

A New Approach to Creating and Managing Large-scale Digital Archives

Sun StorageTek™ 5800 System with Fedora



Rapid growth of digital content is driving the need for a new approach to preserving and managing archived data. With vast stores of data that must be preserved for many decades to come, many organizations are finding it costly to maintain the integrity of their data and are challenged by the complexity of storing content in a way that enables users to easily find and retrieve what they need. Sun has partnered with Fedora Commons to provide a solution that can help organizations store, manipulate, and access their digital resources in a flexible and cost-effective manner. The solution combines the Fedora Commons open source software platform with the Sun StorageTek™ 5800 system to provide a petabyte-scale object store that greatly simplifies the task of preserving massive amounts of data over long periods of time.

Highlights

- Integrated solution with proven success for large-scale repositories at well known institutions
- Link millions of digital objects in Fedora with capacity to seamlessly scale to hundreds of terabytes with the Sun StorageTek™ 5800 system
- High data integrity through automatic block and file level checksums and distribution of file fragments across nodes to ensure no single point of failure
- Extreme data protection with redundant design that allows multiple nodes and drives to fail without any resultant data loss
- Reduced cost and complexity of managing and administering digital archives

Overloaded with digital data

Organizations such as research institutions, libraries, professional societies, and publishing groups have long been tasked with maintaining large volumes of printed information for ongoing research or as historical artifacts. The role of these organizations is changing, however, as people are creating and managing digital content on a scale never before imagined. Libraries that were once the domain of stacks of books and journals are now handling a vast and rapidly growing array of digital objects such as ebooks, research data sets, digital images, and audio and video recordings. Many research organizations utilize High Performance Computing (HPC) applications that commonly generate data sets that fill multiple terabytes of storage. Both the quantity and size of digital objects are increasing rapidly, so it's easy to see how today's archives can quickly grow to petabyte scale.

For many types of digital data, it is necessary to preserve both the information content and its authenticity over a period of several decades or even multiple centuries. The data must survive transitions in staff or user communities as well as replacement of hardware or software systems that contain the data.

Preserving digital content requires active management to avoid potential pitfalls such as bit rot or storage formats becoming obsolete.

IT managers in these organizations are looking for more efficient ways to manage their digital archives and to reduce the risk of damaged or lost data as the archives age. With the digital data boom continuing to accelerate, it is becoming clear that traditional methods of managing data are not sufficient for today's needs. Organizations are looking for large-scale repository solutions that can offer:

- Efficient methods to preserve data integrity and prevent bit rot
- Reduced complexity in executing backups and administering storage systems
- High performance search and retrieval with multiterabyte archives
- An open data format that enables data to survive technology transitions
- Seamless scalability for easy and cost-effective capacity expansion
- Simplified discovery of data and the ability to search or sort based on wide ranging criteria
- Simplified capture of metadata when data is first stored in the repository
- Ability to capture and preserve the relationships between content items

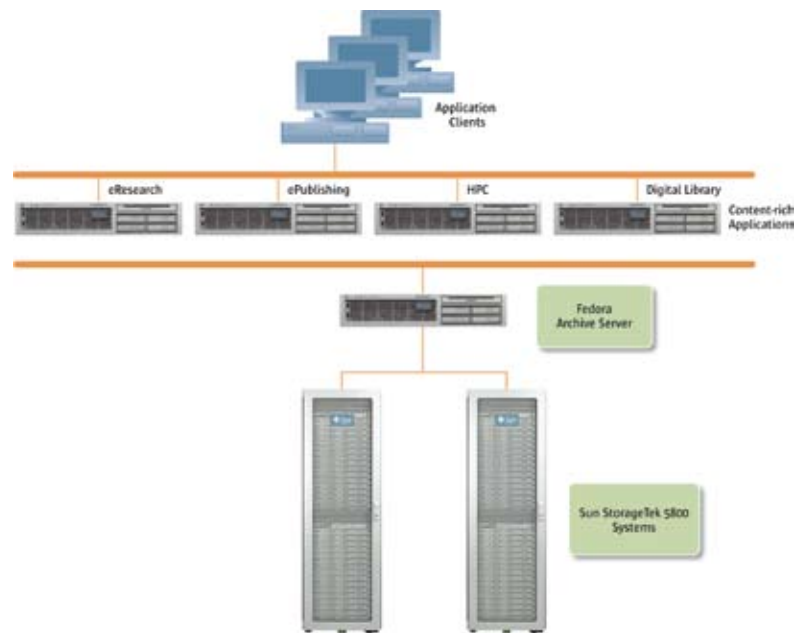
Sun and Fedora active archiving solution Together, Sun and Fedora Commons provide a next generation solution for active archiving of digital data. Organizations can now create, manage, publish, share, and preserve their digital resources in a flexible and cost-effective manner. The solution provides a petabyte-scale object store that can improve data integrity and reduce the cost of preserving data for applications such as:

- Data curation
- Object re-use and exchange
- Archival storage
- Open access publication
- Digital asset management
- e-research
- Digital libraries

The Sun StorageTek 5800 system is a third-generation object storage system that has been designed to address today's large-scale repository needs. It can be used in combination with the Fedora Commons platform to provide an intelligent data repository for applications that generate large volumes of data and must preserve the data over a long period of time.

Fedora is an open source storage repository platform that allows the creation of innovative, collaborative information spaces. It is designed for the longevity and integrity of any kind of digital content, and also offers the ability to inter-relate such content from different sources. Fedora was originally created at Cornell University under research grants from NSF and DARPA. This research evolved into a successful open source project and ultimately led to the creation of the Fedora Commons non-profit organization that is currently overseeing the development and distribution of Fedora.

With the help of researchers at Johns Hopkins University, Fedora was integrated with the StorageTek 5800 system, creating a unique combination that greatly simplifies



Applications that use Fedora can transparently access the Sun StorageTek™ 5800 system data store through the integrated interface.

archival storage. The collaborative work done between university researchers and the Sun storage team has resulted in a new module for Fedora, which seamlessly integrates the Fedora Commons platform with the StorageTek 5800. Any application that utilizes Fedora software can now easily access the StorageTek 5800 system through Fedora as shown in the figure.

The StorageTek 5800 system has been designed from its very inception to meet the needs of large data repositories such as those created under Fedora. It incorporates clustered servers for both processing and storage, allowing the system to be easily scaled as archive needs grow. Fedora has some core requirements for maximum efficiency, all of which are met by the architecture of the StorageTek 5800 system. In addition to scalability, the required attributes of high data integrity, robustness, reliability, and effective metadata management have all been designed into the StorageTek 5800 system.

Success has already been demonstrated with Fedora-based archives using the StorageTek 5800 system in many institutions. The examples below demonstrate the value of this combined solution.

Extending the value of library science at Purdue University

Traditional library science involves the acquisition of catalogs, books, and journals, mostly in the medium of paper. At Purdue University's Distributed Data Curation Center in West Lafayette, Indiana, Assistant Professor Michael Witt is one of the people leading the charge to develop new technologies for preservation and access to information. One of the key objectives of Professor Witt's work is to move upstream from the researchers' publications typically managed in library science and capture the research data that the publications are based on. His recent work has involved the application of Fedora with the StorageTek 5800 system to expand the frontiers of his field.

With access to original research data sets in the system, researchers can perform further analysis and create new views of the data, including views that encompass additional research. This form of curation allows aggregation from widely varying sources and at a far larger scale than previously possible. This is critical because modern science generates vast quantities of data whose value may be unlocked only at some distant time in the future. For instance, in the case of the Human Genome Project, the data is far more valuable than the papers that have so far been published from it.

One of the projects that Professor Witt is involved with is named Purdue e-Data. As the name suggests, it is focused on storing research data sets. Many of these data sets are held within TeraGrid, a computing infrastructure funded by the National Science Foundation that brings together tremendous computational, storage, and networking resources. Examples of the kinds of data sets that have been explored by researchers in the e-Data project include sound recordings, water quality measurements, vehicular signatures, and oncology readings.

Professor Witt has created a local repository for such data sets based on Fedora and the Sun StorageTek 5800 system. The hardware is a full-cell configuration, with two-gigabit Ethernet switches and 16 nodes representing 32 terabytes of raw storage. The architectural framework also accommodates distributed repositories. For example, it can harvest metadata from TeraGrid and subsequently ingest it into Fedora. Middleware allows query and retrieval as a Web service.

The StorageTek 5800 system has been particularly well designed for this application because it handles metadata at a very low level, as content-aware storage. A namespace is configured by the user along

with a metadata schema that is used to describe the data. Data is retrieved by querying this structured metadata in a straightforward manner using a set of Application Programming Interfaces in the Java™ language as well as C, or webdav.

Furthermore, Sun has announced that the Software Development Kit for the device will be made available as open-source, allowing for extensibility and repurposing both as functionality evolves and as data is required in future formats.

Finally, the self-healing design allows for great assurance of the integrity of data objects. Versioning is inherent in the architecture, preventing overwriting of original information. Researchers can therefore have complete trust in the original data, as well as be able to forensically track the provenance of additional layers of data with the passage of time.

“Along with its reliability and scalability, the StorageTek 5800 includes metadata and preservation functionality at a low level. This makes it very well-suited architecturally for digital library and archiving applications.”

Michael Witt
Assistant Professor of Library Science
Purdue University

Open solution brings longevity to digital assets at Oxford University

Oxford University Library Services in England supports over 110 institutional libraries. They are in the process of deploying Fedora as part of a repository framework for all digital assets across these libraries. The repository is being implemented using the StorageTek 5800 system in a two-cell configuration, providing 64 terabytes of raw storage via four gigabit Ethernet switches.

Currently, these digital assets consist of published papers, conference proceedings, and electronic theses, all in native digital format. The output from current and historical digitization programs is also captured in the system.

Neil Jefferies is the R&D Project Manager for SERS (Systems and E-Research Services), which is the IT arm of the university library services. In this role, he is responsible for the Fedora implementation with the StorageTek 5800 system. “There are a number of advantages in using this configuration for the preservation of digital assets at Oxford,” said Jefferies. “First, the architecture is designed for resilience and long-term stability. The self-healing firmware and block-level checksum error-checking ensures that data integrity is not compromised even in the event of degradation or failure of individual underlying disks.” By keeping the processing close to the storage in the same device, validation of data can be done in a reliable fashion.

A second advantage pointed out by Mr. Jefferies is that the system allows archival in native digital object format instead of having to translate it to fit the needs of some other form of file system. In addition, since Sun is offering an open source Application Programming Interface, Sun is virtually guaranteeing that there will be no format obsolescence. There is no constraint to have

“We are pleased to be a Center of Excellence with Sun Microsystems and an early site for using the StorageTek 5800. We have always found the company highly supportive of and responsive to the needs of the academic community, and especially to those of libraries, education and research.”

Susan Stein

Federated Service Manager,
University of Alberta

to hew to a particular metadata schema. This open path to subsequent developments and progress in storage technology help to ensure longevity that is well in excess of conventional archival solutions.

Flexibility of open source proves useful to the Alberta Library

The Alberta Library (TAL) is a non-profit consortium of 300 libraries that provides services to residents of Canadian cities ranging from Calgary to Banff, and Edmonton to Medicine Hat. Susan Stein is the IT operations manager at TAL, responsible for offering services and improving access to the electronic resources across all these libraries.

The consortium stores digital content for its constituent members who would otherwise have no such archival facility. One of its key roles is to facilitate federated searching across all content repositories in the province. TAL has been designing and implementing a solution based on Fedora and the StorageTek 5800 for much of the last year.

The consortium is currently using a single-cell configuration of 16 nodes (32 TB). Ms. Stein says that a benefit of starting with this hardware is the ability to scale seamlessly to a multi-cell configuration as the project gets larger. Another strength of the StorageTek 5800 system’s design is its data integrity and preservation capabilities.

Access to the stored data is made simple due to embedded extensible metadata using simple Application Programming Interfaces (APIs) for later query and retrieval.

By using the open source Fedora platform for design of the information repository, the Alberta Library is able to take advantage of the numerous modules made freely available by contributors worldwide. Of the many modules hosted in the Fedora

Commons collection, those associated with authentication, authorization and digital rights management have been especially useful to TAL. In turn, the consortium is working on contributions in the area of interoperability and linking to various national and international repositories.

The Mountain Parks Project is an example of useful and engaging data hosted by the Alberta Library. In 1915, an explorer climbed all the major peaks in the Rocky Mountains in Alberta, and took photographs and measurements at every stage. This effort resulted in a collection of thousands of images along with almost a hundred thousand glass-lantern slides, and data that was logged in dozens of notebooks. In the late 1990s, researchers climbed the same peaks and took similar photographs and readings. The combined images and data have been made open and available as part of the project, allowing comparisons that have proved useful in studies ranging from glacier movement to climate change. By adding mapping technology and Geographic Information Systems as overlays, this same information can be made pertinent for both high-end research and as part of the educational curriculum.

The flexibility of the open source Fedora platform and the ease with which data can be managed and preserved in the Sun StorageTek 5800 system have given the Alberta Library the opportunity to explore new uses of research data that will benefit an extended community.

The cure for digital overload

The StorageTek 5800 system is an ideal storage solution for large scale repositories built on the Fedora platform. It provides intelligent storage functionality at a very low level, reducing the manual effort required to maintain the data archive and

simplifying searches. Both Fedora software and the StorageTek 5800 system are based on a highly scalable architecture and an open source structure to enable modular growth and protect your investments in data and applications. Together, the StorageTek 5800 system and Fedora software bring extraordinary flexibility to creating, managing, publishing, and preserving digital content.

Intelligent storage

The StorageTek 5800 system provides object-based storage, thus moving the process of managing space and scalability down into the storage layer. There are no RAID sets to manage, no Storage Area Networks (SANs) to architect, and no file systems to mount. Administration thus becomes a very simple task, reducing the cost and complexity maintaining the archive.

Primary storage on traditional systems can be difficult to manage and it can be a challenge to access data due to limited views of it through a file directory structure or through an application using raw devices. As more and more files are added, the directory structure can become extremely complex and difficult to navigate. In addition, access is limited through a single view. The object based storage on the StorageTek 5800 system with its flat namespace offers multiple views of the data through a simple query interface that gives researchers the flexibility to search for the data they need. Searches are based on metadata information that is created when the objects are originally stored.

The StorageTek 5800 system is also self-healing, meaning that it detects failures and reassembles the data away from affected hardware. When failed components are later replaced, Sun software that runs across all nodes automatically detects new replacement hardware and spreads the data across the replacement hardware. All storage

nodes internal to the storage system have equal capacity utilization, including new or replaced hardware.

The software also does automatic data integrity checking and correction and includes block and file level checksums. Checksums are stored separate from the object and are validated upon object retrieval. The data encoding and reconstruction process means that any two disks on different nodes (or more than two disks on the same node) can be lost without losing data. The system can even sustain multiple, simultaneous drive failures without data loss.

Modular growth without disruption

Both the StorageTek 5800 system and the Fedora Commons platform are architected for scalability, allowing institutions to start small and expand as needed.

The scalability of the StorageTek 5800 system results from its modular architecture. A half-cell is composed of 8 nodes, and can be easily upgraded to a full 16-node cell. The system can then be expanded to a multicell configuration as data needs increase. With each new cell, more compute power is also added to the system, making the system scalable to hundreds of terabytes.

Fedora is similarly scalable since it encapsulates all data as XML objects. As the objects then increase in number, they can be stored on devices ranging from a single small server all the way up to an enterprise-wide multiserver system. In this solution, Fedora maintains a registry of data that is stored across multiple nodes. In addition, Fedora captures the relationships between objects, which is often as important as the data itself. Since the relationship information is stored the same way as other data, the StorageTek 5800 system can preserve it reliably for as long as it is needed. Some of these

relationship links are also indexed by the StorageTek 5800 system to enable very fast searches and fast access to data, even for multiple terabyte digital archives.

Investment protection through open source

One of the most challenging aspects of long term archival solutions is that data must survive multiple generations of technology and complete turnover in the staff who manage the data. Open source solutions such as the Fedora Commons platform and the technology of the StorageTek 5800 system provide greater assurance that the digital archives can be preserved regardless of technology upgrades or staffing issues.

Many institutions have adopted Fedora with enthusiasm because of its open source structure, its modularity, and its extensibility. All Fedora Commons code is publicly and freely available on sourceforge.net. Customers can use it in its entirety, or as relevant application pieces. In its tenth release since 2003, the code is tested and robust. In addition, a large and vibrant user community has produced a wide variety of user interfaces, front ends, middleware, applications, and utilities.

The StorageTek 5800 system represents the first commercially available fixed content storage system today that comes with a commitment to open source its code. Since it is based on OpenSolaris™ and Sun's commitment to open interfaces, the StorageTek 5800 system protects against proprietary lock-in. Sun is committed to present open APIs on both the client and the server side. Data is also kept in an open storage format so that future devices that may be based on new advances in technology will be able to read it. Customers can therefore be virtually assured of being able to retrieve their data well into the future. More details can be found at opensolaris.org/os/project/honeycomb/.

Flexible data management

Fedora is built with a flexible framework that allows easy extensibility not just to new formats of data, but also to new storage platforms and devices. New devices can be integrated by simply developing a new Fedora module such as the one that provides seamless integration to the Sun StorageTek 5800 system.

Fedora enables the creation of collaborative, integrated information spaces where any information entity can be linked to any other entity. Multiple data sets can be linked to multiple research papers or other sources via multiple paths. Since this relationship information is maintained within Fedora digital objects and preserved in the StorageTek 5800 system, there is less risk of broken links and it is easy for users to find similar data, even for objects that are many years old.

Users can perform post-facto analysis such as locating all papers that resulted from a single set of data. Or conversely, one can easily find all data sets referenced by a particular paper, author, or institution. This understanding of linkages can result in new hypotheses and research directions that otherwise would not have been possible.

Sun and Fedora — a winning combination

Together, Sun and Fedora bring a breakthrough in innovation for digital content. The Sun StorageTek 5800 system offers the only commercially available petabyte-scale fixed content store designed to be open-source and that meets all the required attributes of simplicity, expandability, programmability and data integrity. Institutions and corporations can now more fully utilize their digital content while reducing the cost and complexity of preserving it throughout its full lifecycle.

Get started now with free downloads

You can get started immediately with free download software from Sun and Fedora. The Fedora Commons code is available for download at fedora-commons.org/developers and the emulator for the Sun StorageTek 5800 is included in the StorageTek 5800 SDK 1.1, which is available at sun.com/download/products.xml?id=465eed06. The StorageTek 5800 emulator mimics the behavior of a StorageTek 5800 system so that you can test your application without having a StorageTek 5800 physically available.

Learn More

To learn more about Fedora and Sun solutions for digital content, visit fedora-commons.org and sun.com/storagetek/.