FILETEK STORAGE MACHINE APPLICATIONS

John G. Burgess
Nancy C. Ramsay

FileTek, Inc.
Rockville, Maryland

ABSTRACT

FileTek's Storage Machine is an automated mass storage system which utilizes optical disks stored in an optical disk library device. The Storage Machine is made up primarily of off the shelf hardware components and storage management software developed by FileTek and designed to be consistent with the MSS reference model.[1]

The Storage Machine is currently being utilized in both the commercial and government sectors by organizations that have a requirement to store, manage and provide large amounts of data in an efficient and cost effective manner. In an effort to illustrate the practical functionality of such a system, this paper describes some of the applications which have been implemented using the Storage Machine.

INTRODUCTION

IBM/MVS, DEC/VMS, and UNIX host systems are currently supported by the Storage Machine, with other host support planned. Host systems are connected to the Storage Machine via the Network Systems Corporation's HYPERchannel network, allowing concurrent access by homogeneous or heterogeneous host systems. This configuration offers the possibility of a central storage repository within an organization for either data or resource sharing.

The Storage Machine manages a transparent hierarchy of storage (see Figure 1). This allows those files that are frequently accessed to reside on magnetic disk which offers minimum response time, while files accessed less frequently are stored on an appropriate level of optical disk in the optical disk drive, in a slot in the optical disk library device, or on shelf storage. As the higher level of storage reaches saturation, data is automatically migrated through the other levels to free space. The Storage Machine software places each data file on the storage level most conducive to efficient access performance based on file retrieval statistics, activity and size, in addition to user parameters.

Data written to the Storage Machine can be retrieved as an entire file or at the record level. FileTek's Virtual Record Access Manager (VRAM) provides the capability to access data by relative record number or by key. In addition, records can be changed, deleted, and appended to a file. The record processing as with file level processing is transparently supported on all storage levels including write-once optical.

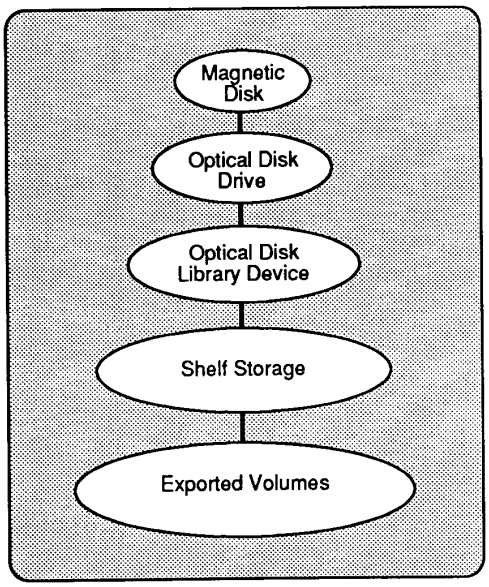

Figure 1. Levels of Storage.

The system provides allocation control to enable the storage of a set of files on a volume surface or sets of surfaces/volumes. Sets of volumes may be moved, exported or imported.

Other features include security, automatic recovery, preemptive priority processing, version control, user data backup and multiple copies. The Storage Machine system is described more fully in "Virtual Library System: A General Mass Storage Archive."[2]

The current platform for the Storage Machine is either a single or multiple MicroVAX(es). Future plans include a

 CH2626-0/88/0000/0096/$01.00 © 1988IEEE

UNIX platform.

STORAGE MACHINE APPLICATIONS

National Aeronautics and Space Administration (NASA)

The Missions Operations and Data Systems Directorate (MO&DSD) of NASA at the Goddard Space Flight Center (GSFC) is responsible for the storage of satellite telemetry data which is used for scientific research (see Figure 2). The data is transmitted via satellite to GSFC ground stations and then over a network to computer systems designed to process the raw data, perform data reduction, and generate data for use by researchers. The anticipated increase in data generated by satellites scheduled to be launched in the next few years,[3] combined with the continuing requirement for increased quality of data, dictate the need for a more flexible and efficient method of storage and access.

NASA has installed the Storage Machine to be used as a repository for the management of a portion of their satellite telemetry data. Users currently gain access to the Storage Machine through an IBM 4341 and a VAX 11/785.

Data from The Cosmic Background Explorer (COBE), scheduled to be launched in the Spring of 1989, will be stored on the Storage Machine. COBE's purpose is to study the "Big Bang," the primeval explosion that started the expansion of the Universe 15 billion years ago. It will carry three instruments to study the diffuse cosmic background radiation from this explosion and from more recent events. The Tracking and Data Relay Satellite will be used to monitor the satellite and to send commands. Raw data, processed data, and orbit ephemeris data will be sent to the Storage Machine through the existing HYPERchannel network (InfoLAN) at GSFC. Several versions of each telemetry data set are expected to be kept on the Storage Machine.

The initial storage requirement for on-line retrieval of data is in excess of 40 gigabytes of data. Access to the data will be available from the VAX 11/785 and the Unisys 1100/82 through a NASA-developed menu driven routine which prompts the users for data about the specific file they wish to access. The file is retrieved from the Storage Machine and returned to the requesting host machine. Users can then proceed to run experiments on the most advanced version of their data and/or return to a previous version on the Storage Machine for changes in data reduction processing. The long shelf life of the optical medium (now projected to be in excess of 40 years) is very beneficial to preserving this valuable information.

The Storage Machine solution provides efficient access to large volumes of scientific data from the different host systems at GSFC. On-line, multi-user access provides the data processing efficiency required by NASA experimenters, while eliminating the costly and time consuming process of reformatting large volumes of data from magnetic tape.

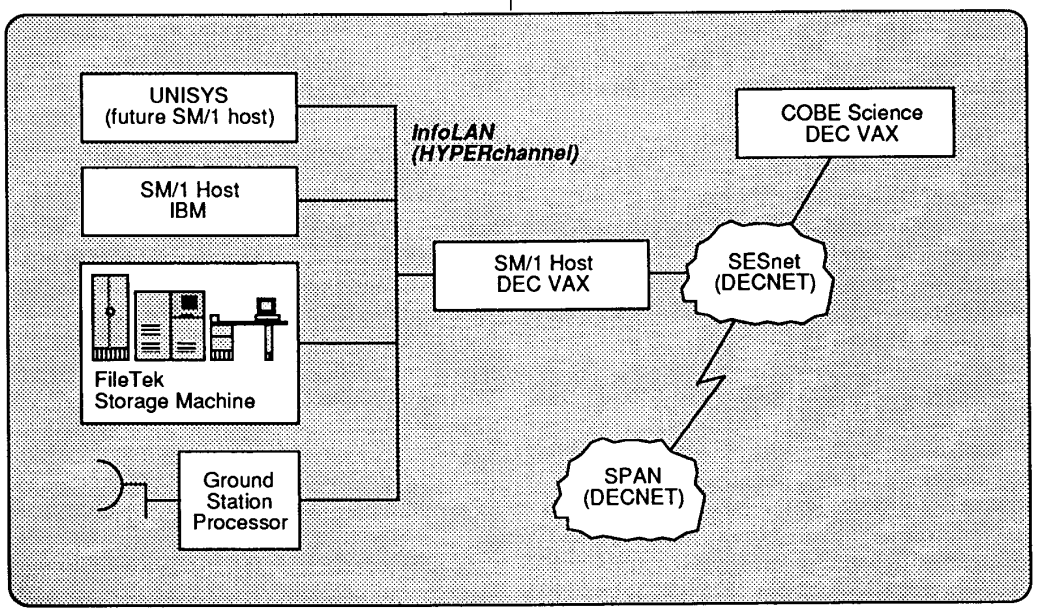

Figure 2. NASA Goddard Space Flight Center Configuration.

Planned applications for the Storage Machine at NASA include the storage of data for the International Solar Terrestrial Physics Program (ISTP) and the Upper Atmosphere Research Satellite (UARS).

After an extensive 30-day test, the Storage Machine was accepted for production operation by NASA/GSFC's Information Processing Division.

United Services Automobile Association (USAA)

United Services Automobile Association, based in San Antonio, Texas, insures over 1.7 million military officers and their dependents. They are the sixth largest insurer of personal automobiles, the seventh largest home owners insurer in America, and the largest direct response life insurer in the world.

There are approximately 1,200 Customer Service Representatives (CSRs) employed at USAA to answer questions regarding customers' past insurance statements. The CSRs need to have access to 7 months of data on-line and 7 years of data are required to be saved on microfiche. USAA's customer statements for a single month consist of 1.5 gigabytes of data. The large volume of data makes it cost prohibitive to retain more than 6 months worth of statements on DASD.

Prior to the installation of the Storage Machine, statements older than 6 months were written to tape, processed into microfiche to be stored in cabinets, and finally, deleted from the on-line IMS database. Customer inquiries regarding statements older than 6 months would require the CSR to send a request to the microfiche clerk to retrieve the information and forward a hard copy of the statement back to the Customer Service Department. As a result of the time delay associated with the retrieval of microfiche records and forwarding of the paper copy of the statement, the inquiries could not be answered during the time of the original phone call. As a result, the CSR was then responsible for contacting the customer by telephone to service the initial request. Additional time was added to the process if the CSR was unable to reach the customer by phone, or if there were further questions regarding another month's statement.

The Customer Accounting System (for customer statement inquiries) resides on the IBM 3090 host computer at USAA. The Storage Machine is connected to the host computer through the HYPERchannel network. Customer statement data over two months old is written to tape. The data on tape is then loaded to the Storage Machine via batch jobs and subsequently deleted from the IMS database. When the CSR requests a statement, the application program attempts to retrieve the data from DASD. If the data is not resident on DASD, the CSR is notified that a search will be made on optical. The application program places a call through the FileTek Callable Interface to connect with the Storage Machine, locate the statement record based on account number and statement month using the VRAM software, and returns the record for insertion into the IMS database. The record is then available for query by the CSR and is deleted from the database at the end of the day.

Customer inquiries using Storage Machine can now be answered in seconds eliminating the need to call the customer back and enabling the CSR to improve productivity. USAA has been able to decrease the amount of data kept on DASD from 6 months to 2 months, with the remaining data on the Storage Machine. Costs related to the labor-intensive nature of microfiche retrieval have been significantly reduced. The statement data remains in computer readable format retaining the ability to perform market research and statistical analysis on the data. The direct accessibility of optical disks, and the systems support for returning only the needed record, enable obtaining the less than 20-second response time required by this application.

Under consideration for future Storage Machine applications are closed claim processing, the storage of acturarial data, and on-line documentation.

AT&T Network Systems

AT&T has developed an integrated, computer-based information system called the Electronic Document Distribution System (EDDS) for the purpose of solving information management problems in the areas of engineering and management. EDDS combines into one system the capabilities for text and graphics publishing, engineering drawing management, and mapping/outside plant applications.

The EDDS allows AT&T to provide paper, film, CAD, and word processing-based documentation in a uniform electronic format for use by internal users as well as large customers like the Bell Operating Customers.

Engineering drawings can be scanned and stored in raster or vector format for edit, retrieval, distribution and printing. Input devices supported include aperture card scanners, A through E size paper scanners, OCR scanners, microfiche and direct input from CAD-based systems. The management of the drawings is supported through the use of high-performance workstations.

Text and graphics publishing capabilities allow the integration of graphics and text on a single page. Output is

defined using PostScript, enabling the use of a wide array of output devices to produce customized documents.

AT&T has elected to use the Storage Machine for EDDS to serve as the central repository for a wide range of information including both record and file level accessing from a variety of host platforms. AT&T has installed multiple Storage Machine systems to support this application.

The Depository Trust Company (DTC)

The Depository Trust Company is a cooperative owned by the financial industry and serves as a national clearinghouse for the settlement of trades and corporate municipal securities. DTC also performs security custodian services for its participating banks and brokerage dealers. "The value of securities on deposit is valued at almost three trillion dollars and the deliveries processed during 1987 had a total value of almost ten trillion dollars. DTC's data processing environment includes three large IBM mainframes, an intermediate sized one as a shared output spool, over 600 gigabytes of DASD..."[4]

One major function requires DTC to maintain a reconciliation and research department responsible for answering participant inquiries, daily stock record control, fast account balancing, and error research for lost certificates. Approximately 150 employees in the reconciliation department receive inquiries and answer requests. Researchers may spend several minutes to several days researching a problem.

DTC utilizes IBM's Report Management and Distribution System (RMDS) as a vehicle for the retrieval, access, viewing and distribution of reports from their IBM host computer. The large size of these reports necessitates their migration from DASD to microfiche after several days. Each researcher keeps approximately 1 to 6 months of various microfiche reports (Daily Stock Record, Customer Account Statements, etc.) at his/her desk. A master copy of all fiche reports is kept in a filing cabinet in each area for 6 months after which it is sent to a warehouse. Inquiries for data older than 6 months require that microfiche be retrieved from the warehouse, extending the research time considerably.

To improve response to customer requests, DTC has chosen to write a Storage Machine software interface to IBM's RMDS. The Daily Stock Record report (approximately 120 megabytes per day) was selected as the first candidate for the RMDS/Storage Machine software due to its large size.

Nightly batch jobs scan the RMDS Directory to read user modifiable fields which hold information regarding the report status. If the report version is old enough to be moved to the Storage Machine, and does not already reside there, it is prepared for transfer. The Daily Stock Record report is divided into 10 segments, based on a CUSIP number for efficient retrieval. Each segment is stored as a separate file on the Storage Machine.

End users retrieve reports using a TSO/ISPF Dialog Management Application written by DTC. This is invoked from an RMDS exit when the requested report is not found on-line. The user choses a report type from the menu, and then specifies the time span needed for his/her research. A listing of all reports that meet this criteria appears on the screen. After the user selects the desired reports, the application verifies that those file are not already resident on the host machine. If the report is to be retrieved from the Storage Machine, the user is notified that the report is being copied back to RMDS and his/her terminal is freed for additional processing. Once the report has been processed by RMDS, the user may display and view the report within RMDS using standard utilities and functions.

By utilizing the capabilities of the Storage Machine, all researchers have access to all reports on-line. Duplication and distribution of fiche and printed reports is eliminated. The time required to respond to customer questions and problems has been significantly reduced.

United States Government

This organization is an element of the Department of Defense and has worldwide responsibilities that include operations, maintenance and financial management functions.

One of the functions of this agency is to retain large amounts of historical data. The data is archived, available for user access on various media such as tape, COM, and paper reports. Depending on the application, the retention period for the data varies from three to seven years. Although the application specifics can not be divulged, some of the areas well suited for the Storage Machine are:

- The contract audit analysis report which requires the passage of 15 to 20 tapes to extract a few records for the report.
- Contract data and financial history data files which must be retained for 7 years and are retrieved frequently.
- Payment ledger report of several hundred pages with multiple copies printed for reference by users who are typically searching for only one record of

information.

- Vendor and inventory master file lists which are currently sent off site for COM production.

The installation of the Storage Machine at this site will reduce the amount of paper reports produced, significantly decrease the time necessary to retrieve information previously on tape from a few days to less than one-half hour, increase user productivity, and create a central repository for the agency's data.

CONCLUSION

The development of applications to use Optical Technology has been slow in coming. The Storage Machine architecture has proven, in these and a number of other applications, to be an effective enabler for the use of this technology. FileTek fully supports the Storage Machine with system engineering, training, documentation and maintenance support.

In addition, FileTek actively supports the IEEE Mass Storage Reference Model and where appropriate, the Storage Machine software can be used as building blocks in integrated system developments.

TRADEMARKS

IBM, MVS, RMDS, IMS are registered trademarks of International Business Machines Corporation; DEC, VMS are registered trademarks of Digital Equipment Corporation; HYPERchannel is a registered trademarks of Network Systems Corporation; Storage Machine is a registered trademark of FileTek, Inc.

REFERENCES

1. Miller, S.W., and M.W. Collins, "Toward a Reference Model of Mass Storage Systems," Digest of Papers, Seventh IEEE Symposium on Mass Storage Systems. p.3, 1985.

2. Burgess, J., "Virtual Library System: A General Purpose Mass Storage Archive," Digest of Papers, Eighth IEEE Symposium on Mass Storage Systems. p.72, 1987.

3. Kempster, L.S., and John B. Martin, "In Search of: NASA Space Data Storage Solutions", Digest of Papers, Eighth IEEE Symposium on Mass Storage Systems. p.27, 1987.

4. Ackerman, E., "Optical Storage in the Securities Industry," Yankee Group Seminar, June 1988.