Grid File System Working Group

Global Grid Forum, Data Area

Administrative Information

Name and Acronym:
Grid File System Working Group (GFS)

Chairs:
Osamu Tatebe, o.tatebe@aist.go.jp
Jane Xu, jxu@us.ibm.com
Arun Jagatheesan, arun@sdsc.edu

Secretary(s)/Webmaster(s):
Yves Denneulin, Yves.Denneulin@imag.fr
Pradeep Padala, ppadala@cise.ufl.edu

Email list:
gfs-wg@ggf.org

Web page:
http://forge.ggf.org/projects/gfs-wg/

Charter

Focus/Purpose
Data in a grid can be of any format and stored in any type of storage systems. There can be many hundreds of petabytes of data in grids, among which a very large percentage is stored in files. A standard mechanism to describe and organize file-based data is essential for facilitating access to this large amount of data. The Grid File System Working Group (GFS-WG) will provide specifications of Grid File System Directory Services and Architecture of Grid File System Services.

The specification of Grid File System Directory Services will describe and manage the namespace of federated and virtualized data from file system resources, access control mechanisms, and meta-data management.

The specification of Architecture of Grid File System Services will specify the hierarchical structure to facilitate federation and sharing of virtualized data from file systems in the grid environment by providing the virtual namespace that will allow association of access control mechanisms and meta-data of the underlying physical data sources.

Scope
The GFS-WG will first focus on the specification of the Grid File System Directory Services. The specification will provide features such as (a) virtualized hierarchical namespaces for files or potentially other type of data (such as live data feeds), (b) efficient and transparent file sharing, and (c) flexible management of inter-organizational data access controls, and (d) ability to describe and manage file-system and application-specific metadata.

Then the GFS WG will provide a specification of Architecture of Grid File System Service. The specification will focus on the organization and infrastructure of file systems in the grid environment. The architecture will make the implementation of upper or lower level Grid services defined by other GGF WGs to be easily applied to file system data sources.

In order to produce the specifications, the GFS-WG will create a quick survey of the state-of-art technologies to facilitate the discussion for the standardization

Goals
1. Recommendation Document: "**File System Directory Services Specification**"
    This document will describe a specification of a grid service to manage virtual file system directories in association with access permission, and other application-specific metadata.
    a) Outline by GGF 10
    b) Draft by GGF 11
    c) Updated draft by GGF 12
    d) Final by GGF 13

2. Recommendation Document: "**Architecture for Grid File System Services**"
    This document will describe a recommended architecture for a grid service to provide a functionality of the grid file system. This service will provide POSIX like I/O operations using the File System Directory Services and other common services.
    a) Outline by GGF 11
    b) Draft by GGF 12
    c) Updated draft by GGF 13
    d) Final by GGF 14

Two pre-study based information documents will be produced before the Recommendation Documents:
    a) **A Survey of the Major Grid File Systems**
    b) **Requirements and use cases for Grid File Systems**

Management Issues

Evidence of commitments to carry out WG tasks
Osamu Tatebe has working on Grid File System in the Grid Datafarm project since 2000. The papers have been published in proceedings of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2002) and in Proceedings of the 12th IEEE International Symposium on High Performance Distributed Computing (HPDC-12). He can commit the time required for this WG supported by a funding agency.

Pre-existing Document(s) (if any)
Ron Oldfield, "Summary of Existing and Developing Data Grids", GGF4, 2002
Reagan Moore, et al., "Data Grid Implementations", 2002
Leo Luan, Ted Anderson, "Grid Namespace for Files", GGF8, 2003

Exit Strategy
The WG will be finished after submitting a recommendation documents Grid File System Directory Services Specification, and Architecture for Grid File System Services Specification.

Any other relevant information

Evaluation Criteria (from GFD-C.3)

Is the scope of the proposed group sufficiently focused?
The group will focus on Grid Services which are applied to data stored in File Systems. The goal is to first produce a specification of Grid Services for File System Directories. The specification will define technologies to support federation to file data, such as virtual file system directory services in association with file system metadata and other application-specific metadata. During this process, the group will produce an architecture document for Grid File System Services exploiting the file system directory services and other services.

Are the topics that the group plans to address clear and relevant for the Grid research, development, industrial, implementation, and/or application user community?
Yes. Several Grid research projects and industrial products in US, Europe and Asia Pacific have worked on the topics, and provided several implementations. There is a strong need to have standard ontology (namespace) and service interfaces for the data grid, especially file-like data sets. The recommendations by this WG will facilitate further implementations and usage in industry and user communities.

Will the formation of the group foster (consensus–based) work that would not be done otherwise?
Yes. There are several research projects that implement the functionalities of the topics to be recommended by this working group. The formation of the group will foster standards that are greater than the work done by any single group. Proposed three co-chairs (IBM, SDSC, and AIST) are from groups that have different projects at multiple institutions across various regions of the world to implement Grid File Systems/Services. These recommendations would be possibly used in US (by projects funded by multiple funding agencies like NSF, NIH, DoE, NARA), Europe (UK Datagrid) and Asia-pacific (Japan, Singapore, Korea, Thailand, Australia, Taiwan, China). Also there has been interest from other groups working from Europe (EDG) and US (PNNL, LBL, and University of Florida).

Do the group's activities overlap inappropriately with those of another GGF group or to a group active in another organization such as IETF or W3C?
No. There is no overlap with other GGF WGs because no other group has a focus on file system directory issues and hierarchical organization of file system data. However, this group will leverage the results of other WG's, such as Data Services work in the DAIS WG, OREP WG, etc. to avoid duplicated work.

Are there sufficient interest and expertise in the group's topic, with at least several people willing to expend the effort that is likely to produce significant results over time?
Yes. There were 67 participants in the BoF session at GGF8, and 42 participants at GGF9. The participants were from various projects and different disciplines and felt that this working group related to their work. There are volunteers for every informational or community practice documents. At least, four projects; Grid Datafarm, IBM Distributed StorageTank, DoE SAM, and SDSC SRB, already indicate to implement the recommendations that will be provided by the proposed WG. Reagan Moore (SDSC), Leo Luan (IBM Almaden), Pradeep Padala (University of Florida), Andrew Mcnab (Univ Manchester), and James Myers (PNNL) have already committed their time for some of the activities of this working group. Andrew Grimshaw (UvA/Avaki) joined to the WG discussion, and provided comments and the information of his project and product. We also intend to involve the user communities from various projects who attended the BoF session in creating the use case document.

Does a base of interested consumers (e.g., application developers, Grid system implementers, industry partners, end-users) appear to exist for the planned work?

3

Yes. The output of the WG will be applicable to every Grid application that wants to have access files or use file systems as data sources to facilitate the Data Virtualization. The implementers of the being proposed OGSA Data Services could be good consumers of the WG in case file systems will be the underlying data sources. The interested consumers could also include companies and institutes like IBM, SDSC and AIST.

Does the GGF have a reasonable role to play in the determination of the technology?

Yes. The GGF is the right place for this work because we are going to provide standard Grid services.