

SciDAC DataGrid Middleware
A High-Performance Data Grid Toolkit:
Enabling Technology for Wide Area Data-Intensive Applications
Quarterly Report October 2002 thru December 2002

Accomplishments this Quarter:

Globus XIO (eXtensible IO system) Proof of Concept Prototype Completed

We did a proof of concept implementation of our XIO design. It was decided that the interface was much too heavy weight, and we simplified it a great deal. There are now no factories, we have a single method for driver specific functions (fcntl), and for performance reasons, we have separated the connection acceptance code into a separate "server" which will make servers built with XIO more scalable.

Globus XIO Coding is progressing

We have begun coding against the revised design. We will make adjustments as necessary based on reviewer feedback, but believe the design is solid and changes should be minimal. This code will be the underlying IO system for the new GridFTP server.

HPSS

Work with HPSS continues and appears to be going well. Timelines mesh well. We are projecting release 5.2 of HPSS (Q1 or Q2 of Calendar 2004) will be using GridFTP as its interface and will be using the new server code. We are working on getting an NDA in place so we can have access to the HPSS source code.

GridFTP at SC2002

GridFTP was used in numerous demos at SC2002. This included GriPhyN in support of Chimera computations, LIGO for data replication, NVO for Chimera runs, PPDG demos, ESG demos (using the striped server), and many others. GridFTP is the newest component of the toolkit and by far the heaviest used with the exception of GSI security.

Investigation of IGrid problems

Significant effort was expended investigating and understanding the issues encountered while trying to utilize the 10 GigE, high latency link between Chicago and Amsterdam during IGrid 2002. The problems were determined to be a combination of HW problems on our Compaq servers, overly conservative network code in the Linux kernel, and fundamental problems inherent in the TCP/IP protocol. A paper has been submitted and accepted. We are working with Compaq (HP) to resolve the HW issues.

Migratory File Services

Designed and implemented a migratory file service built with Condor and NeST. This service adaptively manages computing and I/O resources, allocating each to specialized roles according to application load and network availability. The results are currently in review.

Integration of CPU and I/O Management

Integrated the Chirp I/O protocol into the CMS ROOT I/O library, allowing for the simplified bootstrapping of existing applications into a secure I/O proxy structure, in a manner to that developed for Java applications in Condor.

Scientific Workload Analysis

Developed a method for analyzing the data requirements of integrated scientific applications. Applied to a collection of “batch pipelined” applications, results are currently in review for publication. A preliminary technical report is mentioned below.

Storage Structure Semantics

Developed the naming and synchronization semantics for storage devices that serve as rendezvous points between running processes. They are known as “sparse files.” This is a necessary component for “batch pipelined” applications, such as those mentioned above, which must execute in an unreliable environment. The results are currently in review. A preliminary technical report is mentioned below.

Languages for Simplified Fault Tolerance

Continued development of a closed scripting language for controlled reliability in the face of common failures. This tool, the Fault Tolerant Shell, is now deployed as part of the iVDGL Virtual Data Toolkit.

Continued development, packaging, testing and deployment of the Replica Location Service

Continued development on the Replica Location Service, including debugging for greater robustness, more testing, performance evaluation, and packaging of the RLS with the Globus Toolkit version 2.2.3. The RLS was deployed widely, including an RLS testbed of over 30 machines on three continents that was used for demonstrations during SC2002. The RLS was also used in GriPhyN and Earth Systems Grid demonstrations at SC2002. The RLS saw greater interest and testing from other groups, including PPDG, the European DataGrid project, IBM and others.

Turning the Replica Location Service into an Open Grid Services Architecture Service

This quarter, we developed a first draft of a specification document for an Open Grid Services Architecture Replica Location Service. This document was presented at the Replication Research Group of the Global Grid Forum in October 2002. We proposed an OGSA Data Replication working group of the Global Grid Forum through which we intend to standardize interfaces to replication grid services.

Continued development of Metadata Catalog Service

Continued development and experimentation with the prototype Metadata Catalog Service. We worked extensively with two applications (Earth Systems Grid and Laser Interferometer Gravitational Wave Observatory) to deploy the MCS, load it with application-specific metadata and query the metadata to identify data items by attributes. These application metadata catalogs were used extensively in project demonstrations of ESG and LIGO at SC2002. We identified some deficiencies of the current prototype design that need to be addressed in the next phase of

MCS development. These include: 1) the XML format of many metadata files does not map very naturally to the relational database back-end of the MCS prototype and 2) the MCS needs to support a richer set of metadata queries. We also spent time this quarter evaluating the functionality and performance of alternative database backend technologies, such as native XML databases.

Plans for next quarter

Globus XIO design will be distributed for external review and comment

The XIO design documents will be made available to a limited set of reviewers for comment. We are particularly looking for how they might employ XIO and would it meet their needs. Design specs and implementation will be updated as necessary based on feedback.

Globus XIO available for early Testing

We should have XIO code that works including the framework, file, and TCP drivers. GSI and UDP should be close behind. We also have a student implementing an RBUDP and Tsunami driver (both are reliable UDP protocols from UIC and IU respectively)

Initial draft of server design complete

By the end of March we should have a rough draft of the server design document complete.

Integration of CPU and I/O Management

Currently adapting the secure I/O proxy to communicate with third party storage devices such as NeST. Will unify the bootstrap protocol and the existing NeST protocol. Plan to develop a consistent semantic for the interaction between I/O and CPU managers in the case of failures.

Storage Structure Semantics

Will integrate prototype of the "sparse file" interface into existing storage devices such as NeST. Will deploy application workloads that make use of these sparse files and explore the implications of fault-tolerance.

Languages for Simplified Fault Tolerance

Currently applying results from computability theory to extend fault-tolerant languages to greater generality without loss of safety. Will explore the transformation of procedural structures into declarative structures such as DAGMan for simplified checkpoint and recovery.

Release of RLS as part of the GT3 Alpha release, January 2003

RLS was packaged for release in the GT3 Alpha release, which occurred on January 13, 2003.

Widespread deployment of Replica Location Services

An increasing number of projects (LIGO, ESG, EDG, etc.) are deploying and using the RLS. With the GT3 alpha release of the RLS code, we expect to see the deployment of RLS increase dramatically in the coming quarter. With wider use, we expect to spend extensive time in the coming quarter doing additional debugging, documentation and testing of the software as well as adding some new features.

Development of Replica Location Grid Service Specification and GGF Working Group

We are awaiting final approval on the existence of the Replication Services Working Group within GGF. We will present a new version of the Replica Location Services Specification at the March meeting in Tokyo, Japan.

Re-design of the Metadata Catalog Service

In the coming quarter (and throughout the year 2003), we will be working on a re-design of the MCS based on the lessons learned from the MCS prototype. We will be doing a thorough evaluation of alternatives for database backend technologies, including a variety of relational and native XML databases. We will evaluate the OGSA Database Access and Integration Service to see whether it is possible to use this generic database service as a back-end for the MCS. We will work on mediation of multiple metadata schemas and on federating multiple metadata catalogs. In the next quarter, we will also begin specification of a grid service interface for the MCS. We will continue working with application scientists on ESG, LIGO and other projects to make sure that the new MCS design more adequately meets the needs of applications.

Papers Published or in Progress

Grid-Enabled Particle Physics Event Analysis: Experiences using a 10 Gigabit, high latency Network for a High Energy Physics Application. W. Allcock, J. Bresnahan, J. Bunn, S. Hedge, J. Insley, R. Kettimuthu, H. Newman, S. Ravot, T. Rimovsky, C. Steenberg, L. Winkler
Accepted for publication in the IGrid 2002 special edition of Future Generation Computer Systems

John Bent, Andrea Arpaci-Dusseau, Remzi Arpaci-Dusseau, and Miron Livny, "Migratory File Services for Scientific Applications," currently in review.

Douglas Thain, John Bent, Andrea Arpaci-Dusseau, Remzi Arpaci-Dusseau, and Miron Livny, "The Architectural Implications of Pipeline and Batch Sharing in Scientific Workloads", Technical Report 1463, Computer Sciences Department, University of Wisconsin, January 2003, also currently in review for publication.

Douglas Thain and Miron Livny, "The Case for Sparse Files", Technical Report 1464, Computer Sciences Department, University of Wisconsin, January 2003, also currently in review for publication.

Douglas Thain and Miron Livny, "Error Management in the Pluggable File System", Technical Report 1448, Computer Sciences Department, University of Wisconsin, October 2002, also currently in review for publication.

Giggle: A Framework for Constructing Scalable Replica Location Services. Ann Chervenak, Ewa Deelman, Ian Foster, Leanne Guy, Wolfgang Hoschek, Adriana Iamnitchi, Carl Kesselman, Peter Kunszt, Matei Ripeanu, Bob Schwartzkopf, Heinz Stockinger, Kurt Stockinger, Brian Tierney. Published in Proceedings of the SC2002 Conference in Baltimore in November, 2002.

Presentations Given

John Bent, "Lot Management in Nest," Globus World, January 2003.

Douglas Thain, "Error Management in Condor," PPDG Troubleshooting Workshop, December 2002.

October 2, 2002: Presented RLS design as part of PPDG teleconference on reliable replication.

October 15-17, 2002: Presented update on RLS to Global Grid Forum Replication Research Group. Presented first draft of grid service interface for replica location service.

November 21, 2002: Presented paper on Replica Location Service to SC2002 Conference in Baltimore, MD.

November 18-21, 2002: Gave demonstrations of Replica Location Service testbed at SC2002 conference in Baltimore, MD. RLS was also included in demonstrations of the Earth Systems Grid and Laser Interferometer Gravitational Wave Observatory projects.

November 18-21, 2002: Demonstrated MCS at SC2002 conference in Baltimore, MD, as part of demonstrations of the Earth Systems Grid and Laser Interferometer Gravitational Wave Observatory projects.

December 3, 2002: Presentation of Globus view of grid architecture for data management to IBM data grid researchers at IBM Almaden research center. This included descriptions of the RLS and MCS.

December 3, 2002: Presentation of Plans for GridFTP in GT3 to IBM data grid researchers at IBM Almaden research center.