Data Infrastructure in the TeraGrid

Chris Jordan
TeraGrid Area Director for Data
Texas Advanced Computing Center

TeraGrid™

TeraGrid Basics

- NSF-Funded project, now in 6th year of operation as "TeraGrid"
- 11 U.S. Open Science Supercomputing sites
- Several petaflops aggregate computational power
- Visualization and Data Analysis resources
  - Tiled displays, remote viz, large-memory systems, etc
- Coordinated deployment of software and services to enable mobility in research computation
- Capabilities possible only in distributed environment

TeraGrid™

TeraGrid™

Basic Data Infrastructure

- Parallel File Systems
  - Over 10 PB total, ~1-2PB on average
- Archive (tape) systems
  - Over 50PB capacity total
- Wide-Area File Systems
  - Indiana Data Capacitor
  - SDSC GPFS-WAN
- Data Transfer with GridFTP
  - Striped/Parallel transfers, up to 30Gb/sec peak
- Portal-based Data Management

TeraGrid™

Science Gateways

- Large projects with custom web-based interfaces to TeraGrid resources
- Atmospheric Simulation, Mechanical Engineering, Astronomy, etc
- Often implement composable or pre-composed workflows, including data movement and computation
- Both curated and user-provided data sets
- TeraGrid provides data, compute, viz and software infrastructure through grid services

TeraGrid™

Current Major Deployments

- Distributed Lustre-WAN
  - Storage at 6 TeraGrid sites with ~200TB each
  - 20Gb/sec network capacity per site
  - In Production 3Q 2010
- Data Replication Service
  - Diverse Storage Systems at 5 TeraGrid sites
  - iRODS Data Grid for replication/data management
  - Intended for long-term, reliable data storage

TeraGrid™

Distributed Lustre-WAN

- Wide-Area File Systems
  - Single, Parallel file system available on multiple resources
  - Remove need for explicit data movement
  - Take advantage of TeraGrid network infrastructure
  - Ideal for multi-resource workflows, large shared collections of data used in computations
- In-progress Lustre-WAN deployment
  - Uses distributed (DDN+Dell) storage to minimize latency
  - Introduces potential data management issues
  - Will be available on almost all TeraGrid resources
  - Provides first "TeraGrid-wide" WAN file system capability

TeraGrid™

Data Replication Service - Motivation

- Historically, Archive usage has been controlled by each site and isolated to each site
- Many users cross sites and resources
- Long-time TeraGrid users build repositories of code, simulation input/output, parameter files
- Increasingly, new users have data "collections" or otherwise specialized data-centric needs
- TeraGrid sites can enter and leave the project, but data must remain "in the TeraGrid"
- Growing need for reliable, long-term storage, with data mobility across sites

TeraGrid™

Data Replication Service - Implementation

- iRODS provides fast data transfer and virtual namespace on 6 archive and online resources
- Arbitrary metadata - controlled by users
- Policy-based management automates data placement and replication
- Variety of interfaces provides for easy integration into existing infrastructure
  - Command-line tools on HPC systems
  - Portal integration through GridFTP interface
  - WebDAV for Desktop integration

TeraGrid™

Architectural Plans

- Moving towards allocation system with TeraGrid-wide data services
- Allocation provides usage control and well-defined lifetime for data
- Also improving user support for data management
- Growing need suggests change of funding and usage pattern will become necessary
- Need to expand community/agency support for large-scale, persistent storage

TeraGrid™

TeraGrid and DataNet

- Current DataNet awards do not include significant infrastructure
- DataNet will not handle basic infrastructure needs
- Significant research programs in metadata, integration, sustainability, etc
- Coordinating with all DataNet awardees
- Many other large-scale science projects/data challenges
- Assume greater need for TeraGrid as backend data infrastructure for data-centric science

TeraGrid™

Thanks

- Questions/Comments?

TeraGrid™