

CHALLENGE:

As one of the leading high-performance computing (HPC) sites in the world, the NASA Ames Research Center generates approximately 1.5 PB a month of data related to research and simulations. Its existing data storage solution was nearing end of life and needed to be replaced. The center needed a solution that allowed for fast tape mount times and easy access, was cost effective, and provided a way for them to easily manage the lifecycle of the media.

SOLUTION:

File-based active archive to tape:

- Six (6) Spectra Logic T950 enterprise tape libraries
- Integrated with SGI automatic Data Migration Facility (DMF) and SGI InfiniteStorage high-performance RAID arrays

RESULTS/BENEFITS:

Cost Effective:

- Ames doesn't have to store everything on spinning media or disk drives. Tape costs less in terms of storage per GB and operational costs.

Durable:

- Life expectancy of tape media is much higher than disk drives. Ames can keep data forever due largely to the increasing stability of the media.

Energy Efficient:

- Ames needs to have drives running 24/7; tape drives use less energy year-over-year than disk drives.

CASE STUDY

NASA Ames Research Center

NASA Ames Selects Tape-based Active Archive To Manage and Store High-Volume Data

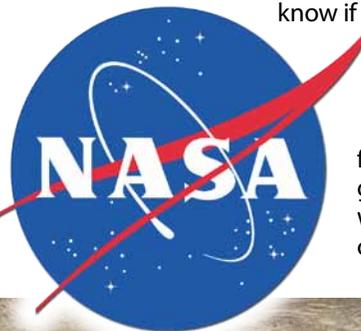
The NASA Ames Research Center (Ames), one of 10 NASA field centers, is located in the heart of California's Silicon Valley. For more than 60 years, Ames has led NASA in conducting world-class research and development. With 2,500 employees and an annual budget of \$900 million, Ames provides NASA with advancements in entry, descent and landing technologies; information technology; next-generation aviation improvements; astrobiology; airborne sciences; and small-satellite programs. Ames has a three mission directorate:

- 1) Science
- 2) Aeronautics
- 3) Human Exploration

Data generated from research in these areas, such as the results of simulations and other projects, is retained

indefinitely so the storage of this data must be robust, reliable and stable. Ames' previous storage solution was more than 20 years old and was coming to the end of its lifespan. Since data is recalled frequently, fast recall time is essential to enable researchers to access data in a time efficient manner. This goal put them on a mission to find a cost-effective data storage solution that allowed for quick tape mount times.

It was also important for the new solution to have the capability to manage the lifecycle of media so Ames' would know if tapes were going bad. This allows the proactive ability to rewrite data to a different tape. Lastly, given the current fiscal environment in the government sector, Ames was very sensitive to the cost of implementing a



Shown at left: NASA Ames testing of the K9 Rover in a new "Marscape" for future Martian exploration.

CASE STUDY Active Archive Alliance



This NASA Ames unmanned aircraft was used in the fight against California's wildfires.

new solution and was looking for something that would be cost effective, yet reliable and scalable.

Ames implemented a new active archive solution using Spectra Logic's enterprise tape libraries to replace its existing solution. The tape library was integrated with the SGI DMF tiered storage virtualization system. With DMF, all storage, whether disk or tape, is perceived by the users and applications as one very large storage pool. All data, whether on tape or disk, is always seen as online, all the time.

As users access data, the system transparently caches it on disk for fast response time. After use, DMF data management policies determine whether that data should remain on spinning disk, or be migrated into the tape tier. The policies can be driven by frequency of access, or by other criteria determined by IT administrators. By tiering data from fast drives to slow drives to tape, Ames can more effectively manage the data and make it accessible as needed.

File-based Active Archive

Active archive software technologies allow existing file systems to expand over disk platforms, tape libraries and other storage technologies. While Ames was effectively using an active archive solution prior, the Spectra tape libraries, which utilize media lifecycle management, afforded them much more flexibility. The

file-based active archive solution is structured using SGI DMF, which creates and automatically manages a tiered virtual storage environment that can significantly reduce equipment and operating costs, improve service levels and lower risks. DMF's metadata server and integrated data movers provide control over the entire system for both policy management and data integrity. This is a typical configuration for most active archives.

Ames has an average of 1,000 users saving and accessing data on a regular basis. Due to the critical nature of its research,

all data is stored forever until the users decide to delete it. Approximately 1 PB of Ames data is cached on disk at any given time. However, this data is only temporarily stored on disk arrays. When the system reaches 80 percent capacity, data that has not been recently accessed is automatically migrated to lower performing tiers or to tape. This migration requires no user interaction. All files still appear to users and applications exactly where they were, no matter where the system may place the files according to data management policies.

"The active archive solution allows us to reduce cost through the use of cost-effective tape media for long term data retention in place of disk drives, while maintaining reliability plus the ability to easily retrieve data. Tape is energy efficient resulting in overall energy savings," said Davin Chan, HPC Technical Director for CSC supporting NASA Ames Research Center.



Data Accessibility and Security

Ames was able to utilize Spectra Logic's Media Lifecycle Management (MLM) to reduce media-related issues through its intuitive reporting of at-risk media that should be retired. The MLM reporting is accessible directly from the library or via remote web access, so Ames can easily identify whether a tape is safe to store data upon by simply checking its health score, which is calculated from its health statistics. Thus, Ames is able to rest assured that its data is safely stored, thus reducing the possibility of data loss.

"The active archive provides our researchers with fast, online access to long-term data and enables virtually unlimited scalability so we can easily expand the storage capacity as our data storage needs grow."

Poised for Growth

Ames has 50 PB of data stored on its system using six tape libraries – three in a primary facility and three in a secondary facility. Ames writes two copies of the data for redundancy – one primary (25 PB), and one secondary (25 PB) – to tape media for a total of 50 PB of storage. The rate of data is growing at approximately 1.5 PB of data per month or .75 PB per month of unique data sets. With a maximum capacity of 115 PB of storage with compression, the Ames active archive is well positioned to accommodate future growth.

Future Plans

While the current active archive can accommodate Ames' needs for quite a while, they are evaluating new technologies such as Spectra Logic's T-Finity tape libraries and LTO-6 media. This will greatly increase the capacity of the tape library solution and allow Ames to continue normal operations for the long term.