TEXAS TECH UNIVERSITY
Information Technology Division™

High Performance Computing Center

Status of the Texas Internet Grid for Research and Education Project (TIGRE)

Alan Sill

TIGRE Senior Scientist, TTU

Open Science Grid

OSG Consortium All-Hands Meeting

5-8 March 2007 | University of California, San Diego

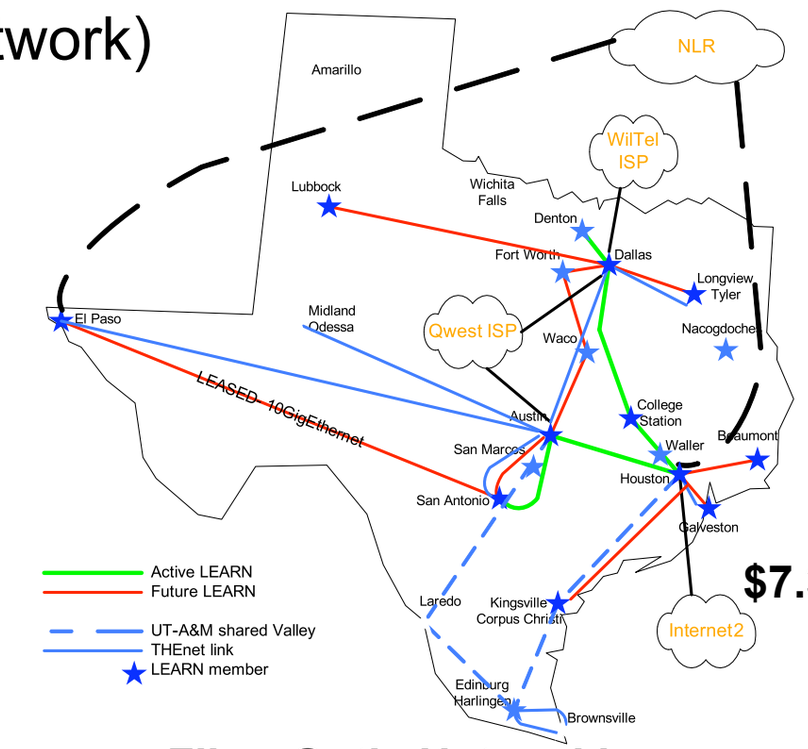
Background and History

- SuperComputing 1998: representatives from five Texas institutions met and agreed to cooperate and exchange notes on a variety of topics of interest in computing. ***High Performance Computing Across Texas (HiPCAT)*** was born as a result.
- TIGRE was proposed as one of HiPCAT's first projects. (Rice, Texas Tech. U. of Houston, Texas A&M, and UT/TACC.)
- Combined by Legislature into one funding bill with LEARN: (Lone Star Education and Research Network)

TIGRE

$2.5M

Grid Software Development



$7.3M

Fiber Optic Networking

Project Organization

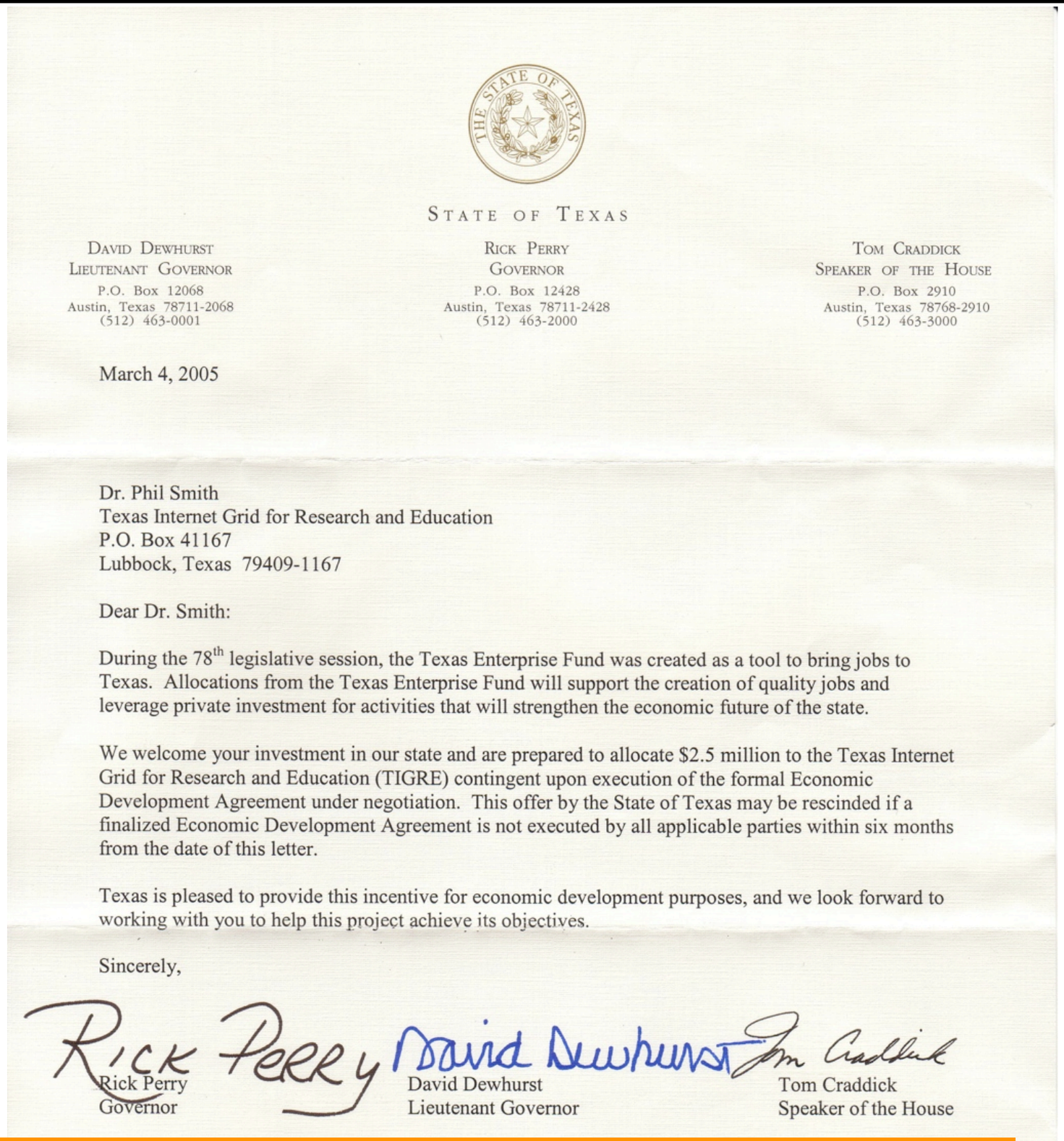
- Targeted application areas were selected by the **steering committee**, which also has responsibility for the overall direction of the project. These initial areas are:
  - Biosciences and Medicine
  - Air Quality Modeling
  - Energy Exploration
- Technical implementation is carried out by the **development team**, consisting of two people at each of the five primary TIGRE institutions. Technical work organized by the development team into **activities** to meet project **milestones**, with work targeted to meet these milestones in collaboration across institutions.

Goals of the TIGRE Project

- Provide a grid infrastructure that enables integration of computing systems, storage systems, databases, visualization labs and displays, even instruments and sensors across Texas.
- Facilitate new academic - government - private industry research partnerships by dramatically enhancing both computational capabilities and research infrastructure.
- Address research areas of interest to the State of Texas in which manifold increase of computing power, data access, and collaboration are necessary.
- Demonstrate new, preferred, enhanced or increased computing and storage handling capabilities offered by a statewide grid infrastructure.

Funding: Texas Enterprise Fund

- Created as a tool to "create jobs in Texas."
- Accompanied by a formal Economic Development Agreement with State.
- TEF intends to "create quality jobs and leverage private investment" to strengthen economic future of the state.
- This mandate differs from that of many grids.



STATE OF TEXAS

DAVID DEWHURST
LIEUTENANT GOVERNOR
P.O. Box 12068
Austin, Texas 78711-2068
(512) 463-0001

RICK PERRY
GOVERNOR
P.O. Box 12428
Austin, Texas 78711-2428
(512) 463-2000

TOM CRADDICK
SPEAKER OF THE HOUSE
P.O. Box 2910
Austin, Texas 78768-2910
(512) 463-3000

March 4, 2005

Dr. Phil Smith
Texas Internet Grid for Research and Education
P.O. Box 41167
Lubbock, Texas 79409-1167

Dear Dr. Smith:

During the 78th legislative session, the Texas Enterprise Fund was created as a tool to bring jobs to Texas. Allocations from the Texas Enterprise Fund will support the creation of quality jobs and leverage private investment for activities that will strengthen the economic future of the state.

We welcome your investment in our state and are prepared to allocate $2.5 million to the Texas Internet Grid for Research and Education (TIGRE) contingent upon execution of the formal Economic Development Agreement under negotiation. This offer by the State of Texas may be rescinded if a finalized Economic Development Agreement is not executed by all applicable parties within six months from the date of this letter.

Texas is pleased to provide this incentive for economic development purposes, and we look forward to working with you to help this project achieve its objectives.

Sincerely,

Rick Perry
Governor

David Dewhurst
Lieutenant Governor

Tom Craddick
Speaker of the House

Project Progress to Date

Goal is to achieve "quick build" toward working status:

YEAR 1

Q1:
Project plan ✔
Web site ✔
Certificate Authority ✔
Minimum testbed requirements ✔
Select 3 driving applications ✔

Q2:
Alpha quality user portal ✔

Q3:
Define server software stack ✔
Distribution Mechanism ✔
Simple demo of 1 TIGRE app ✔

Q4:
Alpha client software stack and installation method distributed ✔

YEAR 2

Q1:
Alpha customer management services system deployed & demonstrated; applications in three application areas *(in progress)*

Q2:
Project-wide global grid scheduler deployed

Q3:
Stable software status (only bug fixes after this)
Required services for TIGRE specified

Q4:
Complete hardening of software
Complete documentation
Finalized procedures and policies to join TIGRE
Demonstrate TIGRE at SC

... So Far Making Excellent Progress!

Details of TIGRE Software

- Both server and client software stacks available.
- Based on the VDT, working in close cooperation with OSG.
- Uses a simplified set including GSI-OpenSSH, omitting much monitoring in favor of GPIR for lightweight status reporting.
- Installed on systems at all five primary TIGRE institutions and also several other locations throughout the state.
- Goal is “one page” installation instructions that can be implemented quickly by newcomers.
- Authentication via X.509 (IGTF + TACC CA recently accredited); authorization local mostly via grid-mapfiles. (TTU uses GUMS/PRIMA.)

Server and Client Software Stack Contents

- Contents of the client software stack:
  - Globus 4.x pre-web services and web services clients
  - GSI OpenSSH client
  - UberFTP
  - MyProxy client
  - Condor-G
- Contents of the cluster/compute server stack:
  - Globus 4.x GRAM4 (web services) server
  - GSI-OpenSSH server (note: as of this version works with PRIMA)
  - GPIR monitoring (added manually after the VDT pieces)
  - The client software stack
- Storage is not yet resolved in TIGRE, but initial investigations using SRM/DRM, SRB/iRODS, etc. are in progress.

TIGRE User Portal http://tigreportal.hipcat.net

Grid Information Browser

Refresh

Parallel Computing Resources

Name	Institution	Department	System	CPUs	Peak GFlops	Memory GBytes	Disk GBytes	Status	Load	Jobs
Alamo	University of Texas Health Science Center at San Antonio	Department of Biochemistry	Rocks i386 Linux Cluster	19	0	8.7	1126	↑		0R-0Q-0O
Cosmos	Texas A&M University	Texas A&M Supercomputing Facility	SGI Altix	128	666	256	4096	↑		40R-38Q-0O
Eldorado	University of Houston	Texas Learning and Computation Center	Eldorado Itanium2 Cluster	124	472	256	2232	↑		0R-0Q-0O
Jacinto	University of Texas Health Science Center at San Antonio	Department of Biochemistry	Microway Linux Cluster	44	66	85.1	8433.3	↑		0R-0Q-1O
Laredo	University of Texas Health Science Center at San Antonio	Department of Biochemistry	Dual Athlon Cluster	32	0	31.6	9509	↑		0R-0Q-1O
Lonestar	The University of Texas at Austin	Texas Advanced Computing Center	Dell PowerEdge Linux Cluster	5200	55000	10400	94900	↑		70R-0Q-168O
Minigar	Texas Tech University	High Performance Computing Center	Dell Linux Cluster	32	230	64	70	↑		0R-0Q-0O
RTC	Rice University	Computer and Information Technology Institute	HP Itanium II Linux Cluster	290	1044	596	7000	↑		31R-0Q-0O
Weland	Texas Tech University	High Performance Computing Center	AMD Athlon MP 2000+	64	69	64	780	↑		0R-0Q-0O
			Total:	**5933**	**57547**	**11761.4**	**128146.3**			

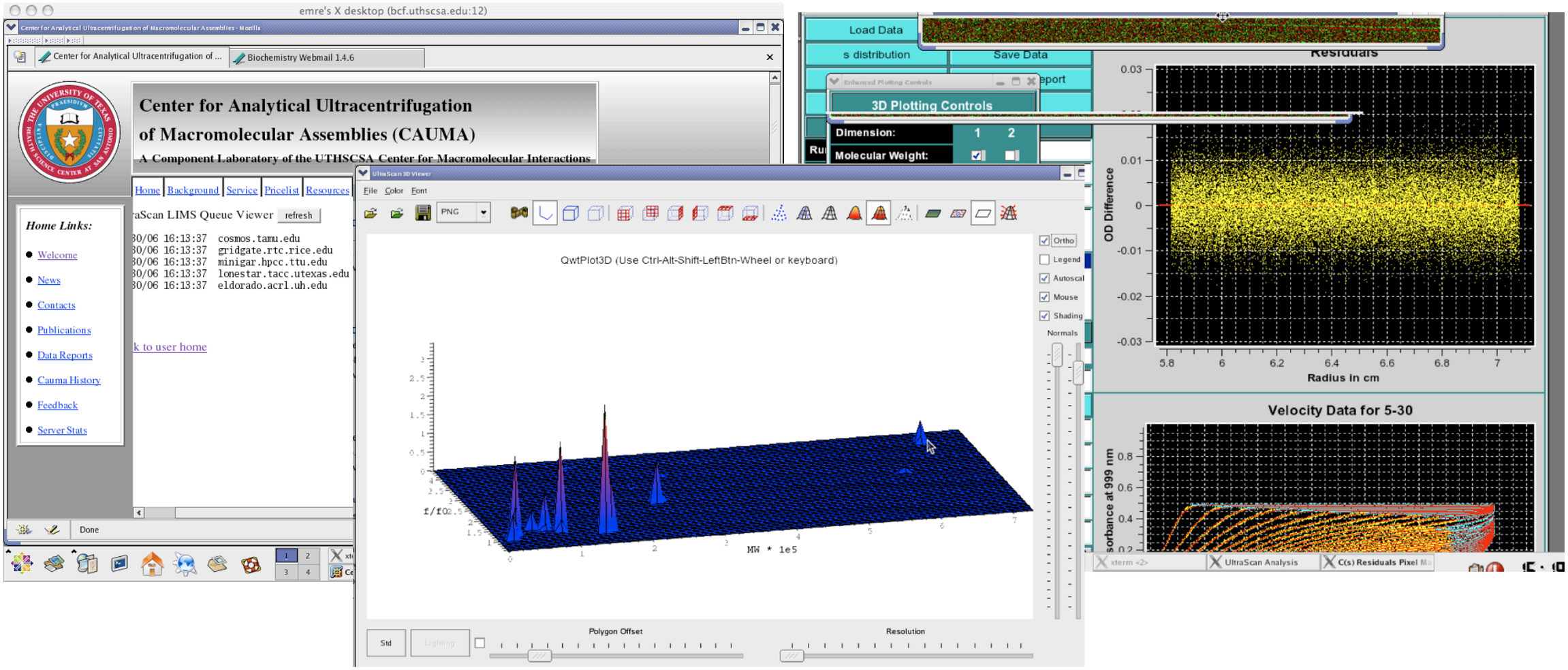
High Throughput Computing Resources

Name	Institution	Department	System	Active PCs	Active CPUs	Memory GBytes	Disk GBytes	Resource Details	Jobs
glb-test	Texas Tech University	High Performance Computing Center	Condor	**0** / 0	**68** / 68	14	1301		
Rodeo	The University of Texas at Austin	Texas Advanced Computing Center	Condor	**60** / 60	**60** / 60	49	1494		
			Total:	**60 / 60**	**128 / 128**	**63**	**2795**		

Jobs Key: **#R** - Number of Jobs Running, **#Q** - Number of Jobs Queued, **#O** - Number of Jobs in an Other State

Application Progress: Biosciences and Medicine

Initial demo was “UltraScan” : an analysis tool for reduction of data from ultracentrifuge biomolecular optical spectra.



Researchers (Borries Demeler, UTHSC-San Antonio w/postdoc Emre Brookes): “We can do science we never did before!”

Other Biosciences and Medicine Progress

- Beginning exploration of cancer radiotherapy modeling involving MD Anderson Cancer Center Houston, Joe Arrington Cancer Center and TTUHSC Southwest Cancer Center (Lubbock).
- Computationally intensive, good fit to grid computing and secure data transfer methods.
- Good fit to educational medical physics modeling needs.
- One of our most exciting near-term applications.

Identity management, security, privacy, license and export control requirements!

Application Progress: Air Quality Modeling

- One of the most computationally intensive application areas in Texas.
- Researchers identified at Texas A&M, U. Houston and TTU; application areas diverse and hard to separate from other general atmospheric modeling topics.
- Relevant code (WRF, MM5, etc.) exists at most institutions.
- Large amount of data generated; archival and redundancy, local access to data are among the relevant issues.
- Decide to start with data redundancy aspects while investigating more general code and framework needs.
- Goal: 31 day pool cache of modeling output results (~2 TB) to be shared among participating institutions.

Application Progress: Energy Exploration

- Also an existing area of computationally intensive use within the state.
- Complications stem from industry sensitivities to shared access to resources and need to protect proprietary data.
- Research topics still nonetheless exist in abundance.
- Working with different researchers at all five TIGRE primary institutions.
- Beginning work with Kalman-filter based modeling methods for oil reservoir seismic data analysis.
- Some license management issues, minimized by degree to which researchers already own or have access to licenses.

Storage Topics

- TIGRE needs to work with a wide variety of storage models and data file sizes, compositions and types.
- Project formulation did not address this topic specifically.
- Atmospheric modeling application area has driven us to consider this topic more quickly than we expected.
- So far we have been working to understand options for SRM-interfaced storage and tools for grid-authenticated parallel transfer, access and storage.
- Plain gridftp, srmcp, srmcopy, RFT and GSI interfaces to SRB and/or iRODS being studied and tested.
- GridImage from NCI CaBIG (CaGrid) project also of interest for DICOM image transfers in radiotherapy applications.

Authentication and Authorization

- Identity management and local authorization for virtual organizations within institutional boundaries is a ***very hot*** topic among institutional CIOs and the security organizations that serve them.
- Within Texas, we have one large TIGRE participant, the UT system, that has adopted Shibboleth for their own internal IdM federated across campuses, others that are looking at it and related products, and some that are not showing any signs of adopting it, as far as we can tell.
- Any solution we adopt has to work with both Shibbolized and non-Shibbolized campuses. So far we use X.509 certificates in combination with local gridmap or GUMS/PRIMA authz; TACC is working on an IGTF member-integrated ("MICS") credential service profile. Other related work is in progress.

Lessons Learned

- It is important to note that the applications being pursued in this project are the result of ***selection of areas by the steering committee*** and not (although they are compatible) from the results of natural selection processes that tend to create collaboration-oriented VOs.
- People tend to need a reason to collaborate. The availability of resources on which to run is more of a problem (in the chicken-or-egg sense) when trying to create sub-VOs and collaborations from scratch.
- In each case we have been able to make progress by ***putting people into the field*** to learn needs of researchers and to try to pull people and resources together to pursue common needs.

Outreach

- Outreach consists of pursuing needs ***beyond your comfort zone*** (by definition, think about the meaning of the word). In a grid context, this means doing things with the tools that we have not done before.
- Explaining has to be ***balanced with learning*** on the part of both the grid development team and the researchers.
- When the tools do not do what we want, we ***have to make them*** do what we (and the researchers) want them to.
- In many cases, this involves outreach to ***university decision makers*** (IT CIOs, Presidents, Provosts) as well as to the reseachers themselves to explain the need for new methods of working together as well as for networks, computers, etc.

What Does OSG Get Out of This?

- In a good partnership, all parties should bring something to the table and should gain some value from the relationship.
- TIGRE partners have
  - Participated in the Integration Test Bed (ITB).
  - Contributed several bug fixes both to Globus and to several other VDT components.
  - Contributed bug fixes and debugging of OSG-specific installation scripts and components.
  - Supported OSG VO testing both directly and on the ITB.
  - Explored several new applications that may lead to new VO involvement both with TIGRE and the OSG (and other large-scale grids).
- In general, this has been a very healthy, strong relationship!

Future Efforts

- Expect to work more closely with regional and other nearby state-wide grids (SURA, LONI, Rio Grande effort, etc.) in the near-term future.
- Project efforts have already led to one modest-scale (order $100K) proposal for commercialization with industry.
- Expect to complete demonstrations in all three targeted application areas soon, with more work coming.
- Need to consider "close-out" of this project to a long-term stable future. (Discussions with the Steering Committee in progress, many ideas being discussed.)
- Value our relationship with other national-scale, international, regional and local projects.

Conclusions

- TIGRE set out to involve support staff and researchers at each of 5 major Texas universities in ***production-quality grid software***.
- We have succeeded in this effort. Demos in targeted application areas are complete or in progress.
- We have a ***simple software stack*** that can serve as an entry point for researchers and institutions into other large-scale national, international, regional and local projects.
- TIGRE has primarily targeted ***outreach***, defined for this purpose as bringing grid software to areas that need it and/or to researchers who are not yet familiar with its use.
- The mandate for this type of work is ***strong*** and extends, in my opinion, into the long term future.

TIGRE

Texas Internet Grid for Research and Education

RICE

A HiPCAT Project

In memoriam

Ken Kennedy, founder of Rice University's nationally ranked computer science program and one of the world's foremost experts on high-performance computing, recently passed away at the age of 61. "Rice has lost one of its great intellectual leaders and a great human being," Rice President David Leebron said. "Ken Kennedy early on realized the power of computers to address real problems that confront people and the earth. His most recent contributions included using bioanalysis to help work on health issues like cancer."

He will be missed.