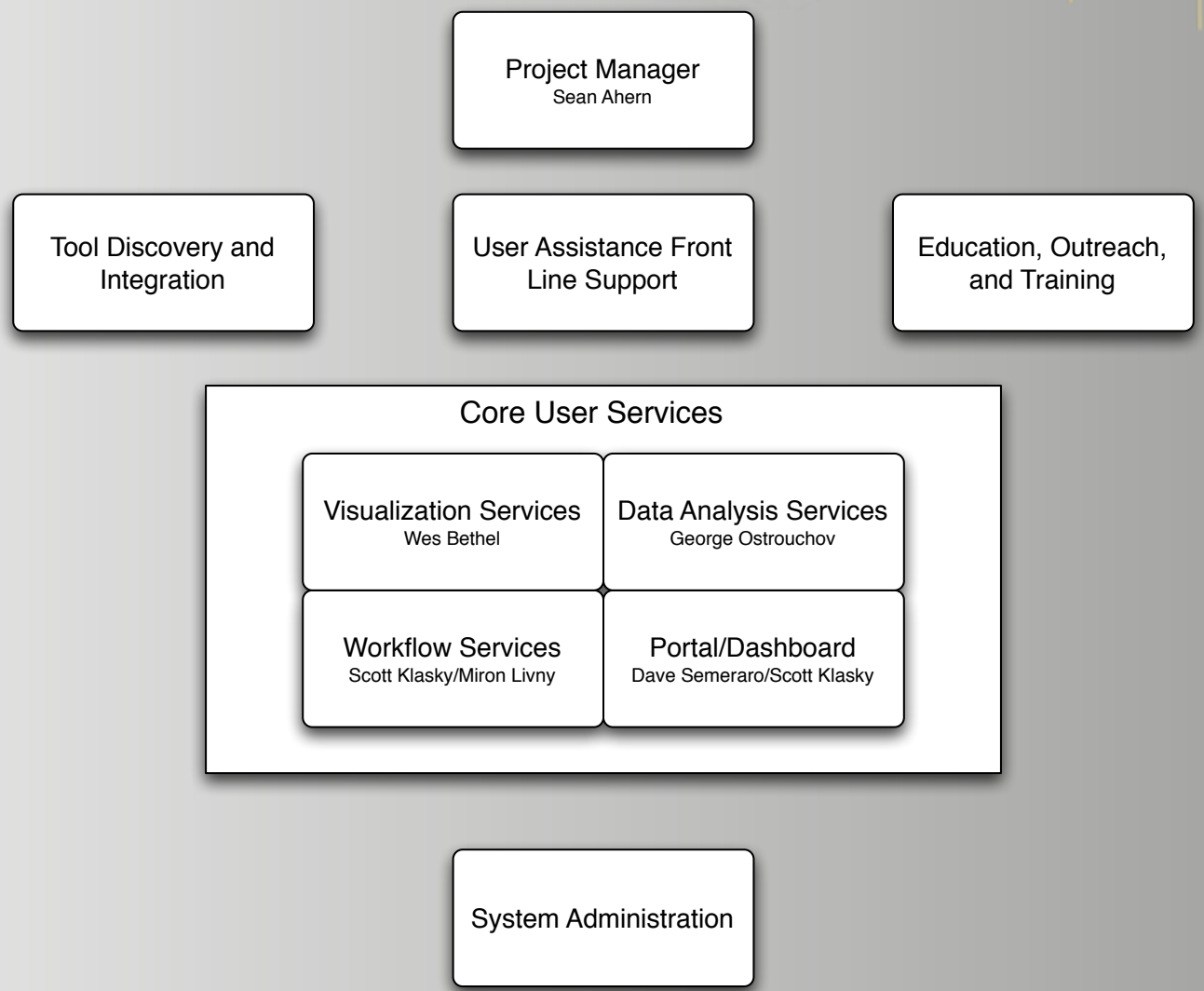
- Help in the construction and automation of scientific problem-solving processes that include executable sequences of components and data movement.
- Scientific workflow systems often need to provide for load balancing, parallelism, and complex data flow patterns between servers on distributed networks.
  - Aiming to solve complex scientific data integration, analysis, management, visualization tasks
  - Error checking & retry
  - Maximizes compute resources / human time
  - “Launch and Forget!”
- Deployed through DAGMan or Kepler



We will provide tiers of software services

- Turnkey tools for data analysis and exploration
  - Tools like VisIt, Octave, R, and ParaView
- Programming environment for the development of home-grown analysis applications
  - Compilers, VTK, OpenDX, Mesa, Chromium
- Portal-based “software as a service” capabilities
  - Allow for the creation of domain-specific community portals.
  - Expected to provide analysis and visualization elements of TeraGrid XD-wide portals

RDAV organizational structure



RDAV Analysis: Three Early Users

- Astrophysics: Supernovae simulation
  - Bronson Messer (UTenn/ORNL)
  - Parameter studies, high-D data
  - VisIt automation
- Experimental Data: Spallation Neutron Source
  - Density estimation from time of flight data (instead of binning/ histogram)
  - Background subtraction in streaming data (and otherwise)
  - 5d to 2d/3d projections
  - VisIt optimization
- NIMBioS: All Taxa Biodiversity Inventory (ATBI)
  - Lou Gross (UTenn)
  - Inventory 100,000 species of living organisms in Smokies
  - Develop novel algorithms for knowledge discovery from ATBI data
- Barry Schneider: "Different customers than the 'cycles' group."