Distributed File System for Clusters and Grids*

Olivier Valentin, Pierre Lombard, Adrien Lebre, Christian Guinet, and Yves Denneulin

Laboratoire Informatique et Distribution-IMAG
51 avenue J. Kuntzmann, 38 330 Montbonnot Saint-Martin, France
olivier.valentin@imag.fr

Abstract. NFSG aims at providing a solution for file accesses within a cluster of clusters. Criteria of easiness (installation, administration, usage) but also efficiency as well as a minimal hardware and software intrusivity have led our developments. By using several facilities such as distributed file systems (`NFSP`) and a high-performance data transfer utility (`GXfer`), we hope to offer a software architecture fully compatible with the ubiquitous NFS protocol. Thanks to a distributed storage (especially multiple I/O servers provided by `NFSP`), several parallel streams may be used when copying a file from one cluster to another within a same grid. This technique improves data transfers by connecting distributed file system at both ends. The `GXfer` component implements this functionality. Thus, performances only reachable with dedicated and expensive hardware may be achieved.

1 Introduction

Current trends in High Performance Computing have been characterized by an evolution from the super computing towards cluster computing for several years [1], thanks to an ever-increasing performance/price ratio. As clusters have started to appears in several different places, be it two rooms in a same institute or faraway countries, aggregating the large power of all those newly-born 'poor man's super-computer' has been the source of lots of works (one of the most famous being Globus, which became OGSA project a few years ago).

Such environments have some drawbacks inherent to their qualities : as they offer as lot of services (first-grade authentications, job management, reservations, ldots), they tend to become quite heavy and complex to use. Yet, all those functionalities are not always required to run in dedicated and trusted architectures based on VPN networks (see the French VTHD project[1]). Clusters evolving in such an architecture, that is clusters linked by means of high-performance links (several gigabit/s), constitutes a 'cluster of clusters' which somewhat heterogeneous characteristics (such as OS, libraries, ...). Thus, to have a useful system,

* This work is supported by APACHE which is a joint project funded by CNRS, INPG, INRIA and UJF. GXfer is a software component developed for the RNTL E-Toile (`http://www.urec.cnrs.fr/etoile/`).

[1] See `http://www.vthd.org/`

R. Wyrzykowski et al. (Eds.): PPAM 2003, LNCS 3019, pp. 1099–1104, 2004.

the requirement of easy installation, easy maintainability and adaptability to commodity hardware appeared soon at the conception phase.

Hence, to summarize the features and characteristics we wanted: a common file tree, shared by all the machines within a grid; a minimal access time to data; working efficiently on commodity hardware; aggregation of the unused disk space of clusters; data availability for all the nodes of a cluster; reading/writing of data allowed; NFS protocol [2] and coherency (temporal coherency). To achieve these aims, we have used two tools developed within our team : the first one being a distributed version of the ubiquitous NFS server for clusters, `NFSP` [3,4], the second one being an inter-cluster transfer tool, GXfer, developed to use efficiently large network pipes without requiring expensive storage systems (SAN based for instance).

After this introductory section, the related works will be shown in section 2. Then, the `NFSG` principles are shown in section 3, followed by a short evaluation of expected results in section 4. Eventually, a conclusion will give hints about future extensions.

2 Related Works

A lot of work has been carried out in the file system field, yet the issues of scalability and data sharing within a grid still is a moot point. Within a local site (LAN for instance), the most prominent is most likely to be NFS [2,5] in the Beowulf world, but cannot solve the constraints of the WAN architectures[2]. Unlike this latter system, the family of systems constituted by AFS [6], CODA [7] try to address certain issues but ignore some others (consistency, high availability, ...) Yet, none of the existing solutions seems adapted to high-performance computing, which often implies the setup of large and expensive machines *à la* GPFS [8] or more complex and intrusive solutions as what may be seen in the rising Lustre [9]. Setting an open, adaptable and efficient solution is still open to developments. Most of the current works consists in optimizing systems so as to provide better performances within cluster-like architectures. The Globus project [10] and its newer versions (OGSA) tackles grid aspects. In a similar way, our works care about the same constraints but with a released level with regards to the security and try to remain compatible with the established standards.

3 The NFSG Proposal

`NFSG` is designed to be a grid file-system. But when we say 'grid', we mean 'cluster of clusters'. In fact, this system should match the needs of several institutions federated into one grid. We think that this strong structure should be taken into account, and thus, that having a two-level system might be a good approach : at the cluster level, a file system that would serve files to local clients, and at the grid level a system to federate the lower level file systems.

[2] NFS4 aims at addressing some of these issues but is still not as widely spread as NFS2 and NFS3.