The Texas Advanced Computing Center at The University of Texas at Austin: UT Grid Design & Implementation

November 2004

NSF Middleware Initiative (NMI) Integration Testbed Case Study Series

Series contact: Mary Fran Yafchak, Southeastern Universities Research Association, maryfran@sura.org.

The NMI Integration Testbed Program provided practical evaluation of NMI components within the context of real projects and application scenarios from June 2002 through November 2004. During that time, NMI Testbed sites collectively submitted over 220 evaluation reports to middleware component developers as direct feedback into the NMI development cycle. Site representatives also actively inspired, promoted and facilitated the integration of middleware throughout their institutions.

The NMI Integration Testbed Case Study Series documents the most significant influences and outcomes of NMI Testbed sites' middleware integration efforts, highlighting intersections with established projects, application contexts and influences, drivers for innovation, decision points and challenges. Through this documentation, the work of these pioneering institutions is captured to provide a breadth of insight and approaches for others to use towards successful middleware development and deployment.

This NMI Integration Testbed Case Study Series is sponsored by the National Science Foundation Middleware Initiative-Enterprise and Desktop Integration Technologies (NMI-EDIT) Consortium of EDUCAUSE, Internet2, and SURA. Additional support was provided by the National Science Foundation Cooperative Agreement NSF 02-028, ANI-0123937.

Copyright © 2004 The Texas Advanced Computing Center at The University of Texas at Austin (TACC). TACC permits use of this content for noncommercial purposes with proper attribution. All rights reserved.

Executive Summary

The Texas Advanced Computing Center (TACC) located at The University of Texas at Austin (UT Austin) supports the advanced computing needs of researchers on campus and is an active contributor and participant in the grid computing area. TACC leverages their grid expertise so that their researchers have access to the richly diverse computing resources in grid-based computing infrastructures.

Since March 2004, TACC has been bringing their grid computing skills and experience to bear on building a campus grid ("UT Grid"). TACC is leading the campus-wide effort that will extend the benefits of grid computing throughout the UT Austin user base. Initial users of UT Grid will be campus researchers, but the long-term goal is to offer UT Grid services to support the needs of all users on campus.

The computational, visualization and storage computing resources of grid environments are, of course, found in non-grid computing environments, but the grid environment enables them to be connected and used together in a way that greatly increases their ready availability. Research endeavors benefiting from grid resources at UT Austin include work in computational fluid dynamics and sharing of data analyses from high-resolution X-ray computed tomography of biological specimens.

UT Grid is being deployed in phases. The first phase will make TACC's computing services available through a hub and spoke design. Future phases will add data, storage and visualization services, while enhancing existing services with additional functionality. During these deployment phases, NMI software will be leveraged to facilitate sharing and scheduling of compute resources across the UT campus.

TACC has worked extensively with NMI software components and is frequently one of the early users of NMI software components. TACC has also developed components that are part of the current NMI release. TACC enhanced their experience with NMI components by participating in the NSF-funded NMI Integration Testbed Program[1]. Their Testbed participation has expanded TACC's experience with NMI components, while sharing experiences and ideas with colleagues from other institutions working within the NMI Integration Testbed.

The maturity of grid technologies like those in the NMI component suite, and the willingness of organizations like TACC to share their knowledge and experiences with others, make partial or full grid deployments attainable for less experienced organizations. The building of UT Grid is a

[1] As part of its overall effort to develop and disseminate software that lets scientists and educators share resources across the Internet, NMI has begun a practical deployment and evaluation effort called the NMI Integration Testbed. Managed by the Southeastern Universities Research Association (SURA) on behalf of the NMI-EDIT Consortium, the testbed consists of eight universities that participate in a closely coordinated effort to deploy and evaluate NMI technologies. http://www1.sura.org/3000/NMI-Testbed.html

powerful example of the benefits that grid technologies like NMI components have when they are implemented by an experienced and user-focused organization.

For information about TACC and UT Grid Design and Implementation, contact Tina Romanella de Marquez at tina@tacc.utexas.edu.NMI

Components Highlighted in this Case Study

The NMI components in this case study series encompass NMI Releases 1 through 4. Information about NMI Releases can be found on http://www.nsf-middleware.org/.

Condor-G

The GRIDS Center's Condor-G is a computation management agent for the Grid. Condor-G is the marriage of technologies from the Condor project and the Globus project.
Home site: http://www.cs.wisc.edu/condor/; Globus (see below)

Globus

The GRIDS Center's Globus Toolkit is an open-source collection of modular technologies that simplifies collaboration across dynamic, multi-institutional virtual organizations. It includes tools for authentication, scheduling, file transfer and resource description.
Home site: http://www-unix.globus.org/toolkit

GridPort

The GRIDS Center's GridPort enables the development of portals and applications on top of underlying distributed and grid computing infrastructure to facilitate computational science.
Home site: http://gridport.net/index.cgi

Grid Packaging Tools (GPT)

The GRIDS Center's Grid Packaging Tools (GPT) is a collection of packaging tools built around an XML-based packaging data format. …. The tools provide a means for developers to easily define the packaging data and include it as part of their source code distribution.
Home site: http://www.gridpackagingtools.org/

GSI-OpenSSH

GSI-OpenSSH is a modified version of OpenSSH that adds support for GSI authentication, providing a single sign-on remote login capability for the Grid. GSI-OpenSSH can be used to login to remote systems and transfer files between systems without entering a password, relying instead on a valid GSI credential for operations requiring authentication.
Home site: http://grid.ncsa.uiuc.edu/ssh/

MyProxy

The GRIDS Center's MyProxy is a credential repository for the Grid. MyProxy provides a set of flexible authorization mechanisms for controlling access to the repository.
Home site: http://grid.ncsa.uiuc.edu/myproxy/

Network Weather Service

The GRIDS Center's Network Weather Service (NWS) is a distributed system that periodically monitors and dynamically forecasts the performance various network and computational resources can deliver over a given time interval.
Home site: http://nws.cs.ucsb.edu/

Storage Resource Broker

The San Diego Supercomputer Center (SDSC) Storage Resource Broker (SRB) is client-server middleware that provides a uniform interface for connecting to heterogeneous data resources over a network and accessing replicated data sets. SRB, in conjunction with the Metadata Catalog (MCAT), provides a way to access data sets and resources based on their attributes and/or logical names rather than their names or physical locations.
Home site: http://www.sdsc.edu/srb/

UberFTP

UberFTP is the first interactive, GridFTP-enabled ftp client. It supports GSI authentication, parallel data channels and third party transfers.
Home site: http://dims.ncsa.uiuc.edu/set/uberftp/

Texas Advanced Computing Center at The University of Texas at Austin: UT Grid Design & Implementation

The Texas Advanced Computing Center (TACC) located at The University of Texas at Austin (UT Austin), supports the third-largest campus user community in the United States. Since March 2004, TACC has been building a campus grid ("UT Grid") to integrate the numerous, diverse resources on campus into a comprehensive cyberinfrastructure for research and education (1). UT Grid will make grid-based services available to the campus's approximately 65,000 computer users. In such a large user community, there are many types of computer users with a vast range of computing needs. Yet the UT Grid can meet most of these needs.

This article looks at key aspects of UT Grid, with a view to drawing lessons important for the grid community in general. The organizations and programs that are part of the UT Grid fabric are discussed, including general lessons learned. Most important in the design of UT Grid was the question of how the specific grid components could enhance the work of UT researchers and other UT computer users. We emphasize in particular the needs of users that dictated the use of specific NMI grid components and reflected the influence of TACC's participation in the National Science Foundation Middleware Initiative (NMI) Integration Testbed program.

UT/TACC Background in Grid Computing

The deployment of a comprehensive cyberinfrastructure such as UT Grid requires extensive expertise and experience. TACC has long supported the advanced computing needs of researchers on campus and in this role has participated in several grid computing projects. These projects gave TACC staff members the skills necessary for building UT Grid.

The University of Texas at Austin (UT Austin) is one of the nine pioneering institutions that are members of the TeraGrid [i] project, the National Science Foundation's ambitious effort to build a national grid for advanced computing. TACC actively participated in defining, developing, and deploying the components of a common software stack for each resource type (storage, computation, and visualization) on the TeraGrid. TACC has also installed a number of grid technologies, including NMI components, across its complement of supercomputing hardware in preparation for the startup of production operation of the TeraGrid. TACC is currently exploring such NMI technologies as the GRIDS Center's

Condor-G[2] to create potential scheduling solutions for use in the TeraGrid.

TACC has long been an active contributor and participant in several other efforts in the grid computing area. For instance, TACC is the developer of the Grid Portal Toolkit (GridPort) (2), a component of the NMI-Release 5 suite, which has been used extensively to build user and application portals in several national grid projects. TACC is developing a command-line interface called GridShell to improve users' access to grid-enabled software.

As part of TACC's active participation in using and developing grid technologies, UT Austin joined the NSF-funded NMI Integration Testbed Program. Managed by the Southeastern Universities Research Association (SURA) on behalf of the NMI-Enterprise and Desktop Integration Technologies (EDIT) Consortium, the testbed consists of eight universities that participate in a closely coordinated effort to deploy and evaluate NMI technologies[3]. Prior to participating in the NMI Integration Testbed, TACC had worked extensively with NMI software components. In fact TACC was one of the early users of NMI software releases, when each component of the software had to be installed separately from all others. Through its participation in the NMI Integration Testbed program, TACC was able to test mature middleware components including the convenient NMI installation package, which enables installation and configuration of a number of NMI components simultaneously. TACC has also enhanced its body of experience with NMI components, such as the GRIDS Center's Globus Toolkit[4] and Network Weather Service[5] (NWS). TACC further benefits from sharing experiences and ideas with colleagues from other institutions working within the NMI Integration Testbed.

Building the UT Grid

TACC is leading the UT Grid project at The University of Texas at Austin. The UT Grid project is an enterprise-wide, collaborative endeavor with many organizations across campus, including Information Technology Services (ITS), which is responsible for supporting the computing needs of the academic departments; the Institute for Computational Engineering and Sciences (ICES), an interdisciplinary research center in computational sciences and engineering; and the Computer Science department. Planning for UT Grid began over a year ago, but the project officially began in March of 2004, when it received a research grant from IBM Corporation.

The grid will initially make resources located at TACC and other campus organizations available to campus users. The initial users of UT Grid will be campus researchers, but the long-term goal is to offer UT Grid services to all users on campus, including staff and students. UT Grid will be deployed

[2] Condor-G information: http://www.cs.wisc.edu/condor/
[3] http://www1.sura.org/3000/NMI-Testbed.html
[4] Globus Toolkit information: http://www-unix.globus.org/toolkit/
[5] Network Weather Service (NWS) information: http://nws.cs.ucsb.edu/

initially as a "hub-and-spoke" network, rather than a peer-to-peer grid, to permit trust relationships to be built among all participants. The hub will consist of TACC computational, visualization, and storage resources located at two geographically separate campuses nine miles apart. A second, parallel test grid is also being constructed to explore new concepts before implementing them on the production grid.

UT Grid will be deployed in phases. The first phase will permit full access to the compute services supported by TACC resources at the "hub", and it will include user interfaces and middleware to support basic grid functionality. These functions include monitoring grid computing resources and submission and management of high-performance computing jobs. Future phases will add data, storage, and visualization services while enhancing existing services with additional functionality. During these phases, NMI software will be leveraged to facilitate sharing and scheduling of compute resources across the UT campus. Over the three-year project period, the additional needs of distributed file and database access, remote and collaborative visualization and steering, and instrument integration will also be addressed.

The UT Grid solution contains several component types: client nodes, resource nodes, and nodes that provide specific grid services. Client nodes have command-line or graphical interfaces for user access from laptop or desktop. Resource nodes (compute and visualization clusters, and network-attached storage devices) provide basic compute and storage functionality to users. Grid services nodes host security, scheduling, and resource monitoring for the campus grid. Each of the UT Grid components has a software stack defined to implement the required grid functionality. Each type uses one of more of the most recently released GRIDS Center components, including the Globus Toolkit, Grid Packaging Tools (GPT)[6] v3.1, MyProxy[7] v5.8, NWS v2.8.1, OpenSSH[8] v2.9, GridPort[9], UberFTP[10], and Condor-G v6.6.0.

UT Grid resource nodes use the GRIDS Center's Globus Toolkit to implement several key grid functions. The Grid Security Infrastructure (GSI) component of the Globus Toolkit is used for authentication of all grid users, and provides mechanisms for system administrators to control whether a user is allowed to access a given resource. The Globus Resource Allocation Manager (GRAM) provides a standard interface for requesting and using UT Grid resources for the execution of jobs. Installing and configuring the Globus Toolkit used to be a very labor intensive and error-prone process. Using the Grid Packaging Tools (GPT) has greatly simplified this process to just a few steps with a significant reduction in the installation and test time required.

[6] Grid Packaging Tools (GPT) information: http://www.gridpackagingtools.org
[7] MyProxy information: http://grid.ncsa.uiuc.edu/myproxy/
[8] GSI-OpenSSH information: http://grid.ncsa.uiuc.edu/ssh/
[9] GridPort information: http://gridport.net/index.cgi
[10] UberFTP information: http://dims.ncsa.uiuc.edu/set/uberftp/

Services nodes provide functionality for monitoring resources and scheduling jobs across resources on the UT Grid. The resource browser component of the GridPort toolkit is used to monitor the current load and availability of UT Grid resources. Network Weather Services (NWS) provides current network state information and estimates of the achievable bandwidths between pairs of grid resource nodes. Job scheduling services on UT Grid use components like Condor-G to forward jobs to grid resource nodes. The UT Grid user portal uses MyProxy to manage users' private key and certificate files, and for users to obtain proxy certificates when they sign on to the grid.

The GRIDS Center package contains several other tools that are very useful to users of UT Grid. Tools such as OpenSSH and UberFTP are used to provide support for remote logins and file transfers using the Globus Toolkit security infrastructure.

Although the NMI Grids components do not provide a complete grid solution today, there are several components, as mentioned above, that can be used in conjunction with other tools to create a grid that can support all the necessary functions.

Applications on the UT Grid

Grid resources can be grouped into several categories. Computational resources enable users to run detailed calculations and analyses of data; visualization resources allow users to create and manipulate graphical data; and storage resources of various kinds allow rapid access to large amounts of data. These resources are, of course, found in non-grid computing environments, but the grid environment enables them to be connected and used together in a way that greatly increases their ready availability. This increased resource availability enables research that would be almost impossible in a non-grid environment. Following is a description of how just two of the many UT Austin user applications will benefit from using various features of the UT Grid.

Computational Fluid Dynamics

Dr. Graham Carey of the Institute for Computational Sciences and Engineering works in the area of computational fluid dynamics (CFD), a research area well-suited to benefit from grid technology. CFD researchers use numerical techniques to obtain quantitative solutions to fluid flow problems, which become extremely complicated when flows become turbulent or the fluids are compressible in any degree (unlike water). Although researchers use several different CFD applications, the steps are often very similar, and optimizing this workflow could potentially benefit a range of CFD applications in addition to MGF, the application Carey and his group developed.

In ordinary usage, MGF is run on a computing cluster that can accommodate its computational demands. To visualize the data, either statically or interactively, it must be loaded into a rendering or other graphics application. Without UT Grid, the application user has to go through a more labor-intensive sequence of manual steps:

1. The user must log in on a high-performance computing cluster to run MGF. Since Carey's group has access to multiple clusters, the user must look at the queues on each cluster or guess which cluster appears to be the most lightly loaded and thus most likely to complete the job the quickest.
2. Once a cluster is selected, the user sets up the simulation and runs the job.
3. The user must monitor the job and wait for it to complete.
4. When the simulation completes, the output data must be transferred (perhaps using a file transfer utility such as *ftp*) to get the data to a visualization machine or cluster for rendering.
5. Finally, the user must log in to the visualization machine to run the rendering software that analyzes the results of the CFD simulation.

For certain types of simulations, these steps must be repeated a number of times, with changes in input parameters, until the desired results are obtained.

Clearly, Carey's group could get more research done if they did not have to perform the manual steps involved in this workflow. UT Grid can simplify their job by automating this workflow. Carey and his students need only to log in to the grid using a Grid User Portal (based on TACC's GridPort), or its command line equivalent, the Grid User Node (based on TACC's GridShell (3,4)) and submit their job to the "grid." The job is then analyzed by a metascheduler such as the Community Scheduling Framework (CSF) (5), which automatically selects the best available (emptiest/fastest) HPC cluster and submits the job. Grid computing workflow tools (such as TACC's GridPort or GridFlow) automatically move the results of the CFD simulation to a specified visualization system and launch the rendering application. In addition, when their research requires them to view the rendered data interactively, the Carey group can use UT Grid's advanced reservation capabilities to reserve time on the interactive visualization system located in TACC's visualization laboratory--or any other such system.

Analysis of High-Resolution X-ray Computed Tomography

Across the UT Austin campus from Dr. Carey is the UT Computed Tomography Lab where a huge database of biological specimen data, called DigiMorph, is being readied for sharing across the campus and beyond. Dr. Julian Humphries, a biological systematist, is the principal investigator for DigiMorph. This application will be among the first to make use of the massive data storage and sharing capabilities of UT Grid.

DigiMorph is a set of biological specimen records collected using high-resolution X-ray computed tomography (HRXCT). Researchers at UT Austin use the HRXCT scanner to make extremely detailed, 3-D images of objects ranging from 500 microns to 3 mm in diameter. The scanned images of a single object (e.g., an animal body, a fossil, plankton) are, on average, about 2.5 gigabytes (GB) in size. When the database was made available for use on the TeraGrid in October 2004, it contained about 500 specimens and totaled nearly 1 terabyte (TB) in size. With estimates of an additional 500 specimens to be added to the dataset each year, plus expected improvements in the CT scanner resolution, the UTCT data collection will more than double in size in the next few years.

Without grid computing, campus-wide and nationwide sharing of valuable biological information of such a massive size would not be possible. The UT Grid solution uses NMI data management software components such as the Storage Resource Broker[11] (SRB) to allow efficient access to the data collection. SRB carries metadata to allow access to parts of the data collection, and efficient data transfer as part of the data services.

Beyond the Nuts and Bolts

The building of the UT Grid is a complex undertaking. The technical intricacies are many, and require the hard work and commitment of many groups of people at UT Austin. By far one of the most important qualities TACC brings to the UT Grid project is their commitment to building a grid infrastructure with the spectrum of end-users in mind. TACC uses several approaches in ensuring that their advanced computing and grid services suit their users' needs. These approaches include creating new grid component software where none exists already and recommending modifications to "off the shelf" user software applications. While these activities are technical in nature, they simply must be based in the software developer having a firm understanding of their end-users' needs and, as we see at TACC, an organizational culture that encourages an end-user focus.

TACC's significant experience in the development of grid software is a prime example of their overall goal to create the most user-friendly grid computing experience possible. GridPort, GridFlow, and GridShell are three examples of software designed by TACC personnel that helps the user get the most from their grid computing experience. GridPort is a portal toolkit to aid in the development of science portals and applications using underlying distributed and grid computing technology. Portals built using GridPort provide a single point for researchers to login to the grid, monitor grid resources, and create and manage their jobs. GridPort designed portals thus help put all of the tools a researcher needs at their fingertips, saving them from having to search around on the grid for what they need. GridPort has been used in several national Grid projects since 1999, and is currently a component of the NMI release.

[11] Storage Resource Broker (SRB) information: http://www.sdsc.edu/srb/

Edward Walker of TACC is designing GridShell for use on the UT grid. TACC staff has found that many of their users prefer to use a command-line interface, rather than a graphical window interface, to run their scientific application software. GridShell software allows the user to enter command-line directions for their software, while in the background, GridShell translates these directions into appropriate sequences of distributed/grid service requests that can run the necessary grid resources required by the researcher's application.

GridFlow is a third example of TACC's own grid-customized software. GridFlow is designed to transfer intermediate results from a computing application process to a remote cluster and launching a visualization application at the remote node. The benefits of this approach for the researcher are that the workflow is automated, and the researcher is able to view intermediate data before completion of the entire simulation.

TACC's approach to ensuring the UT Grid and their implementation of NMI components is user-friendly includes processes that foster an open and accessible communication channel for their users. Direct user involvement in the planning of the UT Grid is encouraged by TACC's engagement strategy, including:

- focused user groups for specific research fields
- inviting users to TACC meetings
- user group mailing lists to keep users apprised of grid development progress
- informal conversations and relationship building with users
- formal training on using NMI components such as GridPort

TACC employs these approaches in order to better define user needs. The TACC staff is skilled in extrapolating individual researchers' needs and approaches to commonly useful scenarios. Limited time and resources mean TACC staff cannot speak in-depth with every user. Therefore, they have to be skilled at understanding how seemingly very different applications used in disparate research fields are actually very similar in the ways they can benefit from grid technologies. TACC staff also search for new ways to reach out to their user groups. For example, in addition to the methods used above, TACC plans to conduct a more formal user survey as a new way to define user needs.

A final insight from TACC's success in deploying grid technologies for their users is TACC's understanding that the "human factor" must be taken into account when planning a grid roll out. The hub and spoke design TACC has used for the UT Grid allows them to leverage the trust relationships they already have with other UT Austin organizations (e.g., ICES), researchers, and educators to integrate their resources into the grid initially as "spokes."

When they first join UT Grid, spoke resources are connected only to the TACC hub, and not shared between organizations. During this period, usage data will be collected for all resources on the grid, which will highlight the uneven usage patterns across resources on the grid. After participants at spoke sites gain confidence in UT Grid software, the usage data can be used to show them the benefits of sharing their resources with other organizations, so that they can trade idle resource cycles at lightly loaded periods to gain additional resources at peak usage times.

Over time, resources can be "connected" to the UT Grid, for willing participants, as peers. Rather than connecting to the TACC hub, these various spoke communities (such as physical sciences, geosciences, life sciences, and arts and humanities) will migrate to a direct connection to the UT Grid and to each other. This migration will be undertaken during a continuous process of user requirements analysis. The order of inclusion of new resources and services will be matched to the ongoing collection of requirements from each new user community.

Conclusion

The building of UT Grid is a powerful example of the benefits that grid technologies like NMI components have when they are implemented by an experienced and user-focused organization. However, organizations don't need to have the breadth and depth of experience that TACC has in order to undertake grid projects. The maturity of grid technologies like those in the NMI component suite and the willingness of organizations like TACC to share their knowledge and experiences with others make partial or full grid deployments attainable for less experienced organizations. The NMI Testbed Integration project has provided a forum for TACC staff to work with a group of colleagues from universities with mixed levels of experience in grid computing. TACC has been able to share with these universities, while at the same time gaining additional exposure to both the challenges and the potential that grid computing has in a variety of institutional settings.

More Information

For more information about TACC and UT Grid Design and Implementation, contact Tina Romanella de Marquez at tina@tacc.utexas.edu.

References

(1) *"The UT Grid Project"* http://utgrid.utexas.edu/
(2) *"The GridPort Toolkit"* http://www.gridport.net/
(3) *"GridShell: A Login Shell for Orchestrating and Coordinating Applications in a Grid Enabled Environment",* E. Walker, T. Minyard, and J. Boisseau, Proceedings of the International Conference on Computing, Communications and Control Technologies, Austin, Texas, August 2004.
(4) *"Orchestrating and Coordinating Scientific/Engineering Workflows Using GridShell",* E. Walker and T. Minyard, Proceedings of the 13th IEEE International Symposium on High Performance Distributed Computing, HPDC-13, Honolulu, Hawaii, June 2004, pp. 270-271.
(5) "Community Scheduling Framework" http://sourceforge.net/projects/gcsf

[i] TeraGrid is a multi-year effort to build and deploy the world's largest, fastest, distributed infrastructure for open scientific research. When completed, the TeraGrid will include 20 teraflops of computing power distributed at nine sites, facilities capable of managing and storing nearly 1 petabyte of data, high-resolution visualization environments, and toolkits for grid computing. These components will be tightly integrated and connected through a network that will operate at 40 gigabits per second—the fastest research network on the planet. http://www.teragrid.org