TeraGrid Visualization and Data Resources and Services

Kelly Gaither
TeraGrid Area Director for Data, Visualization and Scheduling
Texas Advanced Computing Center
The University of Texas at Austin
TeraGrid Visualization Strategy and Mission

• Combine existing resources and current technology…
  • Commodity clusters and commodity graphics
  • Terascale visualization clusters
  • Grid technology
  • Efforts, expertise, and tools from each of the TG sites

… to enable new and novel ways of visually interacting with and gaining insight into science through the analysis of simulations and data.

• TeraGrid deploys 2 visualization resources, but has 9 resource provider sites that do active visualization for the TeraGrid user community.

• TeraGrid also deploys visualization resources for batch, interactive, and collaborative visualization.
## TeraGrid Visualization Software Stack (VTSS)

<table>
<thead>
<tr>
<th>Package</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesa</td>
<td>Mesa is a 3-D graphics library with an API similar to that of OpenGL.</td>
</tr>
<tr>
<td>VTK</td>
<td>The Visualization ToolKit (VTK) is an open source software for 3D computer graphics, image processing, and visualization, compiled with Mesa or OpenGL.</td>
</tr>
<tr>
<td>Chromium</td>
<td>Chromium is a system for interactive rendering on clusters of graphics workstations.</td>
</tr>
<tr>
<td>Netpbm</td>
<td>Netpbm is a toolkit for manipulation of graphic images, including conversion of images between a variety of different formats.</td>
</tr>
<tr>
<td>ImageMagick</td>
<td>ImageMagick is a collection of tools and libraries to read, write, and manipulate images in a variety of image formats.</td>
</tr>
<tr>
<td>ParaView</td>
<td>ParaView is an extensible, open source, multi-platform application for visualizing large data sets. It supports distributed computation models for processing large data sets.</td>
</tr>
<tr>
<td>CMake</td>
<td>CMake is a cross-platform, open-source make system. It is used to compile VTK and ParaView and can also be used as a build environment for user applications.</td>
</tr>
</tbody>
</table>
TeraGrid Visualization Hardware Resources

• UC/ANL Visualization Resource
  – Cluster configuration, 96 visualization nodes, 62 compute nodes
  – Focused on a solution that provides a client-server visualization mechanism through the use of Paraview

• UT/TACC Visualization Resource
  – Large, shared memory machine with 512 GB of shared memory and 16 dual out commodity graphics cards
  – Uses VNC as a means for remote/collaborative visualization
  – Enables any third party visualization package that uses OpenGL
    • Paraview
    • EnSight
    • Amira
    • VMD
Sun Microsystems E25K Server
64 Dual Core UltraSparc IV Processors
512GB Shared Memory

32 Graphics Pipes
QuadroFX 3000G Graphics Cards

540 GB of Disk for Home Directories
8.8TB of Disk for Parallel-IO - Temporary Storage
3.2TB of Disk for Parallel-IO - Long-term Storage

High Performance Fiber Channel Interconnect
High Performance Gigabit Ethernet

Sun Microsystems V890 Login
8 Dual Core UltraSparc IV Processors
32GB of Shared Memory
Visualization Areas of Interest and Expertise by Resource Provider

Indiana University (IU)
- Distributed software rendering
- Data repository and visualization integration

National Center for Supercomputing Applications (NCSA)
- High definition resolution visualization
- Volume rendering
- Visualization web services

National Center for Atmospheric Research (NCAR)

Oak Ridge National Laboratory (ORNL)
- Visualization of experimental data collected from spallation neutron source

Pittsburgh Supercomputing Center (PSC)
- Simulation-time Visualization with hardware rendering directly tied in to the supercomputer

Purdue University (Purdue)
- Remote batch rendering
- Remote interactive visualization
- Hardware-assisted remote rendering

San Diego Supercomputer Center (SDSC)
- Visualization web services
- Volume rendering
- Remote rendering
- High definition resolution visualization

Texas Advanced Computing Center (TACC)
- Remote and collaborative visualization
- Large data visualization
- Feature detection

The University of Chicago / Argonne National Laboratory (UC/ANL)
- Visualization, analysis, and collaboration services
- Remote rendering
- Hardware-assisted volume rendering
TeraGrid Visualization Future

- Development and deployment of community access environment to the TeraGrid Visualization Gateway
- Integration of visualization gateway efforts into a single unified solution
- Development and deployment of hardware-based volume rendering service
- Integration of Access Grid client into TeraGrid Visualization Gateway for collaborative viewing and interaction
- Recruit visualization community to contribute to and use the TeraGrid Visualization Gateway.
TeraGrid Visualization Gateway

• Highlights:
  • Users can use TeraGrid user portal accounts or create community accounts
  • Community users have limited capabilities
  • Initial set of functionality is Paraview and remote visualization portlets

• Snapshots:
  • Initial visualization portlet screen where a user can launch a session
  • Upon launching a session, a user must refresh the job until it is running or can delete the session
  • Once the session is running a user can launch the applet and run a visualization
TeraGrid Visualization Gateway

Welcome to the TGviz Portal

Existing TeraGrid users can load their credentials from the MyProxy server to gain access to TeraGrid Visualization resources. Your credentials must be in both the grid-mapfile on the University of Chicago resource, and available from the MyProxy service.

For instructions on setting up credentials and loading them into the grid-mapfile, see the Setting Up Certificates and DN's section of the TeraGrid Getting Started Guide.

For instructions on loading your credentials into the MyProxy service, see the Proxy Certificates section of the TeraGrid Getting Started Guide.

ParaView Portlet

First time users of the TGviz Gateway should see the instructions for Starting the ParaView Client from the TGviz Gateway

ParaView 2.4.1 is the latest version of ParaView installed on the TGviz nodes. It is recommended that you use the same version for your local client, which can be found on the TGviz Software Download pages.

For additional information on ParaView, including user tutorials, see the TeraGrid pages.

No proxies currently loaded. This portlet will remain inactive until you get a credential.
If this portlet does not automatically update after you get a credential, click the reload button below.
Reload
If you have previously started ParaView during this session, it is no longer running.
TeraGrid Visualization Gateway: Remote and Collaborative Visualization Portlet

• Allows the user to run a remote visualization session on maverick.
• User launches a new session and must refresh portlet until the job is RUNNING
• Once the job is running the user can launch the applet
• User can delete the session when they are done
• Only available to TeraGrid users, not to community account users
To launch a remote visualization you need to have a Java-capable browser, an active account on maverick.tacc.utexas.edu, and set up a VNC password on maverick. For more information please view the Maverick user guide.

This portlet enables you to launch a remote visualization session on Maverick. To start click on the 'Launch Session' button to start a new session and refresh the session until the session is running. Once the job is running there will be a 'Launch Applet' link that will open a new window and take you to a remote visualization session. Once you launch the applet you will be prompted for your VNC password. Upon successful login you will have access to a remote session on maverick.

Launch Session

October 9, 2006
After Submitting a New Session: Launch Applet

Remote Visualization Portlet

Remote Visualization

To launch a remote visualization you need to have a Java-capable browser, an active account on maverick.tacc.utexas.edu, and set up a VNC password on maverick. For more information please view the Maverick user guide.

You can only have one remote visualization running at a time. Please refresh until your current session is running then you can launch the remote visualization.

Session Status: RUNNING

Launch Applet
Delete Session

Note: To launch a new session please delete the current session you have running.

October 9, 2006
The VNC Applet Login
After Logging in to the VNC Applet
Visualization Alert:

• The Visualization Working Group will be hosting a user workshop in Spring 2007 that will solicit data from users throughout the community.
• Stay tuned for more details.
TeraGrid Data Strategy and Mission

• The TeraGrid is focused on the management, movement and collection of data through the deployment and development of resources and tools to facilitate the storage and transfer of terascale data.
• The TeraGrid is working towards a high performance globally available file system on all TeraGrid resources such as GPFS-WAN, LUSTRE-WAN.
• The TeraGrid is also working towards a globally available file system on all TeraGrid resources that is more modest in performance: e.g. xufs, mounted file systems, Amazon S3 storage.
TeraGrid Data Management

• Users control where their data resides. Data can be stored in a user’s home directory, in a temporary location or in archival storage.
• Home directories are permanent, but have relatively small quotas. The data in the home directories is never purged.
• Temporary storage in scratch is shared and has more space than home directories, but depends on use by other users.
• Parallel file systems are fast and large, but are temporary.
• Archival storage offers the advantage of long-term storage of large amounts of data, is accessible at all sites, but offers slower access times.
• SRB is available as a special system for the organization of data.
TeraGrid Data Movement

• Moving data between TeraGrid locations is facilitated by a number of clients that are available at the RP sites, most of which are based on GridFTP.

• There are three GridFTP clients available on the TeraGrid for data transfer:
  – Tgcp: a command line user tool to provide high-performance file transfer, while simplifying efficient copying of files and directories between and within gridFTP enabled resources. Tgcp is a wrapper for globus-url-copy and RFT (Reliable Transfer Service).
  – globus-url-copy: a gridFTP client for transferring files from the command line.

• Transferring data between your local client machine is typically done using common Unix commands.
TeraGrid Data Collections

- There are currently 89 data collections at 5 resource provider sites.
- These collections are listed at http://www.teragrid.org/userinfo/guide_data_collections_table.html
TeraGrid Data Future

• Will be concentrating on a TeraGrid wide deployment of a wide area parallel file system.
• Will be focusing on the rapid transfer of large data stores between TeraGrid sites.
• Will be investigating a variety of ways for provide a user’s view of a centralized file storage mechanism.
• Will be hosting an internal data workshop in January 2007 that will explore these issues in great detail and will produce a white paper.