# Storage and Transfer Solutions Between Clusters

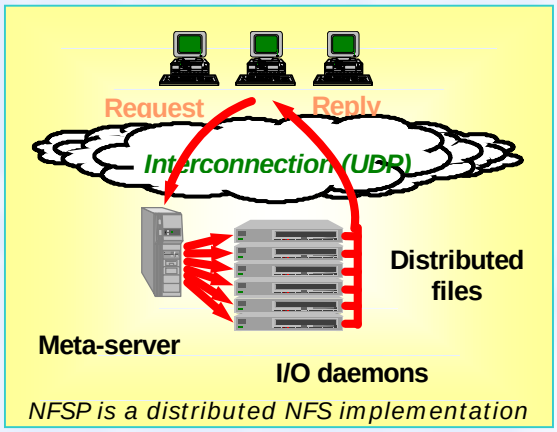
by P. Lombard, Y. Denneulin and C. Guinet

**Abstract : Beowulf (Linux) clusters are more and more used and have been in the TOP500 for several years. Most of the time, off-the-shelf hardware ships with hard-drives but systems using the unused disk space are few and far between. NFSP aims at offering a transparent aggregation of those disks through a NFS layer and at offering improved performances within a cluster. GXfer is a software layer that implements an efficient and adaptable functionality of inter-site transfer thanks to parallel file transfers. Other tools (Ka-Tools, OAR) are also developed to ease the installation and administration of such clusters.**

## NFSP

NFSP stands for *NFS Parallèle* and is a distributed version of the ubiquitous and classical NFS server. Thanks to data distribution, better performances are achieved and unused storage aggregation is offered.

The clients have a single view of the server and requires nothing else but the standard NFS client support available in most (if not all) Unix flavor operating systems.

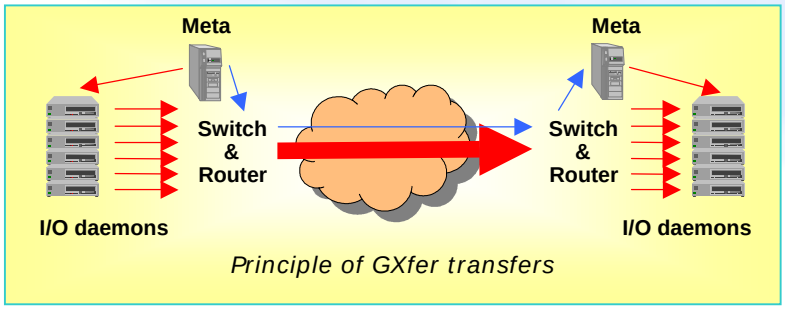



*NFSP is a distributed NFS implementation*

NFSP is an INRIA Open Source software and acts the role of storage component within the INRIA project named E-Toile – project which is carrying experimentations on clusters of clusters.

## GXfer

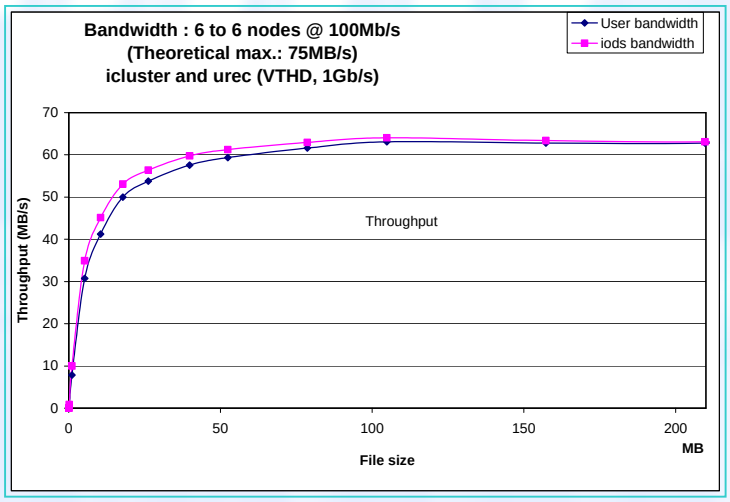
GXfer is built in a way similar to NFSP and separates the handling of meta-data and data. Yet, it is more aimed at inter-sites transfers. Its goal is offering efficient data transfers between sites, especially optimized when distributed file systems and parallel streams may be used.




*Principle of GXfer transfers*

The framework has been designed carefully to be easily extensible to other file systems,

whether they be distributed or not and will be able to cope with several authentication methods. Adding support for new file systems is only a matter of adding a C++ class implementing a simple API. This interface consists in making GXfer aware of the data location, that is where (machine, offset) is the data and where it should be written.




Thus, by using low-cost clusters, high performances may be obtained between several clusters networked by means of high speed links (long distance gigabit pipes). Recent performances have shown with 10 to 10 nodes using Ethernet 100, transfers of 1GB in 10s.

Preliminaries performances are drawn in the curve above and show good performances with standard hardware (100Mb/s ethernet cards, ethernet switches, gigabit routers).

GXfer is also a component of the INRIA project, E-Toile. It is funded by the RNTL (similarly to a NSF project) and is available under an Open Source license and will ship with the complete set of E-Toile tools (scheduler, information system, DSM, applications, …) by the end of the year.

## More Clustering Solutions

Other light weight tools have also been developed by the project APACHE to ease the installation and administration of Beowulf clusters such as in the ID-IMAG + HP I-cluster project, a Beowulf cluster based on completely mainstream and untuned hardware.

*Some of the I-cluster nodes (i-vectra 733MHz)*

### Ka-tools

They are a set of software tools used to ease the installation and administration of clusters and gather facilities of ultra-fast node installation (225 nodes installation in 15 minutes), highly optimized command launching time (hundreds of nodes within a few seconds). The Ka-tools (GPL) are now merged into the Mandrake Linux cluster distribution known as CLIC.

### OAR

OAR is a fast batch system used to schedule jobs within a cluster. OAR is based on standardized Open Source components such as MySQL and is written in perl for easy customization. It uses the Ka-tools to optimize access to huge clusters and is of course licensed under the GPL license.

**Contact : Yves Denneulin <yves.denneulin@imag.fr>**

**WWW : http://www-id.imag.fr/Logiciels/**