Southern California Earthquake Center (SCEC)
TeraGrid Project Description for 2005-2006 TeraGrid Allocations
TG-MCA03S012 (Olsen) and TG-BCS050002S (Okaya)
For TeraGrid Advanced Support

Version Dated: 27 May 2005

Philip Maechling (maechlin@usc.edu)
SCEC IT Architect
SCEC/CME Project Manager

Marcio Faerman (mfaerman@sdsc.edu)
San Diego Supercomputer Center

Introduction:

Scientists from over 54 research institutions throughout the country participate in the Southern California Earthquake Center (SCEC). The mission of SCEC, headquartered at the University of Southern California, is to gather new information about earthquakes in Southern California, integrate this information into a comprehensive and predictive understanding of earthquake phenomena, and communicate this understanding to end-users and the general public in order to increase earthquake awareness, reduce economic losses, and save lives. Funding for SCEC activities is provided by the National Science Foundation (NSF) and the U.S. Geological Survey (USGS). SCEC2 is funded under the following Cooperative Agreements: “Southern California Earthquake Center” - NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. Thomas H. Jordan (SCEC/USC) is the Director of SCEC.

SCEC is the lead organization on a Large NSF Information Technology Research (ITR) Project called the SCEC Community Modeling Environment (SCEC/CME). Principal Investigator on the SCEC/CME is Thomas H. Jordan (SCEC/USC) and Co-PI’s are J. Bernard Minster (UCSD), Reagan Moore (SDSC), and Carl Kesselman (USC/ISI). The SCEC/CME Project performs basic Geoscience and Information Technology research with the goal of improving the practice of seismic hazard analysis (SHA). SCEC/CME NSF Grant is: “The SCEC Community Modeling Environment (SCEC/CME): An Information Infrastructure for System-Level Earthquake Research” - EAR-0122464.

Since the SCEC/CME project provides funding to all the geoscientists working on the TeraGrid Projects, as well as to the IT researchers that participate on these allocation-based projects, we typically credit the SCEC/CME Project as the funding source for the TeraGrid research efforts including last year’s TeraShake Project. This helps credit the IT organizations as well as SCEC. A reference to SCEC as the lead organization on the SCEC/CME is also appropriate. Both SCEC, and the SCEC/CME Project, perform computationally intensive research on seismic hazard analysis.

Previous SCEC TeraGrid Allocations:

In 2004, an interdisciplinary group from the SCEC/CME Project performed a series of large earthquake simulations at SDSC, called the “TeraShake” simulations. This work has been presented at several conferences including SCEC Annual Meeting, SC2004, and Fall American Geophysical Union (AGU) 2004. Articles describing this work have been printed in the SDSC Envision Magazine, Science News (May 21, 2005), and the Science Grid web site. Journal articles describing the geophysical research results are in preparation. These simulations were largely run on the DataStar computer at SDSC. Visualization processing was performed on other computer nodes at that center. This work was run under a TeraGrid Allocation (Olsen) of approximately 100,000 SUs.

The TeraShake Project ran large scale (1.8Billion mesh points) finite difference, anelastic wave model (AWM), earthquake wave propagation simulations. The large geographical dimensions and high resolution (200m grid spacing) of the TeraShake simulations made them among the largest ever run in the seismological community. The TeraShake group saved over 40TB of data, primarily volumetric data, which was among the largest data set ever produced by a single geophysical simulation. The results of the simulations have significant implications for the seismic hazard in southern California.

Along with outstanding help from SCEC/CME-funded researchers at SDSC including Reagan Moore, and Mario Faerman, the TeraShake Project benefited greatly by a Strategic Application Collaboration (SAC) with Amit Majumdar's Scientific Computing Group at SDSC including significant code optimization and runtime support from Yifeng Cui for the anelastic wave propagation software (AWM-Olsen) used in the simulations. The SCEC/CME Project and SCEC have also benefited from outstanding visualization support from Steve Cutchin and Amit Chourasia in the SDSC Visualization Services group.

Current SCEC TeraGrid Allocations for 2005-2006:

Building on the strong Geosciences and IT collaboration of the SCEC/CME Project, and on the success of the TeraShake simulations, SCEC/CME researchers proposed and were awarded two TeraGrid Allocations starting March 2005. These are allocations are called "Southern California Earthquake Center (SCEC) Earthquake Simulation Project" (Olsen - 1,020k SUs) and "Probabilistic Seismic Hazard Map Calculations Using 3d Ground Motion Simulations" (Okaya - 145k SUs).

Each of these projects is attempting a *first* in geosciences computation. In the following sections, we will describe details about these two simulations.

Project 1: Southern California Earthquake Center (SCEC) Earthquake Simulation Project (TeraShake 2):

The original set of TeraShake simulations has generated outstanding results, new scientific insights, and great interest from the SCEC and seismological community. The success of the TeraShake simulations relies on a set of characteristics. First, the simulations used the SCEC 3D velocity model for southern California. The simulations results would not be as realistic without this 3D velocity model. Such velocity models are not available for all regions of the country. Also, the code used (AWM-Olsen) has been through a series of validation exercises, performed by SCEC researchers. These validations exercise include simulations of historic earthquakes. Because the seismograms calculated by these codes for historic earthquakes match the observed seismograms for these events very well, scientists believe the seismograms for "scenario earthquakes" are quite realistic.

In addition, the TeraShake Project brought together a group of geoscience and IT researchers with all the right skills. The group includes experts in wave propagation code, data analysis, data management, high performance computing, earthquake source descriptions, velocity models, scientific visualization, and other areas

To follow on the success of the TeraShake simulations, the SCEC/CME scientists propose to that introduce additional physics based elements into their simulations. Specifically, the TeraShake 2 simulations will include a simulation of a dynamic fault rupture.

Until this time, most earthquake wave propagation simulations specify the earthquake as a kinematic source, which specifies a series of velocity that might be generated by an earthquake. However, dynamic fault ruptures simulate the effects of friction including both static and kinetic friction. Because dynamic ruptures will be consistent with fault friction behavior, dynamic ruptures are expected to be more realistic source descriptions for earthquake simulations. However, dynamic rupture simulation techniques are still under active development and validation, and for this reason, very few, if any, large earthquake simulations have been performed using dynamic rupture-based source descriptions.

One typical problem relating to using dynamic ruptures in earthquake simulations is the need to couple dynamic rupture code to anelastic wave propagation code.

The TeraShake working group proposes to extend their TeraShake work by incorporating dynamic ruptures (also termed spontaneous ruptures) into large wave propagations simulations. The SCEC/CME group contains researchers that have developed, and validated, dynamic rupture simulation codes. By combining dynamic rupture simulations with the TeraShake scale wave propagation simulations, these simulations will represent the most realistic, physic-based, earthquake simulations ever performed.

In order to perform these simulations this year, the group proposes to integrate the dynamic rupture slip information and the anelastic wave code without directly coupling the codes. The group plans to run dynamic rupture simulations and save the resulting kinematic description of the fault slip as output data. This kinematic slip description will then be used as the earthquake source description in a large anelastic wave propagation simulation.

Project 2: Probabilistic Seismic Hazard Map Calculations Using 3d Ground Motion Simulations (CyberSHAke):

Seismic Hazard Analysis is a method used by seismologist, engineering groups, and emergency management personnel to forecast the levels of shaking expect at a particular site due to earthquakes. These ground shaking forecast depend on a number of factors including size and location of earthquakes that will occur, the attenuation of the earthquake waves as they propagate to the site being studied, and the geological properties of the site being consider. SHA studies typically produce a hazard curve indicating the range of ground motions levels (e.g. Peak Ground Velocities) expected and the probabilities for these ground motions, for a site over some period of time. Given a set of these hazard curves, scientists can create a hazard map. Hazard maps are used as inputs to building codes and hazard estimate studies.

Seismic Hazard Analysis is a high profile field within seismology because it has high social relevance. SHA is the interface between the theoretical science of seismology and engineering and public earthquake safety groups. Because improvements in SHA have great potential societal implications such as public safety, building codes and insurance rates there is great interest within SCEC and other research groups on developing new techniques for improving this science.

The computational challenge for SHA is that a Probabilistic Seismic Hazard curve is based on ground motions for many, possibly tens of thousands of earthquakes that may occur. Current SHA hazard curves use an empirically derived attenuation relationship to estimate the ground shaking from an earthquake for sites at distances from the earthquake. Scientists recognize that, due to advances in earthquake wave propagation simulations, significantly more accurate hazard curves could be produced if peak ground motions are derived from synthetics seismograms, rather than from attenuation relationships. Seismogram-based peak ground motion estimates can use 3D velocity models and are reasonably expected to better reflect the geological behavior of a site.

Waveform-based intensity measure relationships (IMRs) haven't been developed yet because of the large number of earthquakes that must be simulated for an earthquake rupture. For a full hazard curve calculations for Los Angeles, approximately 45,000 earthquake simulations may be needed to produce a hazard curve.

Our CyberShake Project plans to calculate this large number of ruptures through the use of a seismological technique called reciprocity. In this technique, large wave propagation simulations are run for the site being studied, and then a large number of ruptures can be simulated very inexpensively. In this technique, you can calculate the effects of many ruptures for a single site. However, if you wish to calculate the effects for a different (even nearby site), you must run another set of large wave propagation simulations.

The CyberShake simulation is a first implementation of a widely discussed, but never implemented, idea of creating a waveform-based IMR. Also, it is a demonstration of the capabilities of the SCEC/CME workflow system. The SCEC/CME project requirements include development of a workflow system. It is important to use the SCEC/CME workflow system on this project.

The results of the CyberSHAke Project are not as visually striking at the TeraShake simulations. The CyberSHAke output is either (a) a hazard curve, or (b) a hazard map. Since these results are basic, non-dynamic images, they lack the visual impact of an earthquake simulation. However, the social significant of the CyberShake results is very real, and there is great scientific interest already within SCEC for the CyberShake project because it is a first attempt at replacing the use of attenuation relationships in PSHA with full waveform modeling.

TeraShake-2 Work-Plan Material:

In the following sections, we outline the basic workplan for the TeraShake-2 effort. The TeraShake-2 effort is lead by Kim Olsen (SDSU) and J. Bernard Minster (UCSD) with significant support from San Diego Supercomputer Center (SDSC) including Reagan Moore and Marcio Faerman, as well as other SCEC/CME researchers including Steve Day (SDSU), Ralph Archuleta (UCSB), and Jacobo Bielak (CMU). There is additional planning material in the TeraShake 2 proposal (URL) and in our TeraShake-2 meeting notes (URL).

TeraShake-2 Success Targets:

1. Scaleable, validated Dynamic rupture simulation software. MPI-based code that can simulate ruptures on 200km+ fault segment.
2. File-based coupling of Dynamic rupture codes and AWM codes.
3. TeraShake scale AWM simulations using Dynamic fault model slip-based source descriptions.
4. Comparisons between TeraShake 1 results (Kinematic sources) and TeraShake 2 results (Dynamic Rupture sources).
5. Simulation results for multiple rupture variations.
6. Map-based animations of ruptures.
7. Spectral acceleration maps and animations for ruptures.
8. Volumetric Visualizations for multiple ruptures, including interactive visualization of some volumetric data.

9. SRB-based digital library containing collections of seismograms, well described with metadata, available to SCEC researchers.

TeraShake-2 Simulation Organization

1. 16 Rupture Dynamic Iterations
2. 14 Wave Propagation Runs, Surface Only
3. 2 Wave Propagation Runs, Surface and Volume
   Where: Rupture Dynamic Iteration consists of
   - 3 Low Resolution Rupture Dynamics Estimate Runs (Mesh w/ 0.2 Gpoints each)
   - High Resolution Rupture Dynamics Run (Mesh w/ 1.7 Gpoints)

TeraShake-2 Workflow Structure

Each Rupture Dynamic Iteration will produce input for a Wave Propagation scenario. The Wave Propagation Runs will produce a velocity vector wavefield dataset as output (surface only or surface+volume), which will be further post-processed, deriving several data products including magnitude, displacement, curl, divergence, seismograms, visualization renderings. The derived data products will be archived for long term preservation in the SCEC Digital Library supported by the SRB.

The workflow sequence described above will require TG support regarding the staging of each intermediate input/output dataset on GPFS. The data produced at each stage would either need to be held on GPFS until workflow completion or alternatively moved back and forth from the SCEC Digital Library archives (currently SDSC samfs) to the TG site processing the data.

TeraShake-2 Resource Demand Estimates (projection for SDSC TG site)

1. Rupture Dynamic Iterations
   a. 18,155 SUs
   b. Up to 240 CPUs on high resolution run
   c. Data Output – 1 TB
   d. Checkpoints – 4 TB

2. Wave Propagation Runs, Surface Only
   a. 16,617 SUs
   b. 240 CPUs
   c. Data Output – 1 TB
   d. Checkpoints – 4 TB

3. Wave Propagation Runs, Surface and Volume
   a. 28,326 SUs
   b. 240 CPUs
   c. Data Output – 41 TB
   d. Checkpoints – 4 TB

4. Visualizations for each Wave Propagation Run
   a. ~ 7000 SUs

   b. ? CPUs

5. Data Transpose for Seismogram derivation
   a. Current serial code benefits from large memory (used DataStar P690 nodes on TeraShake1).
   b. Low SU demand.
   c. Data Input – 1TB for each scenario
   d. Data Output – 1TB

6. Magnitude, Displacement derivations
   a. Low SU demand.
   b. Data Input – 1TB for each scenario.
   c. Data Output – 1TB for each scenario

7. Curl and Diverge derivations
   a. Current serial version will require large memory nodes (DataStar P690s).
   b. SU demand to be estimated.
   c. Large data input, 40TB for each surface+volume scenario.
   d. Large output, 53TB for each surface+volume scenario.

8. Spectral Acceleration
   a. Under Evaluation
   b. Data Output – Relatively Small

TeraShake-2 Target Deadlines

- SCEC Annual Meeting, Mid-September 2005. Present results (visualizations, seismograms, spectral maps and plots) from 2 surface scenarios and depending on results from surface only runs, surface+volume results as well.
- SC2005, November 2005. Demo can include:
  - Multi-site TG run, with distributed registration of results in multiple SRB servers at each site. Demonstrate Web scenario-interface access to multiple Teragrid, SRB sites.
  - Visualization renderings on distributed TG sites
  - Interactive visualizations of volumetric data using Grid-based visualization tools (GVU).
- AGU Fall Meeting. December 2005. Consolidated set of scenarios ready, indicating major issues revealed by the new spontaneous rupture methodology employed in TeraShake2.

TeraShake-2 Technologies

SCEC would greatly benefit from the TG Support Group regarding the evaluation of technologies targeting:

1. **Reliability –** Data Integrity must be preserved along the simulations workflow, post-processing, data access and archival. Integrity preservation becomes much more critical given the high probability of corruption of the large datasets involved, exceeding hundreds of Terabytes. We are currently

using parallelized MD5 strategies for that purpose. Input on possible alternatives is greatly welcome.

2. **Performance and Efficiency –** High performance strategies will be critical in producing results in a timely fashion along next year. Critical issues include remote data access and data access from tape storage systems. Efficient resource utilization is also important, regarding the consumption of the project allocation SUs and also in interest of resource sharing with other projects.
3. **Robustness –** Fault detection and recovery strategies will be needed to support the progress of the project.

We believe that the TG Support group could greatly help SCEC in evaluating the following technology alternatives which could be applied to the project.

TeraShake-2 Data Management

- Global GPFS (transparent GPFS mounts of TG remote sites)
- Storage Resource Broker distributed at each TG site, for collection based reference to GPFS or local file systems, in addition to SRB server support at data archival sites (such as with SDSC samfs, NCSA archives?)
- Remote MPI/IO – SRB integration with MPI/IO developed by Mario Lauria (lauria@cse.ohio-state.edu) in collaboration with the SRB group for post-processing derivations, reducing need to move data across the network. http://www.cse.ohio-state.edu/~lauria/pubs/ccgrid2005.pdf
- HDF5 integration with the SRB. SCEC plans to use HDF5 to describe its scientific datasets with a standardized architecture independent format. The SDSC SRB group and the NCSA HDF5 groups are developing together a “Remote-HDF5” technology, allowing users to access datasets through the HDF5 API, stored at remote SRB vaults.
- Remote-HDF5/SRB integration with Remote MPI/IO technology – This integration would enable an equivalent to a Remote-Parallel-HDF5 technology, allowing remote efficient parallel access to SCEC HDF5 datasets.
- Support for disk-based storage required to perform interactive visualization of volumetric data sets. Tools on the Project support a grid-based interactive visualization of large (>1TB) data sets. The data to be visualized must be on disk and Globus-grid accessible. This would be temporary disk-storage.

A major goal is to build a system which can efficiently read and write datasets from the SCEC Digital Library.

TeraShake-2 Workflow Management:

- The SCEC/CME Workflow System is based on Globus, CondorG, Chimera, Pegasus (Virtual Data System), if needed.
- Workflow Data Management (MCS,RLS, GridFTP)
- Long Term, Permanent Digital Library uses the SCEC Digital Library using the Storage Resource Broker (SRB) at SDSC.
- Application Driven (MPI/IO, explicit file access)

CyberSHAke Work-Plan Material:

In the following section, we outline the basic workplan for the CyberSHAke effort. This effort is lead by David Okaya (USC) and Philip Maechling (SCEC/USC) with significant support from San Diego Supercomputer Center (SDSC) including Reagan Moore and Marcio Faerman, and SCEC/CME funded researchers at USCs Information Sciences Institute (USC/ISI) including Carl Kesselman, Ewa Deelman, and Sridhar Gullapalli (ISI). There is additional planning material in the CyberSHAke work plan that has been developed to manage this effort. That document (http://epicenter.usc.edu/cmeportal/docs/CyberSHAke.pdf) is updated as the work progresses.

CyberSHAke Simulation Organization:

The CyberSHAke simulations can be understood as calculations of a hazard curve for a particular “Site.” A series of calculations are performed relating to one particular site, such as USC University Park Campus, or Los Angeles City Hall. Calculations continue until we are able to create a Hazard Curve for the site being studied.

Our CyberSHAke planning revolves around this concept of a “Site”. We have estimated what calculations are required to produce a hazard curve for a single Site.

Codes are re-usable between sites, but the calculated data sets are not. All portions of a Site calculation must be done specifically for a site. So, our estimates will focus on calculations for a single site. We then multiply these by “Site” estimates by the number of sites desired.

Our current Project Plan is to have 1 Site calculated by 22 June (SCEC/CME All Hands Meeting) and 10 sites calculated by September 10, 2005 (SCEC Annual Meeting). In order to create a hazard map, we want to calculate 100 sites or more by the end of 2005. However, we believe this will require substantially more SUs than we currently have, so supplemental allocations may be required.

CyberSHAke Success Criteria:

- Minimum result - Full ERF-based hazard curves using approximately 45,000 Ruptures with variations, for 10-30 sites.
- Project Goal – Hazard Map (50km x 50km with 5km grid spacing for Los Angeles Area) using 100 sites by the end of 2005.
- Workflow-based job submission accessing SUs on all available resources including multiple TeraGrid sites including SDSC, Argonne, NCSA, Caltech, and PSC and also other computing resources USC HPCC, ISI, SCEC and any other grid-based system available to our collaboration.
- Workflow-based data management including automatic file and metadata registration.
- Collection of SRB-based Strain Green Tensors in SCEC Digital library.
- Computer Science Journal articles describing workflow capabilities.
- Geophysical Journal articles describing Hazard Curve results.

Elements of a "Site" Calculation:

- 2 Wave Propagation Simulations with output volume data.
- 350,000 Rupture Processing Calculations with output seismogram files.
- Other Post processing steps (minimal computation requirements).

Each Wave Propagation Simulation consists of:

- High Resolution Strain Green Tensor (Mesh w/ 1.5 Gpoints) calculation using MPI-based finite difference code.
- Each simulation is estimated to require 10,320 SUs running on 176 processors, with 980MB RAM required per processor.
- Two simulations required per "Site"
- Data output is 4.75 TB of volumetric data per Site.

Rupture Processing consist of:

- Each site requires rupture calculations that produce seismograms.
- Serial FORTRAN code run in a high throughput CondorPool computational model on TeraGrid.
- Volumetric Strain Green Tensor Data (4.75TB) must be accessible during the rupture processing.
- Outputs are collected into "Ruptures" with approximately 45,000 seismogram files per site.
- Seismogram files managed and annotated with Metadata.
- Estimated SUs per Site for Rupture Processing: 8,266
- Estimate Disk Storage for Rupture Files: 3.2 GBytes

Total Estimate Service Units and Disk Storage Requirements

- Total SUs/Site = 10,320 + 10,320 + 8266 = 28906
- Total Disk/Site = 4.75TB

- Minimum Success Criteria Totals = 289K SUs and 47.5TB
- Hazard Map Goal Total = 2890K SUs and 475TB

We believe that our minimum project requirements of 10 sites are possible using our current allocation. We can augment our TeraGrid allocation with SUs at SCEC, ISI, and USC HPCC.

To create a Hazard Map, some optimization and trade-offs will be required. A Hazard Map results will clearly have the greatest impact on the seismological and emergency management community. The scientists may decide, for example, that the volumetric data can be processed, and then removed, in order to produce a map. Once we have more experience with these calculations, we will decide an appropriate strategy.

Workflow Structure

The SCEC/CME workflow system is required in this Project to work on the scale required. A paper describing the system has been submitted to ACM SIGMOD Record and is posted (http://epicenter.usc.edu/cmeportal/docs/SCEC_Workflow_ACM.pdf).

Key capabilities of our workflow system in this Project will include: (a) reasonably responsive, and high throughput, access to 80+ nodes on the TeraGrid, (b) and disk space for our large wave propagation simulation, (b) submission to a wide variety of sites so that we can obtain high throughput for our rupture process, and (c) flexible data movement, and data file management. Our workflow will use GridFTP, RLS, and MCS to achieve this. We also need temporary large disk-based storage for our Strain Green Tensor Files (4.75 TB) which our Rupture Processing will require.

Resource Demand Estimates (projection for SDSC TG site)

- Wave Propagation Runs Volume (Totals for 2 simulations)
    - 20,640 SUs
    - 176 CPUs each simulation
    - Data Output – 4.75 TB (Strain Green Tensor data)
    - Checkpoints – .5 TB

- Rupture Processing
    - 8266 SUs
    - Variable number of processors
    - Data Input – 4.75 TB Disk-based Strain Green Tensor data
    - Data Output – 3.2GB for each site

Target Deadlines

- SCEC/CME All Hands Meeting (21 June 2005). One site calculation should be completed.
- SCEC Annual Meeting (11 Sept 2005) Present results including calculated seismograms, validation results, and Hazard curves for 10 sites.
- SC2005, November. Demo can include:
    - Multi-site TG workflow run with distributed registration of results in multiple SRB servers at each site. Demonstrate SCEC/CME Workflow system including construction of workflows, monitoring tools, and TeraGrid Wide job submission.
- AGU Fall Meeting (December 5, 2005). Hazard Map using 100 Sites.

TeraGrid Technological Support

SCEC would greatly benefit from the TG Support Group regarding the evaluation of technologies targeting. Here are some of the challenges that we can use help with:

- Our large wave propagation simulations on the TeraGrid will take many nodes, and multi-days, with TB of output. How can we get these jobs run quickly and reliably? Access to special queues, or run-time reservations would be very helpful
- Data management of the large volumetric data. Long term repository will be the SDSC SRB. However, we may need help with temporary staging and storage during calculations.
- Data movement will be based on GridFTP. We need GridFTP visibility for our files. Reliable File Transfer on top of GridFTP would be helpful.
- GridFTP access to SRB would be helpful.

- The Chimera/Pegasus Workflow tools available everywhere on the TeraGrid. We utilize a workflow software stack based on Virtual Data System.
- Support for running our Rupture Processing codes on TeraGrid. This code requires access to disk-based volumetric data (Strain Green Tensors) while the rupture processing runs. The disk-based data must be locally accessible (or seem local). This is a high throughput computing challenge.
- Automated File management and metadata management is very important in this workflow due so we need RLS and MCS tools available at as many TeraGrid sites as we can.
- Optimization and Validation of codes on as many TeraGrid resources as possible. We'd like access as many TeraGrid resources as possible. We think our workflow system will help us make good scheduling decisions. So porting our code to all available platforms may be beneficial. Since our SUs estimates indicate that we will use up our SUs before we reach our goal of a Hazard Map, any optimization of our system for running on the TeraGrid would be helpful.
- Obtaining supplementary allocations when needed. Assuming we are obtaining good interesting promising results, once we've used up our current allocation, we would like to continue our calculations towards our goal of a Hazard Map.

Workflow Management Tools

- Globus, CondorG, Chimera, Pegasus
- Replica Location Service (RLS) at computation sites
- Metadata Catalog Service (MCS)
- GridFTP Access to TeraGrid Disk Storage
- Reliable File Transfer Tools for moving large files
- Storage Resource Broker Client for storage of Long Term Results