Data Grid Collaborations

Adil Hasan adilhasan2@gmail.com
Yutaka Kawai yutaka.kawai@kek.jp
Ueng WeiLong wlueng@twgrid.org
Mike Conway michael_conway@unc.edu
Reagan Moore rwmoore@email.unc.edu

Data grid capabilities have evolved from simple organization of distributed collections to now include enforcement of management policies, automation of administrative functions, and validation of assessment criteria. The new capabilities are enabled through the integration of a distributed rule engine into the data management infrastructure. Each policy is mapped to a computer actionable rule that controls the execution of a data management procedure. At each storage location, a rule engine applies the procedures. The procedures, in turn, are mapped to workflows composed from standard functions, called micro-services. The results of the execution of each procedure are stored as persistent state information in a metadata catalog. The state information can be queried to verify assessment criteria.

This policy-base data management approach makes it possible to build generic infrastructure that can support each stage of the data life cycle. The policies and procedures required for simple distributed data management (traditional data grid applications) can be augmented with the policies required for data publication in a digital library or data preservation in an archive. The same data management infrastructure can be used to re-purpose a collection for a new use, by applying the policies required by the new user community. An example of a policy-based data management system is the integrated Rule Oriented Data System (iRODS). This open source software is available from http://irods.diceresearch.org.

The demonstration of the generic capabilities of policy-based data management systems is being done in international collaborations that include groups in Japan, Taiwan, Australia, Europe, and the United States. Data grid applications in Asia of the iRODS technology include the T2K neutrino data grid in Japan, development of gLite-iRODS interoperability at Academia Sinica in Taiwan, and the Taiwan Digital Archives Remote Backup system.

At the Supercomputing Conference 2011 in Seattle, Washington, a demonstration of file reconstruction from five international sources was presented by Yutaka Kawai (KEK). The demonstration involved the federation of five independent iRODS data grids, the distribution into each data grid of a bubble chamber photograph that had been subdivided into 512 parts, and the reassembly of the photograph on the SC'11 exhibit floor. The five independent data grids were installed by collaborators at:

1. Academia Sinica, ASGC (Taiwan) V. Huang, H-W. Chen, H-W Wu, E. Yen
2. French National Institute for Nuclear Physics and Particle Physics, CN2P3 (France) Y. Cardenas, J.-Y. Nief

3. High Energy Accelerator Research Organization, KEK (Japan) Y. Iida, Y. Kawai, T. Sasaki, W. Takase
4. Queen Mary University of London, QMUL (UK) F. Di Lodovico
5. Renaissance Computing Institute, RENCI (North Carolina) R. Moore

The five data grids were federated under the direction of Adil Hasan, to form a testbed for evaluating international data transfer rates. The bubble chamber image reconstruction demonstration was designed to show recovery from storage errors within the distributed environment. As shown in Figure 1, parts of the bubble image were retrieved from each of the five collaborating data grids. When a part of the image was unavailable, the system retrieved the image part from an alternate data grid, and rebuilt the complete file.

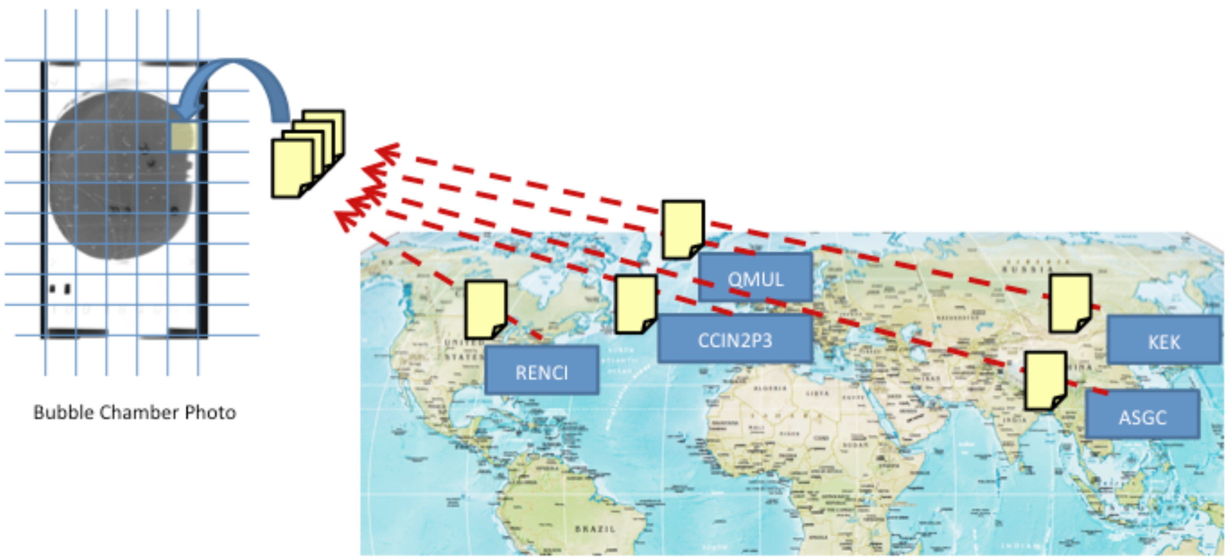

Figure 1. Image reconstruction from five federated data grids

A second goal was the demonstration of the reliable transfer of 800,000 files at a time between independent data grids. Adil Hasan implemented a test environment on a local computer that held 500,000 1-Megabyte files. The files were ingested into the QMUL data grid over a local network that sustained 19 MB/sec. While the rate was limited by the performance of the local systems, the ingestion time scaled linearly with the number of files.

A second test replicated 800,000 files from the QMUL data grid to the ccin2p3 data grid. Ten iRODS copy commands were run in parallel to replicate the data. The replication proceeded without any errors, and the time to replicate a single file remained constant. The test was conducted across a two-day time period, with occasional outages while machines were offline.

The underlying transport mechanisms used in the demonstration relied on the following capabilities:

- Use of large-block transfers within the TCP/IP protocol

- Encapsulation of small files (less than 32 Mbytes in size) within the request to send
- Use of parallel I/O streams to fill the network pipe
- Automated restart across network interruptions
- Support for re-transmission of large files from the last successful buffer

The federation of local iRODS data grids to form an international data management system was a key part of the SC'11 demonstration. This is being augmented through development under way at Academia Sinica of interoperability mechanisms between the iRODS data grid and gLite. Ueng Weilong (ASGC) is developing a Storage Resource Manager (SRM) version 2.2 interface for the iRODS data grid.

The Storage Resource Manager is based on a widely adopted common interface specification used for accessing storage resource elements within the gLite data grid. By implementing the same interface for the iRODS data management system, gLite users will be able to access data within an iRODS data grid. The expected use cases include: 1) Make iRODS an archival system for gLite-based e-Infrastructure, 2) Support flexible lifetime policies for files - volatile, durable, and permanent, and 3) Impose the VO-based resource policy and security control on iRODS for integration with the Grid infrastructure.

An initial implementation has been demonstrated that allows gLite tools to transfer files between SRMs and iRODS. In effect, the interface makes an iRODS data grid look like a classic Storage Element. At the same time, the iRODS data grid capabilities enable the Storage Element to include storage resources located in multiple administrative zones. The iRODS federation capabilities enable the Storage Element to be composed from multiple independent data grids. Policies within iRODS can then enable the automated replication, distribution, and caching of files on an international basis.

Additional interfaces are under development in the Data Intensive Cyber Environments Center at the University of North Carolina at Chapel Hill. The iDrop client monitors the transfer of each file from a local resource into the iRODS data grid, tracks the completion status of each data transfer, and maintains a log of all transfer results. A queue of outstanding file transfer requests is maintained. The iDrop interface supports the dragging and dropping of a local folder onto the data grid, with subsequent monitoring of the transfer of each file. The interface has been used to move 300,000 files at a time from a local resource into an iRODS data grid.

The iDrop interface also supports synchronization of local directories with iRODS data grid directories. The system periodically examines the local directory, identifies any new or changed files, and uploads the files into the data grid. Changed files are registered as versions of the original file, with the versioned files stored in a version directory under a similar path name. By enabling appropriate policies, replication and distribution of the loaded files can then be processed by the iRODS data grid, automating administrative tasks.