Data Migration Facility Development Update

Thomas W. Lanzatella, Cray Research, Inc.

ABSTRACT: *This paper reviews storage management objectives and the DMF 3.0 program. A new and exciting initiative to provide DMF on SGI platforms is described. We conclude with a discussion of the current DMF releases and a product roadmap for DMF covering the next two years.*
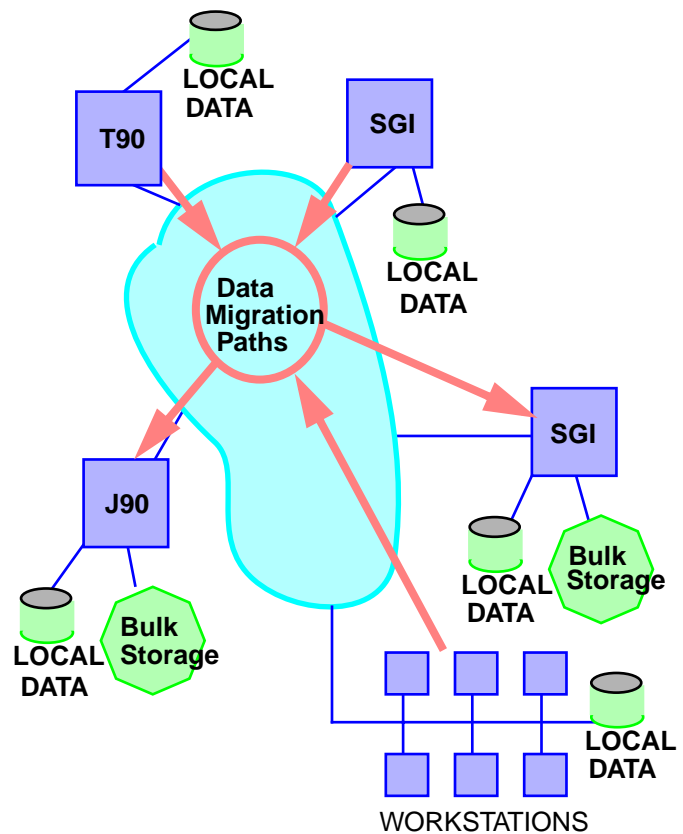
Introduction

Data Migration Facility is established as a reliable and efficient tool for storage management services in the Cray/UNICOS environment. Development objectives for storage management are driven by customer requirements. The evolution of the typical Cray environment to one of multiple-heterogeneous machines connected by high-speed networks require changes in DMF architecture. Distribution of computational and storage resources require that DMF be broken into distributable components. In this way, the critical functions of DMF, i.e., native file system migration and safe storage of data, can be accommodated across a wide area with consideration for the economic properties of storage devices.

Support for heterogeneous environments is a long-standing requirement for DMF. The acquisition of Cray by SGI fuels this requirement by making certain the likelihood of mixed configurations spanning a broad range of performance characteristics. The objectives of the DMF enhancement program are to provide customers with the ability to migrate from the platform of their choice and to store data on the platform or their choice.

Management of native file system capacity continues to be a critical requirement for DMF. This process relies on UNIX kernel interfaces that support file state changes allowing an agent such as DMF to gain control of the data associated with a path in the native file system. To date, only UNICOS has had an efficient, reliable interface for this purpose. In the last couple of years the Data Management Interface Group (DMIG) has promoted an interface called the Data Management API or DMAPI which standardizes an interface supporting native file system migration. The advent of DMAPI makes feasible the opportunity for DMF to manage file systems other than the UNICOS NC1 file system. In particular, since the IRIX file system, XFS, supports DMAPI, there is an immediate opportunity to manage XFS file systems with DMF.

The Migration Domain for Cray/SGI
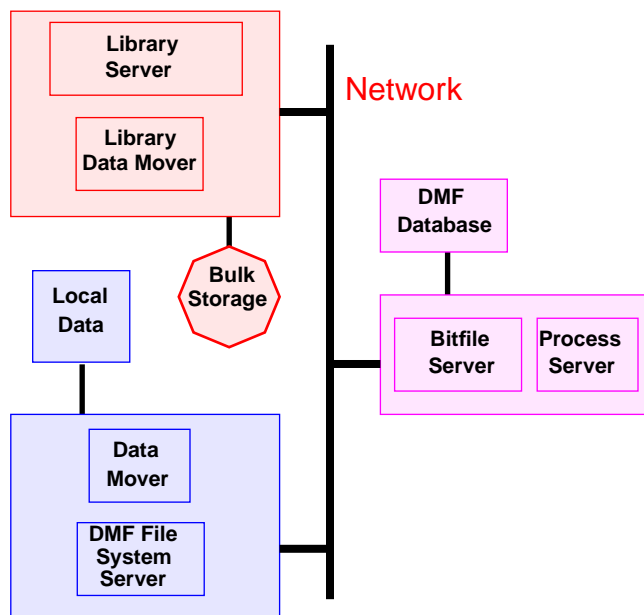
The DMF 3.0 Program

The current architecture of DMF is monolithic in the sense that all DMF components must run on the same machine. The DMF 3.0 program will transform DMF into a storage management tool having the same external characteristics as DMF today but its components will be separable to facilitate remote execution and monitoring.

During this process, we will adopt terminology from the IEEE Storage System Reference Model. The three major components of DMF 3.0 are the File System Server (replacing the DMF client), the DMF Bitfile Server (replacing the DMF server) and the Library Server (replacing the DMF media specific process).

This architecture is portable in the sense that all file system dependencies are localized in the File System Server while all media specific dependencies are localized in the Library Server. The Bitfile Server is considered to have no operating system dependencies.

To ensure that data can be moved from point to point with maximum efficiency, data movers will employ an authenticated, socket-based protocol. A data mover operates on behalf of each file system server and each library server. Multiple File System Servers and multiple Library Servers will be supported.

A single Bitfile Server will dispense bitfile identifiers (formerly called file handles) for the entire migration domain. And finally, the three databases employed by DMF today will be replace by a single database. We anticipate that DMF 3.0 will be available during 2H98.

DMF 3.0 Architecture

DMF for IRIX

The union of SGI and Cray has created a remarkable boost in demand for DMF. Demand is sufficient that we have initiated a port of the current DMF product to IRIX. The primary goal of this project is to offer a fully supported DMF product for the stand-alone IRIX environment. Interoperability with Cray machines running UNICOS/DMF is a secondary objective.

We anticipate availability of DMF for IRIX during 2Q97.

In the area of media support, SCSI versions of Redwood, Timberline and MagStar will be the first supported devices. In the area of robotics Grau, STK and IBM will be the first devices supported. IRIX supports many other manufacturers robots and transports. These will be supported as they are validated for use with DMF.

DMF 2.5 is the product that will be ported. DMF for IRIX will be named DMF 2.6. External differences between DMF 2.5 and DMF 2.6 will be kept to a minimum but there will be differences in areas affected by operating system dependencies. The following list details the major differences:

1) The kernel interface supporting file state transitions will be the DMIG/DMAPI interface.

2) The DMF Automatic Space Management utilities will be adapted to the sparse allocation capabilities of the XFS file system.

3) DMF 2.6 will employ a new SGI product named Open-Vault for tape mounting services instead of the UNICOS Tape Subsystem.

4) XFS dump/restore will be adapted to DMF-migrated files as on UNICOS.

5) The Fast-flexible I/O library (ffio) will be ported to IRIX.

6) IRIX does not have the UNICOS User Database (UDB) feature. DMF features that rely on the UDB will be removed until alternate implementations are found. These features are .keep processing (ARCHLIM) and processes mediated by the ARCHMED field.

7) The fsmon utility will be eliminated and its function will be combined with dmmctl.

8) The DMF tape MSP will be adapted to operate with character-special devices.

9) The dmmode command wil be removed in order to avoid making changes to the IRIX kernel.

10) In UNICOS, the commands ls, find and du were modified to know about migrated files. Rather than modify the IRIX versions of these commands, we will offer new DMF commands dmls, dmfind and dmdu with DMF 2.6

DMF 2.5 - New Database Technology

The primary purpose of the Cray Data Migration Facility (DMF) release 2.5 is to improve DMF database technology. The package used within DMF release 2.5 is the Raima Data Manager (RDM), produced by the Raima Corporation (see http://www.raima.com). RDM provides a database transaction commitment mechanism that protects the DMF databases from damage after an unscheduled interrupt by ensuring that only

whole transactions are applied to the databases. RDM also provides a transaction journaling capability. Transaction journaling is a method of tracking the history of database transactions. In the event of a disk failure or unscheduled interrupt, the system administrator can replay the transaction journals through a special utility which ensures that each DMF operation is reflected by the DMF databases.

DMF performance is not affected by the integration of RDM. Pilot studies show that performance of database operations in DMF release 2.5 is the same as or better than under DMF release 2.4.

Installing DMF release 2.5 requires that you convert your DMF databases with the new dmdbcvt utility. This is a one-way conversion process. Once you have moved to DMF release 2.5, you cannot convert back to an earlier version.

You should schedule a block of time during which you can leave DMF down for the conversion. The dmdbcvt utility runs at a rate of about 130 database entries per second on a CRAY Y-MP machine. Be sure to consider the sum of database entries in the daemon database and in each media-specific process (MSP) database when you calculate the expected DMF down time for conversion.

Prior to DMF release 2.5, the database package had a problem with inefficient reuse of index file space, which required the DMF administrator to compress and rebuild the DMF database on a regular basis. The new database package does not have this problem. Therefore, once you have converted to DMF release 2.5, you will no longer have to compress and rebuild your DMF databases. This improvement simplifies the Cray Data Migration Facility (DMF) Administrator's Guide, publication SG-2135 2.5.

Journal application is now uniform between the tape MSP databases and the DMF daemon database. A single utility is used to apply journals to all DMF databases. Journal application capabilities which used to be offered through dmvoladm and dmcatadm have been moved to the new dmdbrecover utility. This improvement simplifies the Cray Data Migration Facility (DMF) Administrator's Guide, publication SG-2135 2.5.

Support for the old tape MSP and for the station MSP ends with DMF release 2.5. When you convert your DMF databases to the new format, the conversion utility, dmdbcvt, will detect old tape MSP databases and convert them to the new format for the advanced tape MSP. You do not have to convert any of your DMF tapes. The advanced tape MSP reads old tape MSP media. Sites still running the station MSP should convert to the FTP MSP.

In DMF release 2.4 and earlier, the DMF administrator had the ability to change the name of database files from the released path names. In DMF release 2.5, the DMF administrator can change only the name of the directory in which databases are stored. Database names cannot be changed.

Source availability for DMF is ended with DMF release 2.5. For those few remaining DMF customers who have DMF source

licenses, special arrangements have been made to continue providing source.

DMF release 2.5 uses FLEXlm licensing to control access to the software. A license key is required in order to run the product. You will receive a license key from the Cray Research Distribution Center with installation instructions. If you do not receive the license key, contact the Distribution Center.

DMF 2.5 will be installed for its first field-test situation before the end of October. This test will be conducted on the USS file server machine within the Cray-Eagan complex. While a USS field test is a stringent test of DMF 2.5, it does not constitute a customer field test. We are presently searching for a customer interested in running a DMF 2.5 field test. We will not make DMF 2.5 formally available until after a customer field test is complete.

DMF 2.5 runs on UNICOS 9.0 and above.

DMF 2.4 Release News

In DMF 2.4, the dmaudit and dmatvfy utilities were modified to use the csort utility. By using csort, we were able to increase the processing speed of dmaudit by a factor of 4. Since that time we have experienced sufficient difficulties with csort to warrent a different approach to the problem of sorting binary files. In DMF 2.4.5, we have introduced a new sort utility, dmsort, which operates at the same speed as csort but has the advantage of being fully portable to the T3E and to IRIX platforms. All T90 and T90 IEEE sites should upgrade to DMF 2.4.5.

Two security-related problems have been repaired. All sites running. Trusted UNICOS should upgrade to DMF 2.4.5.

UNICOS Disk Quotas - Aggregate Limit Feature

The aggregate quota feature enables an installation to charge for offline files much the same as for online files. In the default (non-aggregate) quota case, the file's owner is not charged for file data which is offline. The total amount of disk space the user can have on line is constrained. An attempt to unmigrate a migrated file can fail because it would violate a quota limit. In the default case, the user can have an unlimited amount of data space in offline files.

With aggregate quotas, file unmigration is always allowed by the quota mechanism. We keep track of offline space and online data together. The user is limited to the total amount of space. This allows the administrator to limit the amount of space used in the offline archive - e.g. a tape silo. The tradeoff is that now the administrator has no quota control over what a user has online. A user who accidentally unmigrates more data that the file system can hold will cause a flury of activity.

A compile time aggregate quota mechanism has been implemented at several UNICOS sites based on code developed by our customer, KFA. The major difference between the released implementation and the implementation at KFA is that in the new implementation, aggregate quota enforcement is activitated by a directive to the quadmin utility.
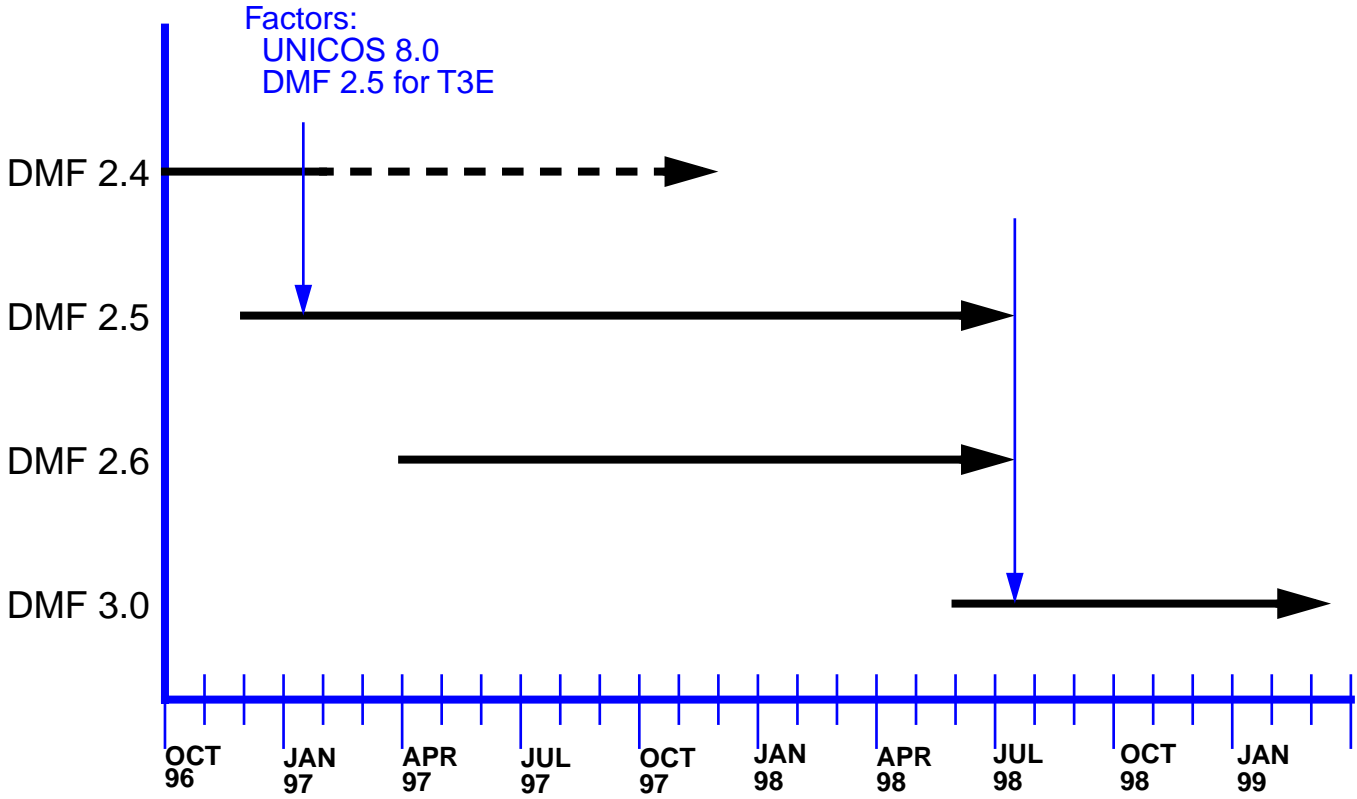
The Aggregate Disk Quota feature relies on UNICOS 9.0.2.3 and DMF 2.4.6. Both of these product should be available during 1Q97.

DMF Roadmap

The next couple of years will see several versions of DMF in use by a growing customer base. We see DMF 2.4 remaining in the field as long as there are machines running UNICOS 8.0. DMF 2.4.6 will be the first offering of DMF for the T3E series of machines. While it would be most prudent for all DMF customers to move to DMF 2.5 for the added safety afforded by the new database technology, this will not be immediately possible for T3E customers or for customers choosing not to upgrade to UNICOS 9.0. We expect DFM 2.5 to be available for T3E during 2Q97.

SGI customers will be able to run DMF 2.6 starting in 2Q97. DMF 2.5 and DMF 2.6 will be the primary products until DMF 3.0 arrives during 2H98.

Data Migration Facility - Product Roadmap