EuroStore[1] – a European Mass Storage Management Project

Martin Gasthuber
Deutsches Elektronen Synchrotron (DESY)
Notkestrasse 85 22607 Hamburg/Germany
FMartin.Gasthuber@desy.de
tel +49-40-8998-2564
fax +49-40-8998-4429

Abstract

A European consortium formed by science and industrial partners have started the EuroStore project to develop and market a Hierarchical Storage Management System (HSM) together with a high performance parallel file system (PFS). The EuroStore project aims to design and develop a high performance file store. It will combine the features of a Hierarchical Storage Manager (HSM) with the performance of a parallel file system to provide a system capable of meeting the requirements of the most demanding applications in industry, commerce, and science. The few existing HSM and parallel file system products do not fully address all the needs of the user community, and in some cases represent a major investment that users are not willing to undertake. This paper will outline the current state of the design and implementation ideas.

1 Introduction

The complete EuroStore project is divided into two major components – the HSM at the bottom and the parallel file system as its client on top. Figure 1 shows the basic overview of the complete system.
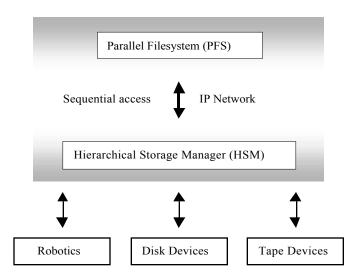

Figure 1: Basic component overview

This simple separation into two major building blocks leads to the separation of development tasks where the parallel file system and the HSM will be developed by two

different teams with minimal dependencies on each other. The complete EuroStore collaboration includes the following companies and science institutes.

- QSW (Industry / UK)
- CERN (Science / French)
- DESY (Science / Germany)
- HCSA (Industry / Greece)
- STAR (Industry / Italy)
- HNMS (Industry / Greece)
- AMC (Industry / Greece)

DESY and QSW will be the development sites for the HSM and parallel file system respectively. The proposed project was approved by the European Commission and will start on March 1, 1998, with a duration of 2 years for the first stage.

2 The two components

The following section describes the proposed development and/or extensions to the base products that will be done during the EuroStore project. Section 2.1 will focus on the parallel file system called PFS and section 2.2 focus on the Hierarchical Storage Management System called HSM.

2.1 Parallel Filesystem - PFS

The parallel file system PFS was initially introduced in the MEIKO CS2 (Computing Surface) massive parallel computer system a few years ago, and allows versatile configurations of hardware and software components among all nodes on the system. This leads to a high performance and high capacity capable implementation.

Due to its implementation (on top of standard VFS layer in the UNIX kernel) any legacy application can use PFS and its additional features like file migration, without any changes. The current implementation runs in the Sun Solaris environment and is compliant with the Sparc ABI.

Throughout the EuroStore project PFS will be enhanced and extended in the following areas:

- HSM integration – file migration capabilities
- Performance improvements – exploiting CS3 hardware capabilities
- RAID-5 capabilities in software
- Easy administration tools and interfaces
- Portability to other OS
- Globally accessible character devices (like tapes) from all nodes in a system and the ability to define logical tape devices mapped to any number of (cheap) physical tape devices connected to any node.

The HSM integration will be the main extension and will be based on the HSM API designed in cooperation with the HSM development team. The approach here is to make use of the already existing stub-file mechanism inside PFS and add a user level migration daemon plus some administration tools. To achieve high performance file access through

the whole chain of components (process – PFS – network – HSM – device) PFS will make use of the already striped layout of files and will access the HSM from all nodes with striped data from those files in parallel. This design makes complex software, dealing with striped tapes etc., obsolete.

2.2 Hierarchical Storage Management System - HSM

Unlike the PFS component, the HSM will be built from scratch exploiting the experience of DESY and other partners during the past years. The overall design goal is to make the software simple and smart. We do not rely on any third party component (except a commercially available database for persistent metadata) like DCE, nor did we plan to introduce new complexities by supporting parallel tape I/O or similar features. The HSM should work in small and large environments, supporting a large number of concurrent data streams with the maximum bandwidth the slowest component in that stream can sustain. The following simplistic list gives an overview of the proposed characteristics:

- Support of device hierarchies of any device type (random or sequential)
- Based on the IEEE Reference Model V5
- Network-centered architecture (*i.e*. unlimited number of 'mover' hosts) in a heterogeneous computing environment
- Full security for all control connection through Idea and other encryption mechanism (no DCE installation required)
- Easy integration/adaptation of new robotic and/or storage devices through simple and open API
- Easy portability (hopefully no work)
- Implementation in JAVA
- Administration/Monitoring/Controlling through HTTP/Java and SNMP. Any WWW browser can be used to connect to the HSM in a nice and secure environment, and already installed network management tools using SNMP can be used to monitor any running HSM components.
- Initially support only for sequential file access through TCP streams
- Support for automated (robot) and manual libraries
- Use of IP protocols for control and data communication
- Logical (and also physical) separation of control and data communications

Besides the listed characteristics we plan to add features in a second stage of the project based on the experience end-users gain while using the initial version. Due to its loosely coupled architecture both components can easily be configured to run without the other complementary component (*e.g*. through the open HSM API).

3 Future Work

Though the official project start is March 1, 1998, we have already made numerous design decisions reflecting basic ideas and implementation strategies. For increase of availability and performance it is planned to add another component to the HSM for caching purposes only. By the time of the conference we will show more detailed and fixed design and implementation strategies.

4 Conclusions

EuroStore will have all required features needed in a high performance and data intensive application in a small, scaleable, manageable size and complexity. By requiring minimal features from the host computer system, the use of modern object oriented languages and extensive use of standards, the system will enforce a smooth deployment in a heterogeneous computing environment.