Replacing DMF for Unicos

An Evaluation of Storage Management Products

Gerhard Rentschler
University of Stuttgart Computing Center
Allmandring 30
D 70550 Stuttgart
Germany
Email: g.rentschler@rus.uni-stuttgart.de

ABSTRACT: *This paper presents describes functions of several different products which could replace DMF 2.4 for Unicos. The products mentioned are DMF for IRIX, HPSS, SAM-FS and ADSM/HSM.*

KEYWORDS: Storage Management, DMF, HPSS, SAM-FS, ADSM, Data Migration

Disclaimer

This paper is not intended to provide a formal comparison between the mentioned products. It presents an evaluation of the functions under the aspect of usefulness for the University of Stuttgart only. We have done this evaluation to our best knowledge and based on the information available to us. University of Stuttgart including all members and employees deny any liability for any result of this paper.

Introduction

The University of Stuttgart Regional Computing Center (RUS) is currently running a Cray M92 with 50 GB DD61 Disks as a general purpose fileserver. DMF 2.4 for Unicos is used as migration software. This system is outdated and has to be replaced. RUS has evaluated HPSS, SAM-FS, ADSM and DMF 2.6.1 for IRIX as a replacement. Evaluation results and experiences will be presented in this paper. Emphasis is put on HPSS and DMF.

I do not intend to recommend a specific system. Which product is best depends on the specific needs and the financial situation of a site. I hope to be able to help in the decision by giving overview information about the products we investigated.

Furthermore, I will report on our experiences with DMF for IRIX.

Current Production Environment

The current production environment consists of a Cray M92 with about 50 GB DD61 disks. Unicos is release 9.0.2.0, DMF is release 2.4. For background storage we use a STK Nearline ACS 4400 tape robot with 12000 slots and 8 STK 4490 tape drives. At the moment the STK silo stores about 3.5 TB of data.

This configuration has some problems:

- High maintenance cost
- High defect risk because of age of hardware
- Disk cache to small compared to stored data
- Slow tape drives
- Low capacity of tape media (400 MB)
- Impossibility of upgrade for addressing increasing requests for data storage

Hardware Aspects

Tape Hardware

As the currently used STK 4490 tape drives are old, slow and don't meet today's capacity requirements we decided to install IBM 3590 tape drives into our STK silo. Installation was without problems. These tape drives meet our requirements. The provide approximately 3 times the throughput and 12 times the capacity of the STK drives. The only feature missing is automatic cleaning of the drives in the STK silo.

Test Systems

HPSS was tested on an IBM RS/6000 model 595 with 2 GB memory, a HIPPI adapter and 36 GB SSA disks. As tapes we used 4 IBM 3590.

SAM-FS was tested on a Sun Enterprise 450 with 2 IBM 3590 drives.

ADSM was tested on several systems. For backup and archival we used a IBM RS/6000 model 380 with 2 STK 4490 tapes. ADSM was tested on a IBM RS/6000 model 25T with a 8 GB disk.

DMF for IRIX was tested on a SGI Challenge DM with 2 processors and 512 MB main memory. The tapes were 2 IBM 3590.

HPSS

The High Performance Storage System (HPSS) was developed in a collaborative effort of major US national supercomputer laboratories and IBM Worldwide Government Industry, Houston, Texas. It is designed to meet the requirements of supercomputer centers. There is virtually no limit on number and size of files and storage capacity.

HPSS is a complete storage management solution. This fact distinguishes it from the other solutions mentioned in this paper.

The following features characterize HPSS.

Platforms

Server software is currently available for IBM RS/6000 running AIX 4.2. CERN in Geneva (Switzerland) is doing a port of the mover function to DEC Alpha. Parallel FTP client is available on a number of platforms. Please ask the vendor for information about specific platforms.

Network-Centered Architecture

HPSS is not just a single fileserver. There can be a large number of systems performing different functions and delivering services to other servers and clients distributed geographically and connected via a TCP/IP network. Towards users HPSS presents a single name space for data. Capacity and performance can be improved by just adding new servers to the network. The architecture of HPSS is based on the IEEE Mass Storage Systems reference Model, Version 5.

Third party data transfer

HPSS allows transfer of data directly from network attached storage devices (e.g. Maximum Strategy disk arrays) to the client using IPI 3 protocol. Other storage devices are attached to a mover system. In this case data is transferred between mover and client using TCP/IP via high speed networks like HIPPI. Multiple movers can be involved in the same data transfer. HPSS servers other than the movers are not involved with the transfer. Typically, HPSS uses different paths for commands and data.

Parallel operation

For striped devices, HPSS allows transfer of data in multiple streams simultaneously. For example, it is possible to use striping over several different movers and have a transfer with a different stream for each mover. This can be done with a special parallel FTP which is provided with HPSS.

Data integrity
HPSS uses Encina as transaction monitor to ensure data integrity in case of failures. For security services DCE is used.

Class of service
HPSS uses multiple storage hierarchies to allow cost effective storage of data according to user needs. Migration can use up to 5 levels including disks.

User access
HPSS allows access to data via the following interfaces.

- Standard FTP: using this interface it is not possible to get the advantages of HPSS
- NFS version 2: has the same disadvantages as FTP.
- Parallel FTP (PFTP): depending on the devices and the network involved this method can be very fast.
- Client API: also very fast. It allows programmers to access data in HPSS directly from the programs via library calls. The calls are modeled after POSIX.1 specifications. However, to use HPSS features existing programs have to be modified. Furthermore, the API client needs an Encina client on the same machine. For most current supercomputers no Encina client exists. In a future release of HPSS this requirement will no longer exist.
- DFS: this interface allows access to HPSS data via a DFS client. It is currently under development and will be available with the next release.

Management
Management of HPSS can be done via a GUI interface. This interface can be used from any system. It is not necessary to log on to servers to perform administration tasks.

Cost of HPSS
The financial requirements of HPSS are high There is a fixed price for the first year and a lower price for maintenance in the following years, regardless of the servers, clients and amount of data. In addition, significant expenses for hardware are necessary.

More information
More information about HPSS can be found at:

http://www.scsc.edu/HPSS

and

http://www5.clearlake.ibm.com:6001/

DMF for IRIX
After the merger of Cray and SGI the developers of DMF for Unicos ported the software to SGI's IRIX operating system. DMF for IRIX is called DMF 2.6 and is functionally equivalent to DMF 2.5 for Unicos. There are a few differences. DMF is not integrated in the system commands like ls or find. Instead, a different set of commands is provided, e.g. dmls, dmfind. Furthermore, the IRIX version uses the standard DMAPI system interface. DMF 2.5 on Unicos supports aggregate quotas. DMF 2.6 for IRIX does not care about quotas at all.

DMF is a hierarchical storage management system. It can be configured to manage a standard IRIX XFS filesystem. DMF is not a distributed storage concept. It runs on the fileserver. Access to the data over a network is possible via NFS. Limitations to number and size of files are those imposed by the XFS filesystem.

If the DMF server capacity is exhausted a second DMF server has to be set up. Multiple servers cannot be integrated, hardware devices cannot be shared. There are multiple different filespaces which must be managed.

Only the data part of the files is copied to background media. Metadata information is kept in the DMF databases and in the inodes. To achieve data integrity regular backups of the filesystem and the metadata information have do be done manually. According to the manual, DMF has to be stopped during backups. To avoid recall of the data to be backed up there is a special option on the xfsdump and xfsrestore command. The recycling of the tapes also has to be done manually.

Unfortunately, there is no tape management facility on IRIX like TMF for Unicos. This makes the manual backups a bit complicated, because it is not possible to request a tape mounted on the STK tape robot via IRIX command. Furthermore, Unicos has a feature called NFS UID-Mapping. There is nothing comparable on IRIX. DFS could help. But according to DMF documentation DMF is not supported on filesystems exported to DFS.

Pricing of DMF for IRIX is moderate. It is not dependent on hardware used or volume of data.

More information can be found at:

http://www.cray.com/products/software.storage/dmg/index.html

There is also a link for downloading the software for evaluation purposes.

SAM-FS

The Storage and Archive Manager (SAM-FS) was developed by Large Storage Configurations Inc. (LSC). It runs on Sun Solaris 2.4 or later. It is not a distributed storage concept. It runs on the fileserver. Access to the data over a network is possible via NFS.

If the SAM-FS server capacity is exhausted a second SAM-FS server has to be set up. Multiple servers cannot be integrated, hardware devices cannot be shared. There are multiple different filespaces which must be managed.

SAM-FS is a hierarchical storage manager. There is only one migration level. Up to 4 copies on background media are possible. There is no limit to the number and size of the files. SAM-FS uses a special type of Filesystem. Inodes are kept in a file .inodes which serves a the database for SAM-FS. The filesystem can be configured for striping.

Data on the background media is in standard Unix tar format. So it can be read without SAM-FS. It is possible to access the data directly from the background media.

Tape merging (or recycling as it is called in SAM-FS) is done automatically. For backup of metadata SAM-FS provides a special command.

The price for SAM-FS depends on the type of tape devices and the number of volumes in the library. For example, for STK Redwood or IBM Magstar drives SAM-FS is more expensive then for DLT drives. The cost for the software is high, but lower than HPSS.

More information can be found at:

http://www.lsci.com/lsci/products/samfs.htm

ADSM

Adstar Distributed Storage Manager (ADSM) is IBM's backup and archive system. For AIX, Solaris and Irix it also provides a HSM option. In contrast to DMF and SAM-FS ADSM has a server-client architecture. An HSM client can use an already existing ADSM server for migrating and restoring files.

HSM functionality of ADSM is comparable to DMF and SAM-FS. ADSM may be slower, especially when migration is done over the network. Furthermore, ADSM has a much more elaborate database. This may also cause a performance penalty.

ADSM/HSM uses the DMAPI standard interface. The limit to the number of files is 2^{64}, the limit to the size of a file is 2^{63}. There is no architectural limit to the storage capacity.

ADSM's big advantage is the integration of HSM and backup. Backup of a migrated file causes a copy operation of the file on the server only. Furthermore, these backups can be restored to any Unix filesystem on any machine, provided the access permissions are set accordingly.

Pricing for ADSM is rather moderate, if an ADSM server license already exists. In this case 1 license key for the HSM service for the server and a normal backup client license for each client is required. If the server license doesn't exist the price for ADSM will significantly increase.

More information can be found at:

http://www.storage.ibm.com/software/adsm/adsmhome.htm

Evaluation Results

RUS has to provide services to the supercomputing center and to the university users. Requirements of both groups are very different. So we decided to use different solutions in both areas.

The Supercomputing solution

In the supercomputing area we found that HPSS was able to meet most technical requirements. Performance is very good. We achieved transfer rates between the NEC SX4 and HPSS disk of 30 MB/sec. From the NEC to HPSS 2-way striped IBM 3590 tapes we got 15 MB/sec. A big advantage is the scalability of HPSS. An HPSS system can be expanded by adding new hardware (servers, disks, tapes) without interruption. There still is one single namespace for files. This is not possible with the other solutions.

Unfortunately, we have to make our budget to fit HPSS' requirements. In the meantime we need another solution. We decided to use DMF for IRIX for the following reasons:

- An SGI system with peripheral devices is available
- DMF know-how exists from Unicos
- Low price

For ADSM we buying an RS/6000 would have been necessary. Using the university ADSM service was not possible because of the large amount of data we would have to send over the network. SAM-FS software was much more expensive than DMF. Buying a large Sun system would be necessary. Functionally, SAM-FS does not have significant advantage over DMF.

The university solution

For our university users we currently offer backup and archival services based on ADSM. We are going to offer a HSM service based on ADSM/HSM. The filesystem managed by ADSM/HSM can be accessed by users via DFS. To avoid transfer of migration data over the network the disk hardware will be attached to the ADSM server.

DMF Experiences

In summer 1997 we started a beta test of DMF for IRIX. Results were quite satisfactorily. However, a few remarks are necessary.

In our opinion, DMF was declared generally available too early. When it was delivered to customers the following functions were missing:

- No –a option on xfsdump. Therefore, no backup of the migrated filesystem was possible because each file backed up got recalled from DMF. This function was not available before early 1998.
- No integrity management tools for DMF databases. Tools like dmaudit were delivered end of 1997.
- Tape merging still doesn't work satisfactorily. We have an open problem record.
- For doing the backups we need a tape management system like TMF urgently. Alternatively, a tpmount command which uses DMF's tape interface would help. Also we need the possibility set tapes offline and online in DMF.

We would like to see the following features in DMF:

- Backup integrated into DMF. DMF should backup everything it needs automatically.
- Automated tape merging
- Aggregate Quotas

Summary

HPSS is a complete storage management product designed for supercomputer environments. It's price may be to high for many budgets. SAM-FS and DMF for IRIX are excellent hierarchical storage management systems. The advantage of DMF is the quality proven in long years of usage on Unicos systems and the moderate price.

ADSM/HSM is a hierarchical storage managemement software integrated in a backup and archival system. If a backup and archival system is needed for workstations and PCs HSM can be added for a small fee.