



CONTRACT NUMBER 508830

DEISA
**DISTRIBUTED EUROPEAN INFRASTRUCTURE FOR
SUPERCOMPUTING APPLICATIONS**

European Community Sixth Framework Programme
RESEARCH INFRASTRUCTURES
Integrated Infrastructure Initiative

Installation of AFS clients on the non-IBM DEISA machines
Deliverable ID: D-SA2-3B

Due date: October, 31st, 2005
Actual delivery date: November 25th, 2005
Lead contractor for this deliverable: RZG, Germany

Project start date: May 1st, 2004
Duration: 4 years

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	X
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Content

Project and Deliverable Information Sheet.....	Erreur ! Signet non défini.
Document Control Sheet.....	Erreur ! Signet non défini.
Document Status Sheet	Erreur ! Signet non défini.
Document Keywords and Abstract.....	Erreur ! Signet non défini.
Table of Content.....	2
List of Figures.....	2
List of Tables.....	2
1. Introduction.....	3
1.1 Executive Summary.....	3
1.2 References and Applicable Documents	3
1.3 Document Amendment Procedure	3
1.4 List of Acronyms and Abbreviations	3
2. Description and configuration	4
2.1 Introduction.....	4
2.2 AFS on non-IBM architectures	4
3. Long-Term monitoring of the core-DEISA file systems.....	5
4. Additional Progress.....	5

List of Figures

Figure1	Configuration for the test of Hierarchical MR-AFS.....	8
---------	--	---

List of Tables

1. Introduction

1.1 Executive Summary

The Service Activity 2 within the DEISA project deals with the connectivity of all DEISA-sites on the file system level. Two strategies are pursued in parallel:

- Deploying IBM's GPFS. See deliverable D-SA2-3A [2].
- Implementing a distributed file system structure for heterogeneous environment.

This document discusses the advances of the AFS implementation at DEISA since the last Deliverable, especially the integration of non-IBM architectures into the DEISA-AFS.

1.2 References and Applicable Documents

- [1] DEISA home-page: <http://www.deisa.org/>
- [2] Deliverable D-SA2-3A
- [3] Deliverable D-SA2-1B
- [4] Deliverable D-SA2-2B
- [5] Acronyms and Abbreviations:
<http://cgi.snafu.de/ohei/user-cgi-bin/veramain-e.cgi>

1.3 Document Amendment Procedure

Not applicable.

1.4 List of Acronyms and Abbreviations

AFS	Andrew File System, used in the open-source implementation OpenAFS
AIX	Advanced Interactive eXecutive (IBM's derivative of UNIX OS)
Altix	Multi-Processor compute node from SGI
AMD	CPU-vendor
DEISA	Distributed European Infrastructure For Supercomputing Applications
FS	File System
GPFS	General Parallel File System, proprietary FS from IBM.
IBM	Computer Manufacturer.
Intel	CPU-vendor
Itanium	64bit based CPU manufactured by Intel.
Linux	Free, open source UNIX clone.
MR-AFS	Multi-Resident AFS; enhanced AFS, providing hierarchical storage management
Opteron	64bit-CPU manufactured by AMD.
OS	Operating System
SGI	Computer Vendor
UNIX	Operating system family

2. Description and configuration

2.1 Introduction

Common file systems provide one of the foundations of the integrated DEISA environment. The idea behind having a common file system is that deeper integration of the DEISA sites will make it easier for developers to write upper layer software spanning the whole DEISA infrastructure.

The expected functionality such as easy installation, usage, reliability and transparency to the user by AFS has been proven in the deliverable D-SA2-1B [3], its security model and performance has been depicted in the last deliverable D-SA2-2B [4].

Since the installation of AFS at three of the four AIX core sites (IDRIS, CINECA, RZG), the AFS-client has been tested on other architectures deployed within the DEISA-infrastructure.

Since the MC-GPFS will support not only IBM-AIX systems but also SGI-Altix and other Linux systems, the MC-GPFS is a high performance solution for a global file system (c.f. D-SA2-3A [2]). Thus the main work in SA2 was done for MC-GPFS and the AFS solution was only continued with a small amount of time.

2.2 AFS on non-IBM architectures

The AFS-cell "deisa.org" is up and running in production since May 2004. It can be accessed from all AFS-clients inside and outside the dedicated DEISA network. On the DEISA test machines of three of the four DEISA core-sites the AFS client is installed and AFS is visible. Besides IBM, there are mainly three other hardware architectures used at the DEISA sites: 32-bit Intel-compatible PCs, 64-bit Opteron systems and 64-bit Intel Itanium based systems.

These architectures run mostly Linux. The AFS-client depends on the combination of the hardware and the version of the Linux-kernel, the heart of the operating system.

Many Office PCs and Linux Clusters are based on 32-bit Intel based or compatible CPUs. These are the most tested machines and the AFS-client can be seen as stable on these systems.

New Linux clusters are nowadays often based on the 64-bit systems, like AMD-Opteron or Intel-Itanium systems. At RZG, the AFS-client is currently running on more than 128 AMD-Opteron nodes with 2 CPUs each node. After solving some initial technical problems, mainly due to problems with 64-bit addressing, the AFS-client can now be seen as production-ready for these systems.

The most exotic architecture the AFS-client supports within DEISA, is the SGI Altix shared memory machine. Up to hundreds of CPUs share a single kernel and thus a single AFS-client. This means that the access pattern to and from the AFS shared file system is much different from any of the other machines described above. In fact much greater demands are made by such a kernel to the AFS-client.

In the bid to integrate an SGI Altix machine, deployed at SARA and also at LRZ, into the DEISA infrastructure, RZG has thoroughly tested the AFS-client on a SGI Altix machine consisting of 64 CPUs located at RZG. It has proven its stability in the production environment so that no fundamental problems are to be expected on the larger SGI Altix machines at SARA and LRZ.

3. Long-Term monitoring of the core-DEISA file systems

In order to understand the complex situation in the upcoming DEISA production environment, the network and the global file systems have to be monitored on a very elementary level. A simple and simple to use monitoring agent has been installed on one of the RZG machines. To simulate a regular user, this monitoring agent creates a random queue of events. Such an event is usually a file copy between two DEISA sites, but it could also be a network test, for example. This gives the possibility to coordinate the network tests with the file system tests so that they do not interfere with each other. The queue is automatically filled up by the agent with respect to some current parameters; these parameters include, among others, the size of the file to be copied, the mean period between two events, and the number of events that are allowed to run simultaneously.

A simple client-tool, started from any host within the DEISA-network, can fetch the status of the queue and change the parameters for further events, so that every DEISA administrator has the possibility of checking the status of the monitoring tool and its next steps. The results of these events are logged into a file, and are thus available for further analysis.

4. Additional Progress

Upgrade of the AFS-client to the new OpenAFS release 1.4

The development of OpenAFS is making progress on both the AIX and Linux platforms. With the help of employees of IBM and others, the OpenAFS software is advancing to the new stable-release for Unix. The latest development version incorporates many features presently not available on all Unix platforms. The presently available release-candidate is being tested on AIX power5 platforms at RZG.

Proof of concept for the hierarchical storage management within MR-AFS

MR-AFS as it is employed at RZG, provides beside the classical features of a distributed, global file system the possibility to migrate files to other storage media, for example from disk to tape, if the disk space diminishes. This feature has not been implemented yet within the AFS-cell "deisa.org". However, as a proof of feasibility, the MR-AFS features of the AFS-cell "ipp-garching.mpg.de" have been tested and used from AFS-clients belonging to DEISA. This was possible because the security model of AFS allows a user to authenticate her- or himself to different AFS-cells simultaneously.

The following Figure 1 shows the principle setup for the test of the MR-AFS features. Actually data was copied from the AFS-cell "deisa.org" into a migrating path of the AFS-cell "ipp-garching.mpg.de". The server then migrated that data to tape, from where it could later transparently be retrieved and copied back to the AFS-cell "deisa.org".

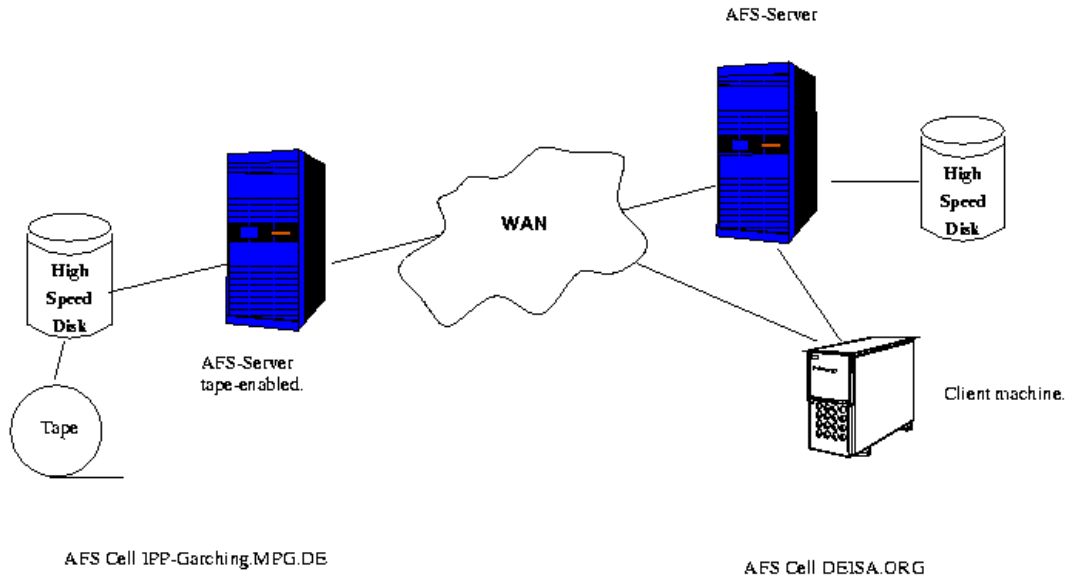


Figure1 Configuration for the test of Hierarchical MR-AFS

The results of these first tests were very promising and show that this feature is a possibility for dealing with mass-storage within DEISA. As soon as the AFS-cell "deisa.org" is configured as MR-AFS cell the data in specific directories could be transparently moved to and from tape. This configuration however requires beside the software, the appropriate licensing and the hardware (connection to tape devices) capable of the migration.