LBNL-49327

A Beta Test of Linear Tape-Open (LTO) Ultrium Data Storage Technology

Shreyas Cholia* and Nancy Meyer

National Energy Research Scientific Computing Center (NERSC)
Ernest Orlando Lawrence Berkeley National Laboratory
Berkeley, CA 94720

*S Cholia is also at
IBM Global Services – Federal
Houston, TX 77058

December 2001

Abstract. NERSC is participating in several HPSS (High Performance Storage System) research and development projects as part of the Probe testbed. One of these projects involved beta testing of the IBM 3584 UltraScalable Tape Library, which uses the new ultra-high-density Linear Tape-Open (LTO) Ultrium tape drives. Ultrium tape cartridges have a capacity of up to 300 GB of compressed data, greatly reducing the number of cartridges needed to store massive scientific datasets. NERSC's preliminary performance testing indicates that LTO Ultrium technology, with compatible products and media available from several vendors, may be a viable alternative for computer centers seeking higher-density archival storage media with a small footprint and relatively low cost per drive.

This work was supported by the Director, Office of Science, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy under contract number DE-AC03-76SF00098.

Introduction: The Probe Testbed for HPSS Storage

Probe is an HPSS (High Performance Storage System) research and development project undertaken jointly by the Oak Ridge National Laboratory's (ORNL's) Computer Science and Mathematics Division and by the National Energy Research Scientific Computing Center (NERSC) at the Lawrence Berkeley National Laboratory. The collaborative effort of the two national laboratories provides a wide-area distributed-storage testbed for HPSS research.

Probe allows testing of new storage hardware and software for HPSS systems. Hardware devices such as servers, network interfaces, switches, and storage devices can be tested using Probe, and new software including protocols and database technologies can be developed using the Probe testbed.

The Probe configuration at NERSC as of February 2001 is represented in Figure 1. The box "Tape Robot Beta Test" represents the IBM 3584 UltraScalable Tape Library.

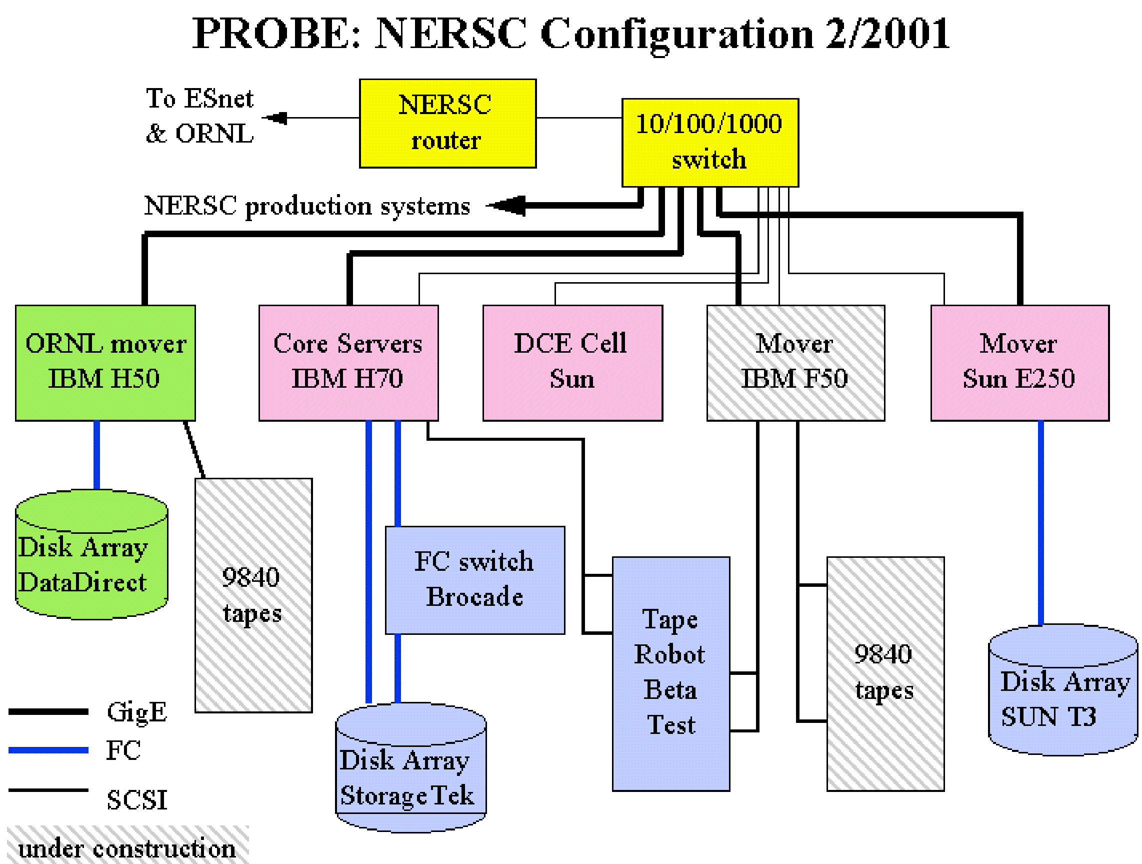

Figure 1. Probe configuration at NERSC.

Background and Specifications

Linear Tape-Open (LTO) technology was developed jointly by Hewlett-Packard, IBM, and Seagate as an open-standards, licensed technology, which means that users will have multiple sources of compatible products and media [1]. The Ultrium format is the single-reel, high-capacity implementation of LTO technology. The dual-reel LTO Accelis format provides faster access than Ultrium, but at the cost of lower capacity.

NERSC tested the IBM 3584 UltraScalable Tape Library, a standalone device that uses LTO Ultrium tape drives for automated tape handing and storage. The library's scalability allows one to increase capacity by adding up to five additional storage units (called *expansion frames*) to the *base frame*. The base frame supports up to 12 Ultrium tape drives, and contains 281 cartridge storage slots and a 10-slot input/output (I/O) station for moving cartridges to and from the library. Cartridge capacity ranges from 100 to 300 GB, depending on data compression. Each expansion frame has up to 440 cartridge storage slots and supports up to 12 LTO tape drives. Each frame (base and expansion) has a footprint of about 29 by 60 inches and a height of 71 inches. The fully configured, six-frame library supports up to 72 drives and 2,481 storage slots, with a total capacity of 248.1 TB of uncompressed data or 496.2 TB with 2:1 compression. The beta test system installed at NERSC in June 2000 consists of the base frame with four LTO tape drives.

Each Ultrium tape drive has the following performance characteristics, according to IBM documentation [2]:

- 15 MB/sec native sustained data transfer rate
- 30 MB/sec sustained data transfer rate at 2:1 compression (the drives have built-in data-compression capability)
- 80 MB/sec burst data transfer rate for Ultra2, low voltage differential (LVD) SCSI drives
- 40 MB/sec burst data transfer rate for Ultra SCSI high voltage differential (HVD) drives
- 20 sec nominal load-to-ready time
- 18 sec nominal unload time
- 95 sec average search time to first byte of data.

These performance characteristics came with the following clarification:

> [T]he actual throughput is a function of many components, such as the host system processor, disk data rate, block size, data compression ratio, SCSI bus capabilities, and system or application software. Although the 3584 UltraScalable Tape Library is capable of up to a 7.8 TB/hour sustained data transfer rate (at 2:1 compression), other components of the system may limit the actual effective data rate. [2]

The IBM 3584 features storage area network (SAN)-ready multi-path architecture, which allows homogeneous or heterogeneous open systems hosts to share the library's robotics without middleware or a dedicated server acting as a library manager. The library is controlled by Small Computer Systems Interface (SCSI) commands issued to the drives. The library's storage slots and tape drives can be partitioned into as many as 72 logical libraries. Hosts can then run separate applications in each logical library. Or, instead of partitioning the system into logical libraries that communicate independently with separate hosts via separate control paths, any single logical library (or the entire physical library) can be configured to be shared by two or more hosts.

A silicon chip in each tape cartridge remembers where the last write was in order to relocate it quickly. The chip also keeps statistics on the tape to enhance reliability.

HPSS Code Development and Configuration

As part of the NERSC LTO Ultrium beta test and the HPSS 4.3 development effort, additional HPSS code was developed to support the LTO technology. The new code was developed and implemented by Shreyas Cholia (IBM/NERSC), Benny Wilbanks (IBM), and Deryl Steinert (ORNL). The additional code involved:

- A new HPSS physical volume repository (PVR) that manages the IBM 3584 tape library and is responsible for tape mounts, dismounts, imports, and exports
- Modifications to the HPSS mover to support the IBM 3580 drives
- Miscellaneous changes to HPSS in order to recognize a new drive and library type.

When the code was completed, the IBM 3584 library and drives were configured on the Probe HPSS system. The HPSS LTO PVR talks to the LTO library using the SCSI medium changer interface in the AIX tape driver. Communications occur directly over a SCSI control path, thus the library must be connected to the same node that runs the PVR. The PVR talks to the System Master Catalog (SMC) device special file for the LTO library. The PVR operates synchronously: a request in the PVR must wait for a previous request to return from the library before it can be issued. LTO drives must be configured in HPSS such that their drive addresses match their SCSI locations in the LTO library. These drives are controlled by a SCSI mover. The Probe LTO configuration was set up like this:

- HPSS core server node running HPSS 4.1.1 with special modifications for LTO
- HPSS LTO PVR controlling the library through /dev/smc0
- HPSS LTO tape mover controlling four LTO drives: /dev/rmt0 – /dev/rmt3
- A tape driver interface (level 6.0.4.0) to talk to the IBM 3584 robot and the 3580 drives
- Dual-channel SCSI Ultra2 LVD card with channel 1 connected to drives 1/2 + LTO library and channel 2 connected to drives 3/4. The drives are daisy-chained together on a single channel, and the library SCSI control path is shared with the SCSI drive data path across channel 1.

Test Results

Library Test Results

NERSC tested the performance if the IBM 3584 Library robotics, and found a mount time of 25–30 sec and an unmount time of 27–32 sec. Mount time includes loading the cartridge from its slot in the frame to the drive, and getting the cartridge ready for the first read or write. Using a Perl script that makes calls to tapeutil, we continuously issued mount/dismount commands. (Note that these were AIX, not HPSS mounts/dismounts.) The test randomly picked a tape and mounted it in a drive, then dismounted it. As soon as a mount or dismount was completed, the next one was issued.

With the library under continuous mount/dismount stress for three days, there were no failures. The test continued from 6:30 p.m. on Friday, August 25, 2000 to 1:59 p.m. the following Monday, with a total of 8,980 mounts and 8,978 dismounts, or 133 mounts per hour.

Library inventory time was tested at < 60 sec per frame. The time was measured by using the date command in AIX like a stopwatch just before and after inventory.

Drive Test Results

Testing of the write and read performance of the LTO Ultrium drives in HPSS was done using three methods:

- Using the HPSS scrub tool, which allows you to create/read/write files directly in HPSS, as well as to specify the data format (compressed/uncompressed) and the buffer size being sent to the drive.
- Migrating and staging files between an HPSS disk cache and the LTO tapes. Files already in HPSS were either migrated from disk to tape or staged from tape to disk. The read/write times were measured by examining the log files generated by migration and staging.
- AIX tar tests were conducted independent of the HPSS tests.

The HPSS storage class was configured as follows:

- Media block size = 256K
- Blocks between tapemarks = 3796 (I/O block size = 971 MB)
- VV block size = 1 MB

Read and write performance results on a 1 GB file are shown in Table 1. Our testing found that the LTO technology works best when the data is kept streaming. If we had to perform any operations that involved stopping and starting the tape heads, there was a severe performance degradation. We achieved streaming performance by the following methods:

- Use very large values for blocks between tapemarks — approximately 1 GB. The writing of a tapemark causes a start/stop on the tape, which significantly reduces the read/write speeds.
- Use large buffer sizes when writing to tape. In the scrub tool, we used 100 MB buffers to ensure streaming.
- Use large files for best performance (over 100 MB).

Table 1
LTO Ultrium Tape Drive Write and Read Performance

	Write Performance	Read Performance
AIX tar		
Compressible data	14.5 MB/sec	25.0 MB/sec
Noncompressible data	9.0 MB/sec	14.0 MB/sec
HPSS scrub		
Compressible data	24.6 MB/sec	30.4 MB/sec
Noncompressible data	10.2 MB/sec	12.4 MB/sec
HPSS migrate/stage		
Compressible data	24.6 MB/sec	34.5 MB/sec
Noncompressible data	10.2 MB/sec	13.8 MB/sec

A test of seek times using the scrub utility produced the results shown in Table 2. The seek time was obtained by subtracting the known write time from the total time of the seek/write process.

Table 2
LTO Ultrium Tape Drive Seek Performance

Write append to EOT (seek over 17 GB of data before writing)	141 MB/sec
Forward seek (fsf command over a 100 MB file)	15 MB/sec
Backward seek (bsf command over a 100 MB file)	8 MB/sec

Discussion and Conclusions

As scientific applications produce increasingly massive amounts of data, scientific computing centers must consider deploying denser storage media. With a capacity of 40 GB per cartridge (a size commonly used today), it would take 25,000 cartridges to store 1 PB of data, while it would take only 3,333 LTO Ultrium cartridges. With a transfer rate in the range of 30 MB/sec, the IBM 3584 library could transfer 1 PB of data in a week using all 72 drives.

Beta testing of the IBM 3584 library at NERSC indicates that the LTO Ultrium technology may be a viable option for computing centers seeking higher density storage media. The cost per tape drive is about a third the cost of many widely used drives today, and the library has a small footprint.

LTO Ultrium technology seems most appropriate for storing many large files at one time. Drive performance is good for large amounts of streaming data, but the start/stop on the tape significantly reduces the read/write speeds. LTO Ultrium technology is not recommended as an online storage mechanism but rather as an archival or backup storage system that can achieve optimal read/write conditions.

Acknowledgements

This work was supported by the Director, Office of Science, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy under contract number DE-AC03-76SF00098.

---

[1] http://www.lto-technology.com/
[2] *IBM 3584 UltraScalable Tape Library Planning and Operator Guide,* draft version, May 2000.