

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

Accomplishments this Quarter:

Completed first design iteration on Globus XIO (extensible IO system)

Globus XIO will provide a protocol pluggable read/write/open/close interface for IO. This will allow an application written with XIO to easily change the underlying transport protocol. The new GridFTP server will be built using XIO and this will allow us to easily provide a non-TCP GridFTP implementation, one of our project deliverables. We will also provide a GridFTP driver for XIO allowing other application to more easily add GridFTP support.

Participated in IGrid 2002

Collaborated with Caltech on a demonstration involving analysis of physics data on the show floor in Amsterdam that was stored in Chicago. Striped GridFTP was used to move the requested data sets via StarLight over a trans-Atlantic 10 Gigabit Ethernet link. The performance was significantly less than expected (about 1 Gbs). Investigations of the root cause of the problem are underway and a paper describing this is underway.

Transport Activities in GGF

The GridFTP working group is nearing completion of the specification for V1.0. We project having the final version submitted prior to GGF7. Also held a BOF proposing the formation of a research group on data transport issues. There was fair attendance, but a lack of people willing to do volunteer to do concrete work. We are continuing to pursue formation of this group.

Technical Meeting to discuss HPSS / GridFTP integration held

IBM / HPSS has committed to using GridFTP as the interface to HPSS. A technical meeting was held to discuss possible issues, interaction logistics and timelines. We are targeting release 6.2 of HPSS in early to mid 2004 for production release of HPSS with GridFTP as its access protocol.

Continued development and testing of the Replica Location Service

During this quarter, we continued our development and testing of the Replica Location Service. Features added to the RLS this quarter include: 1) Improved packaging and documentation 2) Implementation of bloom filter compression 3) Implementation of a Java version of the client API that wraps the existing C client; 4) Implementation of a gridmap file for the RLS servers that supports GSI authentication.

Deployment of RLS into collaborator Testbeds

The Earth Systems Grid project has deployed a Local Replica Catalog as part of their testbed. The ESG LRC contains several thousand mappings of file replicas from the PCM and CCSM climate data sets. The RLS is also being deployed throughout the extensive testbed of the European DataGrid project. We continue our testing efforts with Work Package 2 of the EDG project, and the two projects plan to participate jointly in a demo at the upcoming SC2002 conference.

New Initiative to Turn the Replica Location Service into an Open Grid Services Architecture Service

In August and September 2002, we began discussions internally and with IBM on the possibility of turning the Replica Location Service into a Grid Service compatible with the Grid Services Specification of the Open Grid Services Infrastructure working group of the Global Grid Forum.

A two-day meeting that focused on opportunities for collaboration between Globus and IBM on data grid architectures and services was held at USC Information Sciences Institute on 12-13 September 2002.

Metadata Catalog Service Prototype Completed and Used in Earth Systems Grid

During this quarter, we completed implementation of a prototype of the Metadata Catalog Service. The MCS service maintains metadata that describes the contents of logical files in a data grid. The Metadata Catalog Service (MCS) allows users to query based on attributes of data rather than data names. In addition, the MCS provides management of logical collections and logical views of files. Our prototype uses a web server front end and a relational database back-end to implement the schema described in the Metadata Catalog design document that was produced last quarter. The original MCS design was expanded in the current quarter to provide better support for XML metadata, including shredding the data and storing it efficiently in relational tables as well as providing the capability to reconstruct the original XML files. The prototype implementation includes a Java client API for loading and querying the MCS. An MCS has been incorporated into the Earth Systems Grid testbed and currently holds metadata for the PCM and CCSM climate models. The ESG MCS will be used as a central part of the ESG prototype for SC2002.

Grid Reliability Workshop at USC Information Sciences Institute

A two-day workshop on reliability issues for grids was held at USC Information Sciences Institute on 10-11 July 2002. Approximately 15 people from several groups including Globus, Condor, San Diego Supercomputing Center and Fermi Lab attended the meeting and participated in discussions on grid reliability issues and approaches for improving reliability.

Developed “migratory file services” for scientific applications.

This system performs caching and some prefetching for applications running in a remote environment. Academic paper in progress.

Continued maintenance and development of the NeST grid storage appliance.

More accounting and maintenance features added as well as enhanced support for migratory users.

Additionally, the following task were also accomplished by the Condor team:

- Applied the fault-tolerant shell (<http://www.cs.wisc.edu/~thain/research/ftsh>) to solve reliability problems in the CMS distributed terascale facility (DTF) production run.
- Developed the “sparse file” abstraction for reliable I/O interactions in a distributed system. Academic paper in progress.
- Hardened the Pluggable File System (<http://www.cs.wisc.edu/condor/pfs>) according to user needs.
- Improved error-handling facilities and documented them in a technical report in progress.
- Hardened the Java I/O support in Condor in order to deal with the vagaries of various Java implementations. Now in production use at several sites.
- Modified and re-deployed the FTP-Lite package (http://www.cs.wisc.edu/ftp_lite) to accommodate changing FTP protocols.

Plans for next quarter

Continued development and testing of Replica Location Service

We plan to continue development and testing of the RLS in the coming quarter. However, to some extent, the current set of features has frozen until we get additional feedback from our user community about whether additional functionality is required, so our efforts will be focused more on testing and debugging the code. Planned development for the coming quarter includes making the RLS GT3 security compatible, which requires that the RLS accept the new GSI certificates that will be developed as part of the Globus Toolkit version 3.

Widespread deployment of Replica Location Services

We will focus on widespread deployment of RLS in the coming quarter. We will deploy a fairly large RLS testbed as part of a demonstration of RLS for the SC2002 conference. This testbed will include approximately twelve replica location index nodes and a larger number of local replica catalogs. Development for the testbed includes creating specialized information providers for the MDS information system to reflect the state of RLS catalogs in the testbed.

In addition to the RLS testbed, we will continue to work with other applications, including ESG, LIGO and PPDG, on deploying RLS systems for their testbeds.

Development of Replica Location Grid Service Specification and GGF Working Group

In the coming quarter, we plan to develop a first draft of a specification document for an Open Grid Services Architecture Replica Location Service. This draft specification will be developed under a proposed OGSII Data Replication working group of the Global Grid Forum.

Continued development of Metadata Catalog Service

We will continue developing and experimenting with the MCS. We plan to spend some time in the next quarter evaluating the preferred back-end technology for the MCS. In particular, we will evaluate whether the current relational back-end should be retained or replaced with a native XML database back-end. In addition, we will work on extending and generalizing the current Java API to support general XML queries. We will continue working with applications such as ESG and LIGO that are actively deploying the MCS and evaluating its functionality and performance.

Proof of Concept implementation of Globus XIO

We will implement a rough proof of concept version of XIO to evaluate our design. It is likely we will then iterate the design to refine it based on our experience. We also intend to release this refined design to a group of "trusted friends" for review and comment.

Investigate Performance Issues Discovered at IGrid 2002

We intend to keep the 10 GigE trans-Atlantic link operational for further testing and investigation in to what is needed in order to utilize high bandwidth, high latency links. Hopefully, we will be able to utilize this knowledge to obtain better performance during the Bandwidth challenge at SC2002.

Papers Published or in Progress

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

John Bent, Douglas Thain, Andrea Arpaci-Dusseau, Remzi Arpaci-Dusseau, and Miron Livny, "The I/O Requirements of Batch Parallel Sequential Jobs", in preparation.

Douglas Thain and Miron Livny, "The Case for Sparse Files," in preparation.

Douglas Thain and Miron Livny, "Fault Correctness in the Pluggable File System," in preparation.

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. To appear in Proceedings of SC2002 conference. Final copy submitted July 26, 2002.

High-Performance Remote Access to Climate Simulation Data: A Challenge Problem for Data Grid Technologies. Ann Chervenak, Ewa Deelman, Carl Kesselman, Bill Allcock, Ian Foster, Veronika Nefedova, Jason Lee, Alex Sim, Arie Shoshani, Bob Drach, Dean Williams and Don Middleton. Submitted to Parallel Computing Journal.

Data Grid Architecture. William Allcock, Ann Chervenak, Ewa Deelman, Ian Foster, Carl Kesselman, Peter Kunszt, Erwin Laure, Arie Shoshani, Heinz Stockinger, Kurt Stockinger, Douglas Thain. In progress.

Presentations Given

Aug 2002 John Bent, "Flexibility, Manageability, and Performance in a Grid Storage Appliance," HPDC-11, Edinburgh, Scotland

Aug 2002 Douglas Thain, "Error Scope on a Computational Grid," HPDC-11, Edinburgh, Scotland

July 2002 Douglas Thain, "Errors on the Grid: Theory and Practice," GriPhyn Reliability Workshop, Marina del Rey, California

10-11 July 2002: Workshop on Grid Reliability at USC Information Sciences Institute

20 July 2002 Presented "The Globus Toolkit and OGSA" at Sun High Performance Computing Consortium Meeting in Glasgow, Scotland

22 July 2002: Presented Giggle Replica Location Service to Replication Research Group at Global Grid Forum Meeting in Edinburgh, Scotland.

22 July 2002: As Working Group Chair of Global Grid Forum Replication Research Group, led working group discussion of work of several groups on data replication in grids.

23 July 2002 Conducted GridFTP working Group meeting at Global Grid Forum Meeting in Edinburgh, Scotland.

24 July 2002 Conducted BOF session for formation of Data Transport Research Group at Global Grid Forum Meeting in Edinburgh Scotland.

August 2002: Presented Replica Location Service and Metadata Catalog Service to LIGO group meeting.

12-13 September 2002: Meeting between Globus and IBM on opportunities for collaboration on data grid architectures and services. Presented overviews of the Globus approach to data grid architecture, the Giggle Replica Location Service, the GridFTP protocol and the Reliable File Transfer Service.