Science Gateways Overview

Data Analysis on Ranger

October 23-24

TeraGrid Resource Providers

Background

- Science Gateways: domain specific portals to resources

- Allow researchers to focus on their science and not the "plumbing" involved with using the resources

- They mask the complexity of distributed resources for example
  - Data Collections
  - Computing Resources (hardware and software)
  - Remote Instrumentation & Sensor Data

- The provide a focal point with relevant supporting information
  - References to relevant papers & presentations
  - Web based documentation & training
  - Education & outreach activities

Some Examples

- Official list
  - http://teragrid.org/programs/sci_gateways/

- Purdue NanoHub
  - http://www.nanohub.org/

- Network for Earthquake Engineering Simulation
  - https://central.nees.org/

- Some Cornell examples
  - Cornell Computational Biology Service Unit**
    - http://cbsu.tc.cornell.edu/
  - High Resolution Daily Temperature and Precipitation Data for the Northeast
    - http://compag.tc.cornell.edu/sciencegateway/
  - Legacy Pulsar Data
    - http://arecibo.tc.cornell.edu/legacypulsardata/Default.aspx

** Not a TeraGrid Science Gateway

Creating Your Own Science Gateway

1. Read TeraGrid Science Gateway Primer
   - http://www.teragridforum.org/mediawiki/index.php?title=TeraGrid_Science_Gateways_Primer

2. Request a TeraGrid allocation for your project
   - http://www.teragrid.org/userinfo/access/allocations.php
   - http://www.teragrid.org/userinfo/access/proposal_gateway.php

3. Once you have an allocation fill out Community Account & Gateway Form
   - http://www.teragrid.org/programs/sci_gateways/apply/

4. Contact Science Gateways Coordinator for additional help
   - gateway-info@teragrid.org

TeraGrid Resources Available to Science Gateways

- Resources
  - Compute, data, software, visualization availability and support
  - Different types of accounts
  - Allocation process
  - External relations

- Services
  - Information
  - File
  - Job
  - Account
  - Requirements
  - Security
  - Accounting

Effectively using TeraGrid Compute Resources

Not all TeraGrid computing resources are created equal

- Hardware
  - Tightly coupled parallel systems
  - Grid computing (high throughput)
- Software
  - Job Submission Interfaces
    - Different sites have different job schedulers and different submission interfaces
  - Job Scheduling Interfaces
    - Interactive or on-demand computing resources are limited today on TeraGrid
    - Dedicated computing time for time-critical projects requires additional effort
  - Software Applications can very
    - Common TeraGrid Software Stack (CTSS)

Effectively using TeraGrid Data Resources

- User Home Directories
- Temporary Scratch Space
- High Performance File Systems
- Long-Term, Archival (Tape) Storage
- Storage Resource Broker
- Community Software Areas (CSA)

TeraGrid Security& Audit Requirements

- Make Use of Community Accounts
- Institute a User Registry (Users accessing TeraGrid via the Gateway)
- Maintain An Audit Trail
  - Includes usage information for each user
  - Restrict/remove access for user on demand (security protection)
  - Ability to limit user usage
- Collect User Authentication/Authorization Information
  - Includes login attempt/failure logging
- Include User Attributes in Community Credentials
- Collect Accounting Statistics
- In an Emergency, Provide Administrative Gateway Access to the TG
- Security WG

TeraGrid Reporting Requirements

- Identify researchers who have benefited from the Gateway
    - Important for requesting future TeraGrid resources
    - Useful for your own funding efforts.
    - Ensure credit to TeraGrid and the Science Gateway is given by users in their papers & presentations.