Sun Open Archive Framework and Fedora Repository Solutions

Technical Brief

June 2009

Table of Contents

Introduction

"Sun clearly has heard and understood the critical aspects of archiving and, to provide the means to durably store and access our valuable digital assets, the core of the archive must be open! In partnership with the archival and open source communities, Sun has delivered a scalable, tested set of software and hardware components that enables the construction of trusted digital archive solutions. With full understanding that this is a never-ending quest with considerable work yet to be done, Sun has proven that ongoing partnering with these communities, in a continuous process to improve open archive technology, is the best way to move forward."

Daniel Davis
Systems Integration Specialist
Fedora Commons Affiliate
Cornell University

Archived information can no longer be placed in a 'store and ignore' location. Value of information greatly increases when it is shared and reused over time and across disciplines, therefore, archive solutions must now include the ability to broadly and rapidly access information as well as rapidly ingest information. The Sun Open Archive Framework addresses rapid growth and broader use of this information which is driving the need for this new approach to preserving and managing archived data. In addition to the challenge of storing and accessing the information comes the need to assure its integrity and to take the data through new technologies over time. Sun has partnered with Fedora Commons to provide a solution that can help organizations store, share and reuse their digital information in a flexible and cost-effective manner. The solution combines the Fedora Commons open source Fedora Repository software with Sun storage hardware and software products. Together, these provide multi-petabyte scale archives that greatly simplify the task of preserving massive amounts of data over long periods of time and keep the information sharable, reusable and sustainable.

Archive Crosses Many Industries

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. Today, this printed information is now digitized and being stored on computer systems. This task to archive digital information has now spread to all industries including Healthcare, Media and Entertainment, Financial Services, and Government. People and systems are generating digital information on a scale never before imagined. Some information, such as MRIs and even eMail is born an archive and must be treated as such. Regulatory compliance has created new archive requirements. Many research organizations utilize High Performance Computing (HPC) applications that commonly generate data sets that fill multiple terabytes of storage. Even if the data is 'massaged' into a single conclusion, the original information must be saved for years as supporting evidence of that conclusion. It is easy to see how today's archives can quickly grow to the multi-petabyte scale.

As information becomes sharable and reusable, it also becomes more valuable. Research data that is generated for a single calculation or research project, when made available to other cross-disciplinary researchers to apply different calculations or formulas, results in additional conclusions greatly improving time to market of new products. Making test results of healthcare broadly and securely available to many doctors as well as the patient reduces the number of tests run and reduces errors in diagnosis. This information must survive transitions in staff or user communities as well as technology changes over a period of several decades or even multiple

centuries.

All organizations with IT staff are looking for more efficient ways to manage their digital information and to reduce the risk of damaged or lost data and mitigate the risk involved in technology or vendor obsolescence. It is clear that traditional methods of managing information are not sufficient for today's needs. Organizations are looking for large-scale archive solutions that can offer:

Save, share and reuse information possibly for hundreds of years

- Generate revenue through reuse of information
- Share information between unrelated applications for broader more encompassing conclusions and decision making
- Meet compliance regulations through timely information access

Cost Savings

- Save up to 90% on the cost, 70% on energy and 50% on space
- Simplify Administration reducing management costs
- No License Fees
- No vendor lock-in

Migration through space and time: Designed for Change

- Easy migration through replication to new storage technology as storage hardware changes
- Dynamically make copies of data on multiple and different media tiers based on policy

Scale Capacity and Performance non-disruptively and with ease

- Easy to expand capacity on systems with scalable storage arrays
- Dynamic use of tape for petabyte(PB) scaling

Performance optimization through a unique solid state disk implementation

Freedom of choice

- Storage Preservation Software with a broad choice of industry standard storage hardware

Sun Open Archive Framework

Sun's Open Archive Framework brings together the components required to provide the highest level of data availability and preservation in addition to future-proofing data from technology changes over time. Through the use of open standards based, community supported software, preservation features such as data protection, integrity checking and policy-based management are provided in the storage preservation software component. Independent of this storage preservation software, the end-user is empowered with choice and flexibility to use industry-standard storage device building blocks to create a cost-effective and scalable archive storage infrastructure. To complete the Open Archive Framework, the end-user has flexibility in a wide choice of archive applications to manage their digital assets, making them searchable, reusable and sharable, thus, increasing their value and turning their digital assets into a differentiator. This Open Archive Framework consisting of storage preservation software, flexibility in choosing industry-standard disk and optionally

tape where it makes sense, and freedom in choosing from a wide variety of archive applications provides archive solutions that are designed for economical and innovative change and will reliably take customer data into the future.

Four Layers of an Archive

Figure 1 describes how information moves through the four key layers of a customer's archive solution. Digital assets are first created, then managed by Fedora, then preserved and stored. The Sun Open Archive Foundation provides the preservation and physical storage layers in this data flow.

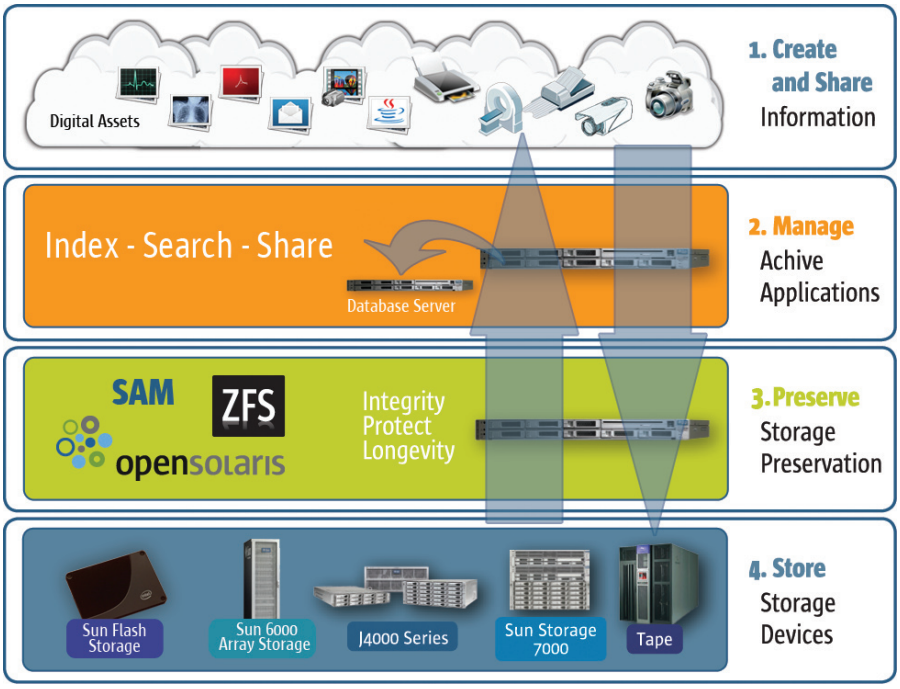

Figure 1: Four Layers of an Archive

Customer Applications and Data (creation)

Understanding the customer's applications and data is critical to architecting an archive solution that addresses their specific needs. The value of data increases as its availability broadens today as well as its availability in the future. Re-use of digital assets across communities and applications and, most importantly, across time, has the potential of becoming a revenue-generating asset through increased speed of research results and results in added protection from potential litigation. Examples of customer data for archive include: office documents, medical images, emails, research data, web content, scanned documents.....

Understanding the data characteristics also helps architect the proper solution. How much data, how big are the objects or files, what are the potential retention

periods, what are the sharing opportunities, what security is needed, are just a few of the questions to be asked.

Archive Applications (management)

The archive application increases the value of information by making it searchable and by creating relationships between the objects. Archive applications ingest and retrieve this information, creates searchable metadata and applies the defined security rules, making the information available to any application and takes the data into the future. Whether the assets are born digital or they become digital over time, Sun has a broad portfolio of partner applications to archive the assets. Many of these partners specialize in specific industries such as Virtual Research Environment, Media and Entertainment, IT and Service Providers, Research Computing, Education, Healthcare and Government.

Storage Preservation Software Layer (preservation)

This abstraction layer pioneered by Sun provides all the necessary storage preservation, policy and persistence required of archived digital assets and is agnostic to the physical storage. Examples of this include data integrity verification and repair resulting in "19 Nines" certainty, checksumming and protection (Raid, snapshot, clone), analytics, predictive self-healing sensors, policy based data migration and simplified management. Based on open software and standards, the customer is not locked into a proprietary API or a specific vendor. As hardware technology advances, digital assets are easily migrated from old technology to new while maintaining complete integrity and persistence. This differentiates Sun's Open Archive from traditional architectures where the storage preservation is locked into the hardware itself.

The Storage Preservation Software Layer protects against data corruption and data loss with advanced data integrity functionality and assurances.

The following software components are included in this layer

Management and Control Software implemented by OpenSolaris, ZFS and SAM provides:

- Integrated data management with ZFS
- Integrity verification and repair
- Checksumming and protection
- Compression
- Analytics to quickly identify problems-resulting in preemptive repairs
- Delivers required access and availability requirements
- "19 nines" data integrity
- Predictive self-healing sensors that continuously perform background system diagnosis, including silent error detection and correction, reducing downtime by 32-52%.

Automated Hierarchical Storage Manager (HSM) implemented by the choice of Sun Storage Archive Manager (SAM) which provides: open archive (gtar) format tiered storage including tape, multiple archive copies, remote archiving, business compliance features for long term data retention (including WORM support), automated policy management, and support for ZFS volumes.

Connectivity includes a wide array of current network protocols that are supported through OpenSolaris at no additional cost. As technology advances and improves and new protocols are developed through open communities in the future, they are easily adopted by the Open Archive Foundation.

Physical devices: Sun Industry Standard Building Blocks and Tape (storage)

Because the Sun Open Archive Foundation does not lock the Preservation Software to a particular hardware platform, the customer is free to use whichever Industry Standard Storage best meets the their requirements. The Sun storage disk and Solid State Disk (SSD), configured in a unique manner called a hybrid storage pool, is a set of cost-effective building blocks for storage used to create an open, flexible storage infrastructure. In addition to disk and server technology, this layer also includes the option to include tape devices managed by SAM, which provides the ability to scale capacity to multi-petabytes of archived information and yet it remains available, sharable and reusable. Through Sun's innovative implementation of solid state disk (SSD) technology, archive application IO performance is dramatically improved.

Fedora Repository

Fedora Repository is an open source archive 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 Repository.

Fedora Commons provides an advanced platform for deploying preservation-enabled repository systems that can seamlessly be integrated with existing applications and IT infrastructure. All content, and the metadata essential for making sense of this content, is kept in ordinary files which can be replicated to remote locations for added safety. The metadata is stored as XML which is open and accessible for the foreseeable future. In the event of a disaster, the entire Fedora Repository can be rebuilt from those files eliminating the risk of any single point of failure. As content formats change, Fedora provides support for migrating files to new formats, option-

ally keeping the previous versions for security. To help manage formats, Fedora is not limited to MIME types but can use format identifiers from any format registry or scheme at any desired precision. Fedora has an advanced capability that enables adding customized services to content types that hides implementation details and provides a stable way for your application to access the content—even as it is migrated to new formats—and easing the introduction of new technologies over time. Changes to content or metadata are recorded in audit trails, also in files, and are combined with cryptographic checksums to help ensure authenticity. With an eye to the future, Fedora enables plugging in viewers and emulators so that obsolete formats can be accessed even if the original software is no longer available.

Sun and Fedora Open Archive Solutions

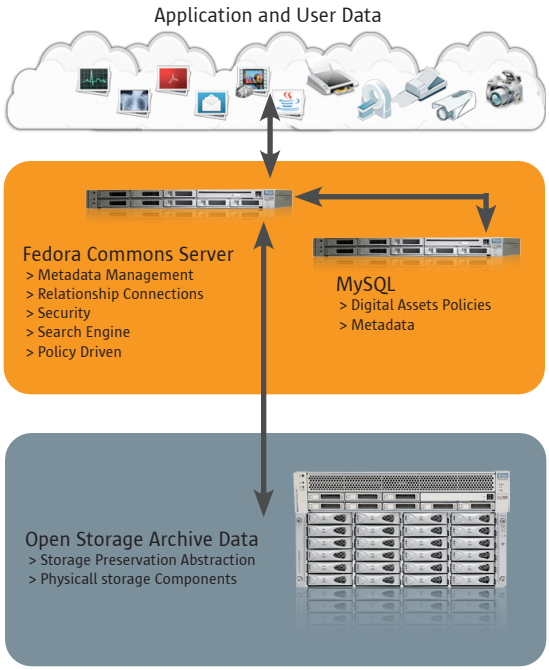

Figure 2: Sun and Fedora Open Archive Solutions

Together, Sun and Fedora Commons provide a next generation solution for archiving information. Organizations can now create, manage, publish, share and preserve their digital resources in a flexible and cost-effective manner. Figure 2 describes this integration.

The solution provides a multi-petabyte scale solution that can improve data integrity and reduce the cost of preserving information for applications such as:

Virtual Research Environments

While we all know that computers help in medical research and healthcare delivery,

how many of us realize just how much technical data and personal information is generated and archived? Research studies generate vast amounts of varied data including text-based client consent forms to genomic mapping and sequencing data.

The value of research is multiplied when the data it produces can be easily found, shared, and reused by other researchers. In addition, medical or drug research is regulated, therefore all data must be managed and stored using Standard Operating Procedures (SOPs) developed in conformance with federal and good clinical practice guidelines.

The Sun and Fedora Repository Open Archive Solution provides a comprehensive solution for corporations and research communities conducting bioscience research. It provides research groups with a collaborative web environment for the creation and stewardship of research which also includes the storage, transformation, long-term preservation and distribution of data and information.

Government and Digital Repositories

For customers who need to store information as well as make it available globally, the Open Archive Framework provides a clear direction. The massive amount of information that is being generated by both the Government and Education space begs for the ability to search and share and reuse. Open source applications such as Fedora Repository have been architected to handle billions of small and large files, creating and managing the metadata. The National Science Foundation has offered a grant to several organizations, both Government and Education, for a project that will ensure interoperability between many different archive applications, including Fedora Repository, further ensuring the information's sharability and reusability.

Health Industry

Medical information and records are mostly siloed today, locked into a hospital or doctor's office. Customer data (Xray, Cat Scan, MRI) is generated by devices with their own management software, such as the archive application GE Healthcare PACS (picture archiving and communication system) software.

Key trends are emerging that are driving the healthcare industry to utilize technology as a means of improving care that cannot be solved through siloed information. Some of the key trends include the need to share patient care and patient information across various service vendors which would maximize the service that is provided. It would also reduce and prevent the number of medical errors and increase the timeliness and accuracy of diagnosis. Broadening the communication channels to include the patient as well as all of their medical service providers alone will improve healthcare. This can initially be accomplished simply by providing the ability to connect existing systems and addressing the growing size and number of records that must be maintained throughout and beyond the life of the patient. Secure yet

easy access to health records from multiple locations is a key part of the current administration's health care initiatives Sun's Open Archive Framework is architected to meet that requirement.

Broadcast/video

The opportunity in this industry covers several market segments that can be investigated for an Open Archive Framework solution. These segments include Broadcast and Film, Cable and Satellite, Gaming, Internet Services, Media and Publishing, Music and Sports. All segments need to store, manage, distribute & consume digital content. In the Broadcast and Film industry, content such as film can be resurrected, new technology applied to improve the film and it can be re-distributed for new revenue. In the Media space, content can be shared with multiple distribution channels, again growing revenue from a single source.

Fedora Repository and Sun Storage Reference Architectures

General Archive Architectures

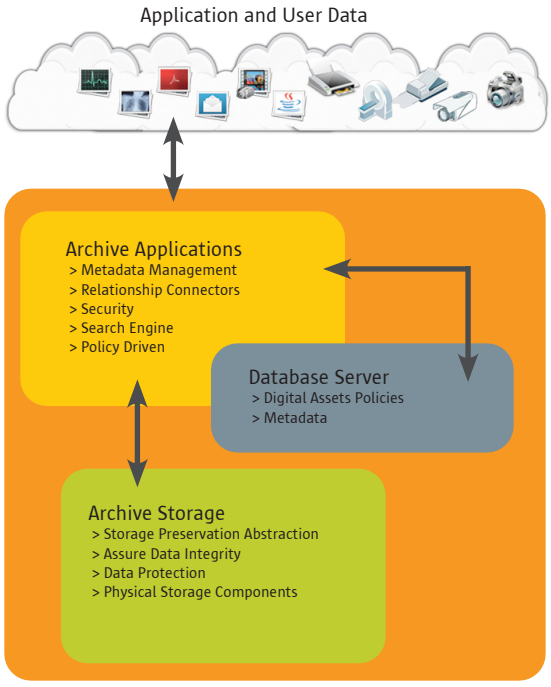

Figure 3: General Archive Architecture

As shown above, all Archive applications follow a basic architecture for preservation of any type of digitized information. Variations are based on such items as the type of data, its source, use of commercial or open source application code, use of a database or use of proprietary indexing, capacity, ingest rate requirements and ac-

cess rate requirements. In addition, other features and differentiators could include full text indexing, retention, linking objects together, improved metadata, improved security, improved search capability, industry specific capabilities and many other features. There are also variations for the storage where the information ultimately lives. Features that should be taken into consideration includes proprietary vs. open standards storage, retention period requirements, ingest and access rates, protection from local and site failure as well as tampering and intentional data corruption.

Fedora and Sun Open Archive Architectures

This section describes the key components of Sun's Open Archive with the Fedora Repository software. Two configurations have been tested and have different characteristics in performance and in manageability. Selection of your architecture will be determined by your specific requirements which include

- Expected archive capacity and growth rate
- Archived document characteristics
- Ingest, access and update workload ratio
- Cost sensitivity
- Required performance

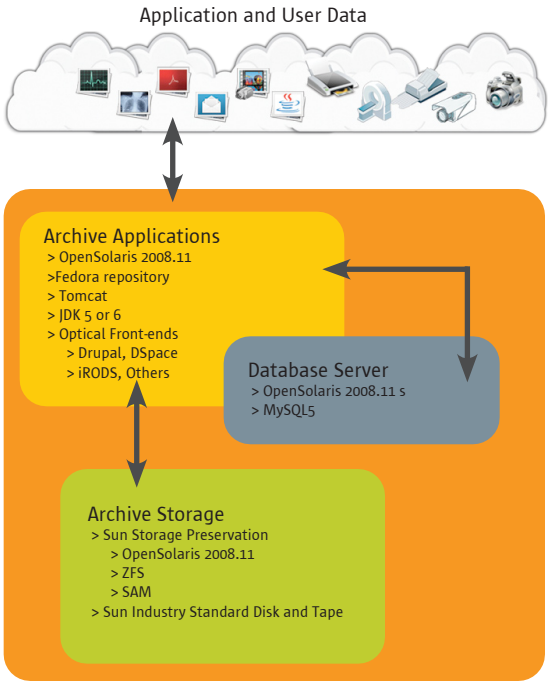

Figure 4: Fedora Repository and Sun Software Architecture

In the Reference Architectures as shown in Figure 4, Fedora Repository software is used to capture, generate appropriate metadata and store digital objects from any source and then access them with the appropriate security in place. The platforms that this archive solution requires include:

- Fedora Repository Server
- Database Server
- Sun Storage to store the objects

An additional software application such as open source Drupal, Dspace, iRODS or others can run on the Fedora Repository Server and be used as a front end to Fedora Repository. A project from University of Prince Edward Island (UPEI) called Islandora ties Drupal to Fedora as an end-to-end solution. Drupal is a popular open source social publishing system that blends web content management and social media capabilities (such as blogs and community discussion forums). It provides an ideal, easy-to-use user interface to the Fedora Repository when combined with the Islandora module. Islandora is an open source module for the Drupal web content management system which allows Drupal to act as a web-based front end to the digital repository and preservation platform from Fedora Commons. Islandora uses Content models to determine what mimetypes are allowed to be ingested into Fedora Repository and what to do with the object on ingest. For instance, a PDF content model may tell the module to create a thumbnail and ingest the thumbnail as a datastream along with the actual PDF datastream. The module also enables viewing and management of Fedora Repository objects. This includes functions such as ingesting, purging, adding data streams, searching and browsing by collection.

Fedora and Sun Open Archive Test Environment

Sun and Fedora Commons have tested several reference configurations for performance, functionality and scalability. The goal of the test environment was to characterize Fedora Repository ingest, access and update activity using different storage architectures in order to help a customer select the proper configuration that will meet their requirements. For this testing, a data generator was used on a separate server to drive input into the Fedora Repository Server. This tool is based on the open source The Grinder utility. The tool itself is available on http://grinder.sourceforge.net). The scripting to automate the load steps will be made available as open source in the near future.

The second server supports the MySQL database on which Fedora Repository updates the metadata during the ingest process and accesses during the search process. Other databases have been tested by Fedora Commons, however, this testing used MySQL and it is a well proven and recommended open database in order to have a complete open software solution.

The storage was tested in two different configurations with different performance results. One configuration directly attached Just A Bunch Of Disks (JBOD) to the Fedora Repository Server as in Figure 5. The second configuration used network attached storage, Sun Storage 7000 as shown in Figure 7 or Open Storage as shown in Figure 6. In order to select the proper configuration for your requirements, it is important

to understand ingest and access requirements, capacity requirements and ease of management requirements.

Sun and Fedora with Direct Attached Storage

This solution provides the highest ingest rates of small (15k, the size of a small PDF or text file), medium (20M, the size of a typical YouTube video) and large (700MB, a CD image) objects. A single Fedora Repository server with JBODs attached can scale to 96TB RAW storage. To scale this solution to a larger capacity, additional Fedora Repository servers can be added with direct attached storage to implement a repository federation.

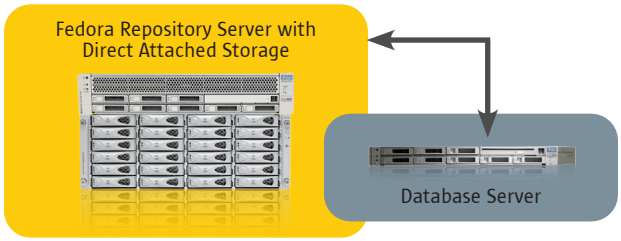

Figure 5: Test Environment #1: Fedora Repository Server with Direct Attached Storage

Test Environment #1: Fedora Repository Server Direct Attached Storage

Layer	Server	Software	Storage
Application	*Sun Fire x4150* *- 2 4-core Xeon CPUs @ 3.16GHz* *- 32GB memory* *- 4 x 146GB internal SAS disks* *- 4 x GigE Ethernet*	*Fedora 3.1 (w/Tomcat)* *OpenSolaris 2009.06*	
Database	*Sun Fire x4150* *- 2 4-core Xeon CPUs @ 2.66GHz* *- 32GB memory* *- 4 x 146GB internal SAS disks* *- 4 x GigE Ethernet*	*MySQL 5* *OpenSolaris 2009.06*	
Storage		*OpenSolaris 2009.06* *ZFS*	*Sun J4200 JBOD disk array* *- 12 7.2K 250GB SATA II drives* *- SAS interface* *(utilizing ZFS pools)* *(direct connect to App Server)*

Sun and Fedora with Network Attached Storage

This architecture provides the highest scalability of storage capacity. There are two options for network attached storage. One option is Open Storage utilizing a storage server (X4100) with JBODs (J4400) attached. The second option is the Sun Storage 7410(C) Unified Storage System which provides higher capacity, higher availability and ease of management. Both options use OpenSolaris/ZFS to protect and provide data integrity.

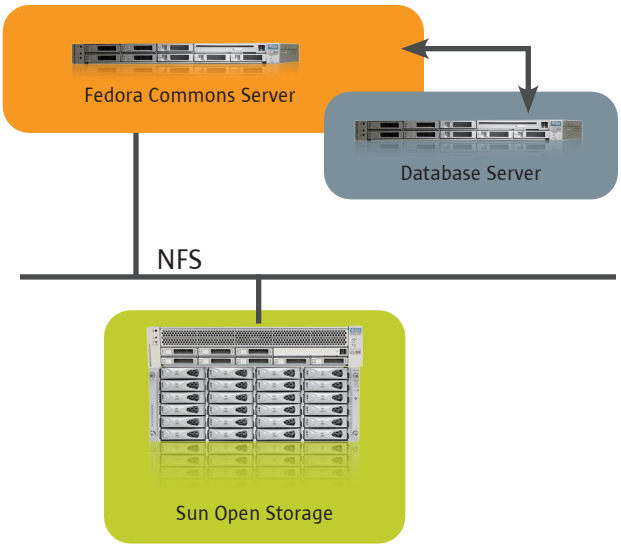

Figure 6: Test Environment #2: Fedora Repository Server and Open Storage

Test Environment #2: Fedora Repository and Sun Open Storage

Layer	Server	Software	Storage
Application	*Sun Fire x4150* *- 2 4-core Xeon CPUs @ 3.16GHz* *- 32GB memory* *- 4 x 146GB internal SAS disks* *- 4 x GigE Ethernet*	*Fedora 3.2 (w/Tomcat)* *OpenSolaris 2009.06*	
Database	*Sun Fire x4150* *- 2 4-core Xeon CPUs @ 2.66GHz* *- 32GB memory* *- 4 x 146GB internal SAS disks* *- 4 x GigE Ethernet*	*MySQL 5* *OpenSolaris 2009.06*	
Storage	*Sun Fire x4150* *- 2 4-core Xeon CPUs @ 2.66GHz* *- 32GB memory*	*OpenSolaris 2009.06*	*Sun J4200 JBOD disk array* *- 12 x 250GB SATA II 7.2K drives* *- SAS interface*

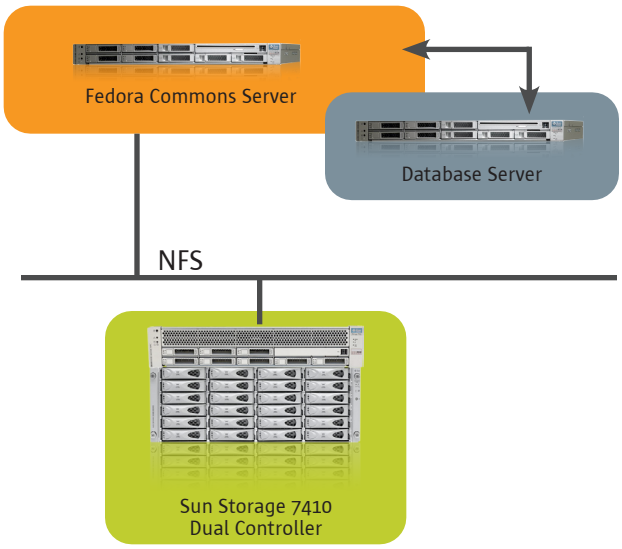

Figure 7: Test Environment #3: Fedora Repository Server Sun Storage 7410 Dual Controller

Test Environment #3: Fedora Repository with Sun Storage 7410 Dual Controller

Layer	Server	Software	Storage
Application	*Sun Fire x4150* *- 2 4-core Xeon CPUs @ 3.16GHz* *- 32GB memory* *- 4 x 146GB internal SAS disks* *- 4 x GigE Ethernet*	*Fedora 3.1 (w/Tomcat)* *OpenSolaris 2009.06*	
Database	*Sun Fire x4150* *- 2 4-core Xeon CPUs @ 2.66GHz* *- 32GB memory* *- 4 x 146GB internal SAS disks* *- 4 x GigE Ethernet*	*MySQL 5* *OpenSolaris 2009.06*	
Storage	Testing to be completed in the near future		

A test report with test results of the three configurations listed above and interpretation of those results has been published by ISV-Engineering and can be found on website http://blogs.sun.com/err/resource/FedoraSizing.pdf. These results can be combined with ingest requirements, available budget and system management requirements to select the appropriate solution.

Additional Architecture Options

Fedora Repository Server

The Fedora Repository server is to be selected based on performance requirements. The Sun X4150 was used during testing and was not a bottleneck. During a test run using internal disks (not a realistic configuration, this was run to challenge the server, not the storage), the server was still not a bottleneck.

Fedora Repository Server Options

Component	Recommended HW
Server	*X4100 or X4200Server.*
Memory	*16GB when not running the Mulgara Triple Store Software* *32GB when the Mulgara Triple Store Software is turned on*
Internal Disks	*8 146GB SAS Drives.* *> 2 mirrored for the OS* *> 4 mirrored for Fedora Repository* *Configure using ZFS Pools*
Software	*OpenSolaris release 2008.11* *ZFS* *Fedora Repository SW version 3.X* *> Mulgara Triple Store* *JAVA 5 or 6*

Database Server

The Database server is used by Fedora Repository to store metadata and used during searches and access. The recommended database is MySQL, however, other databases, including Oracle 8.1 and higher and Postgresql are supported for production use and Apache Derby is an option in Fedora Repository 3.2 for non-production uses (development, testing, demo).

Database Server Options

Component	Recommended HW
Server	*X4100 Server.*
Memory	*16GB* *32GB*
Internal Disks	*8 146GB SAS Drives.* *> 2 mirrored for the OS* *> 4 mirrored for the database. The size of the database is determined by the number of objects in the repository, not the capacity used. If necessary, this server can be configured with J4200 or J4400 to support this database.* *Configure using ZFS Pools*
Software	*OpenSolaris release 2008.11* *ZFSDatabase Selection:* *> MySQL (used during testing)* *> Oracle 8.1 or higher* *> Postgresql* *> Apache Derby for non-production work*

Other Notes and Observations

The Solid State Disk use in a Hybrid Storage Pool configuration was tested. The ingest rate comparison makes it clear that specific type and size of ingested records are not appropriate for this pool. Further testing will be completed and documented to further clarify this for architecting a specific solution for a specific set of requirements.

Additional Test Plans and Architectures.

Fedora and HSM: It is clear that some archive solutions can reach the multi-petabyte size. This can be accomplished in two very different ways. One is to scale the Fedora Repository server horizontally and add additional Fedora Repository servers which can then be federated for a single search. A second method is to accomplish this on a single Fedora Repository Server through the use of SAM as a method to move data from primary disk to tape and releasing It from primary disk, leaving a pointer in the metadata and yet make it visible to Fedora Repository as if it were on disk. Additionally, a second copy could be saved on disk or remotely. Up to 4 copies can be maintained by SAM.

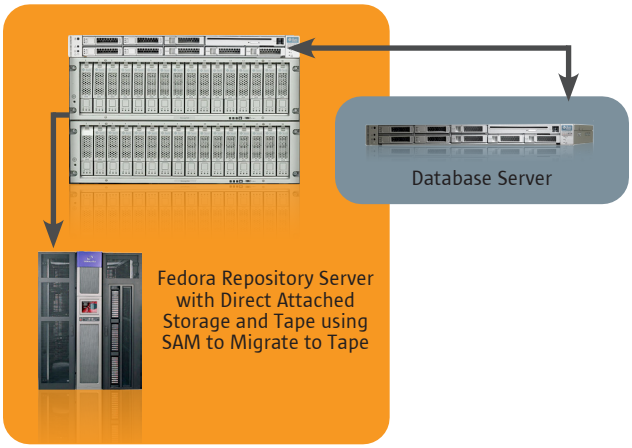

Figure 8: Use of SAM to migrate data to tape

Single Box Solution: An additional test will be with a single system solution as shown in figure 9. This solution will include all software and storage in a single storage system. When integrated by a qualified independent storage or system vendor, customers have the opportunity to purchase an archive solution preloaded and ready to define the interface to their data. The midsize market as well as groups within a large corporation can take advantage of this.

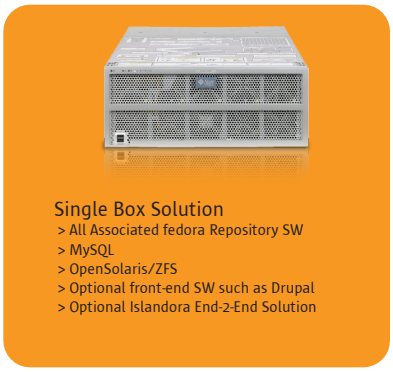

Figure 9: Single Box Integrated Archive Solution

Conclusion

The Sun Open Archive Framework is leading in the new wave of Archive by creating solutions that future-proof information from technology changes and provide the highest level of information availability and preservation. The result is information that increases in value over time by making it sharable, reusable and sustainable. The known facts are that people move on and technology changes but information grows by the second and must be archived and yet available for years, sometimes hundreds of years. Bringing together Sun's broad and powerful storage hardware and software portfolio with Fedora Commons Repository Archive Software will take information through time, make it reusable for current and future applications and make it sharable across various disciplines.

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA Phone 1-650-960-1300 or 1-800-555-9SUN (9786) Web sun.com

© 2009 Sun Microsystems, Inc. All rights reserved. Sun, Sun Microsystems, the Sun logo, are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and other countries. Information subject to change without notice.

Printed in USA 06/09 563208 SYWP14930.A4-0