



Sun QFS and Sun Storage Archive Manager (SAM)

Harriet Coverston

Distinguished Engineer
Sun Microsystems, Inc.

September, 2009



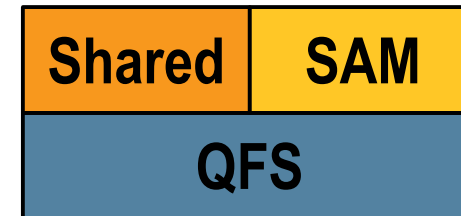
Sun Storage Software

Advanced Data Management Software

- Sun QFS – Shared File System
 - > High performance parallel SAN file system
 - > Native Linux Clients
 - > http://www.sun.com/storage/management_software/data_management/qfs
- Sun Storage Archive Manager (SAM)
 - > Policy-based automatic data protection
 - > Tiered Storage
 - > http://www.sun.com/storage/management_software/data_management/sam

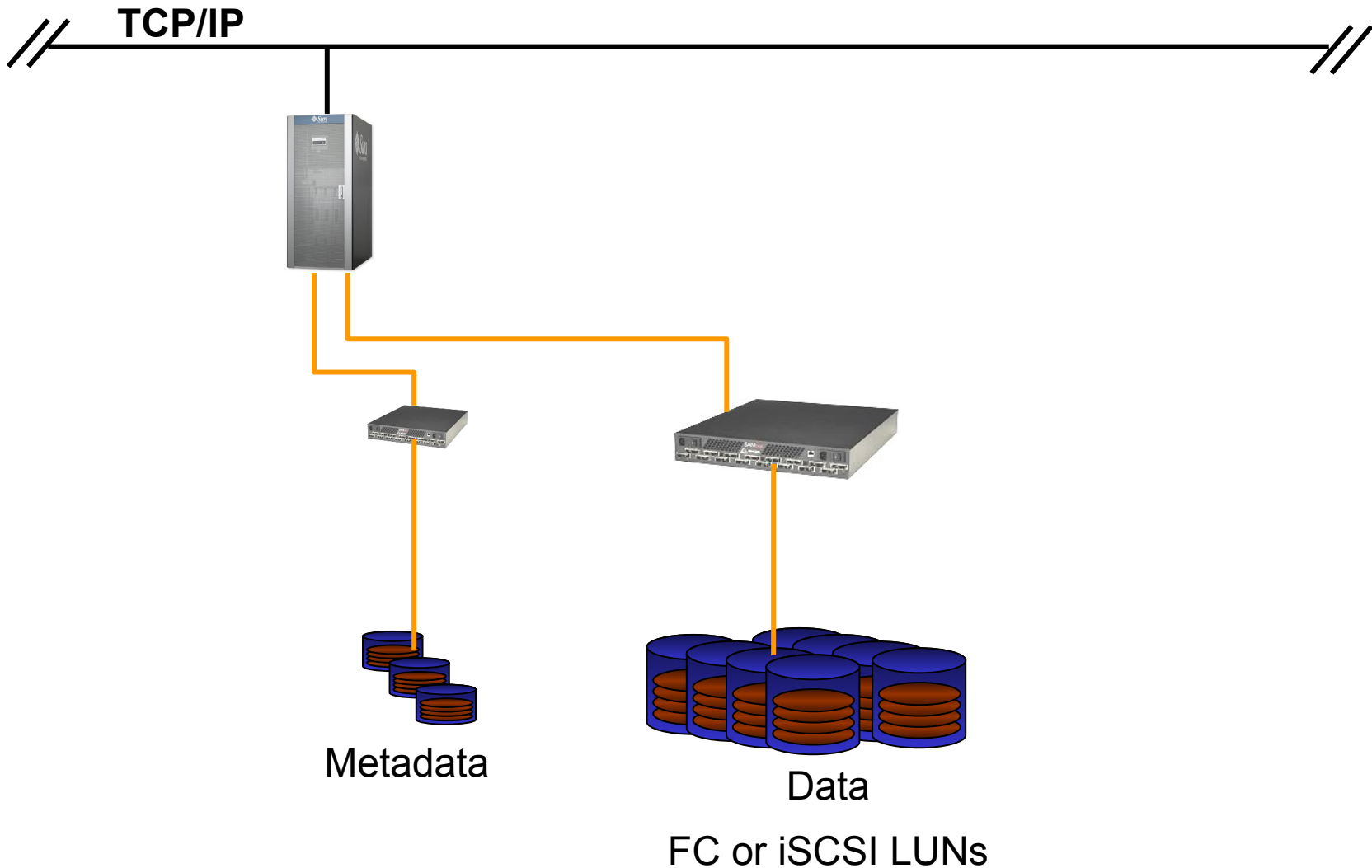
How the Software is Deployed

QFS is the file system foundation for all configurations

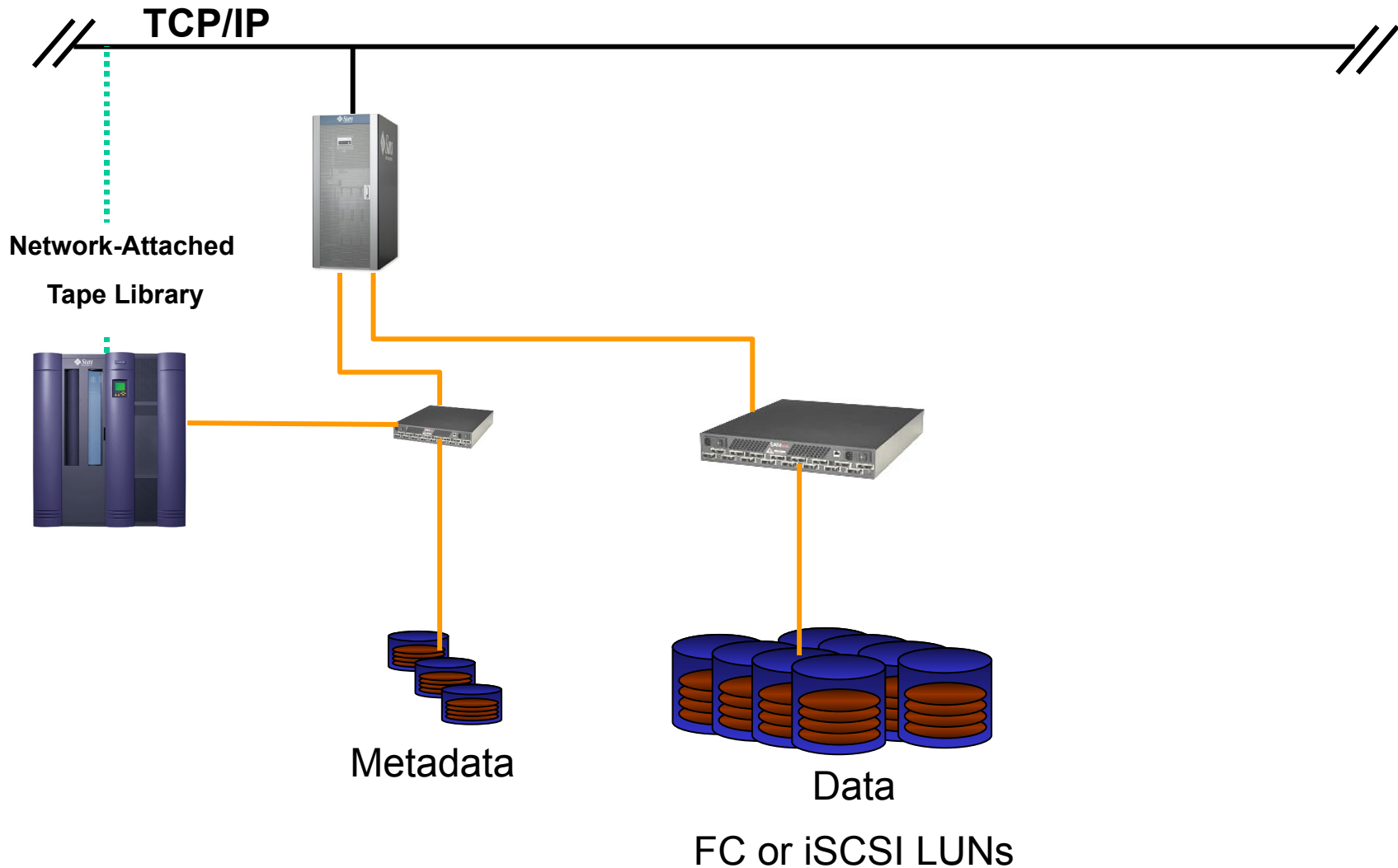


- Archiving File System
 - > Storage Archive Manager (SAM) provides the archive service for local or shared QFS file systems
 - > Remote archiving services can be configured for disaster recovery
- Shared SAN File System
 - > QFS Shared describes the file sharing services provided by the QFS file sharing protocol configured to enable file sharing across multiple nodes
 - > QFS clients can be configured as multi-reader or multi-writer

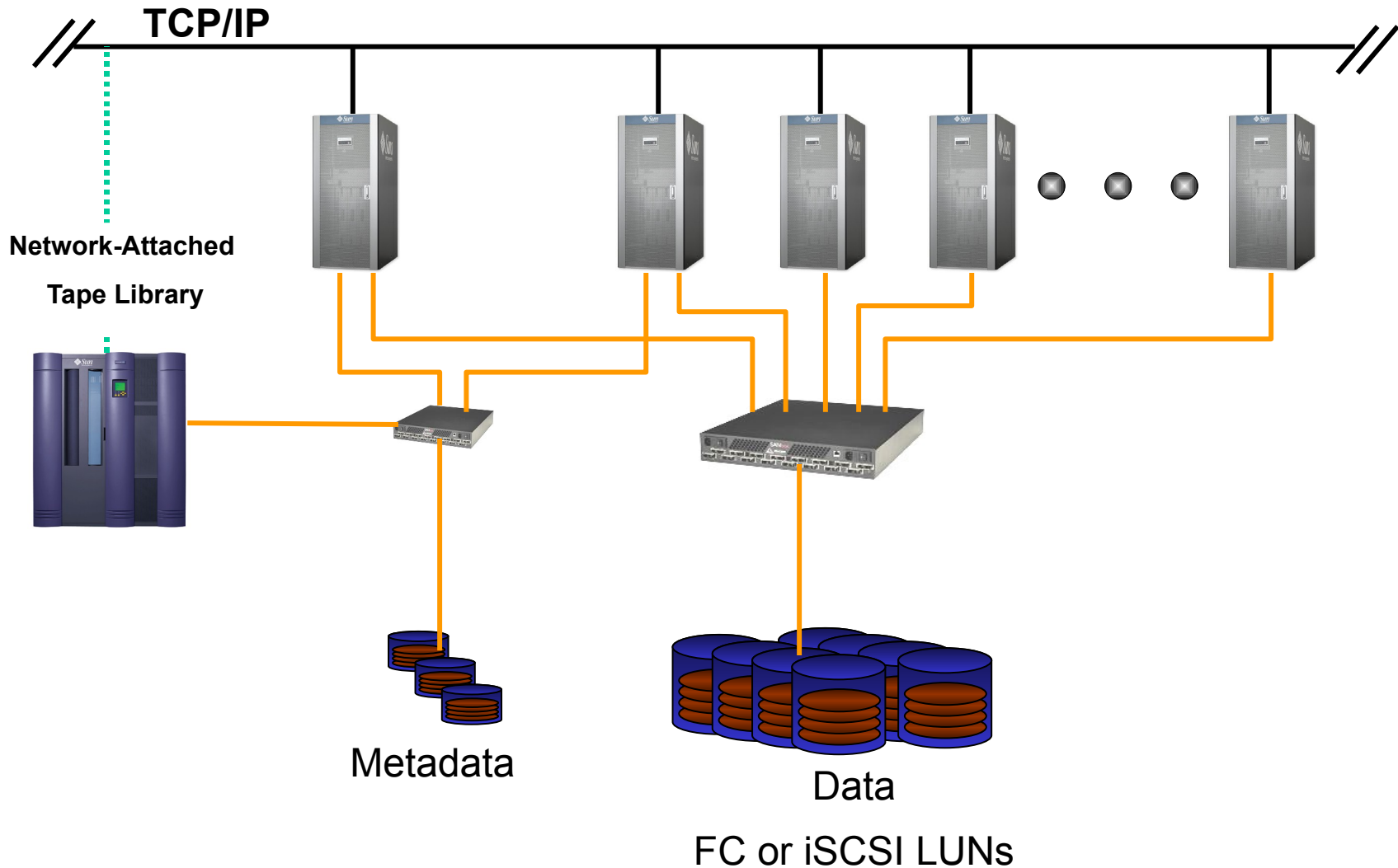
QFS – Standalone File System



SAM-QFS – Archiving File System



Shared SAM-QFS – SAN File System



Open Source Software

Build Community

- Source is open!

<http://opensolaris.org/os/project/samqfs>

<http://blogs.sun.com/samqfs/>

- SAM APIs are open!

> Allow users to manage data in SAM-QFS from within an application program

<http://developers.sun.com/solaris/articles/libsam.html>

- Discussion lists

> Discussion list for general topics or issues

<http://mail.opensolaris.org/mailman/listinfo/sam-qfs-discuss>

> Development alias for specific questions about source, source contributions, or code reviews

<http://mail.opensolaris.org/mailman/listinfo/sam-qfs-dev>



Data Management Challenges

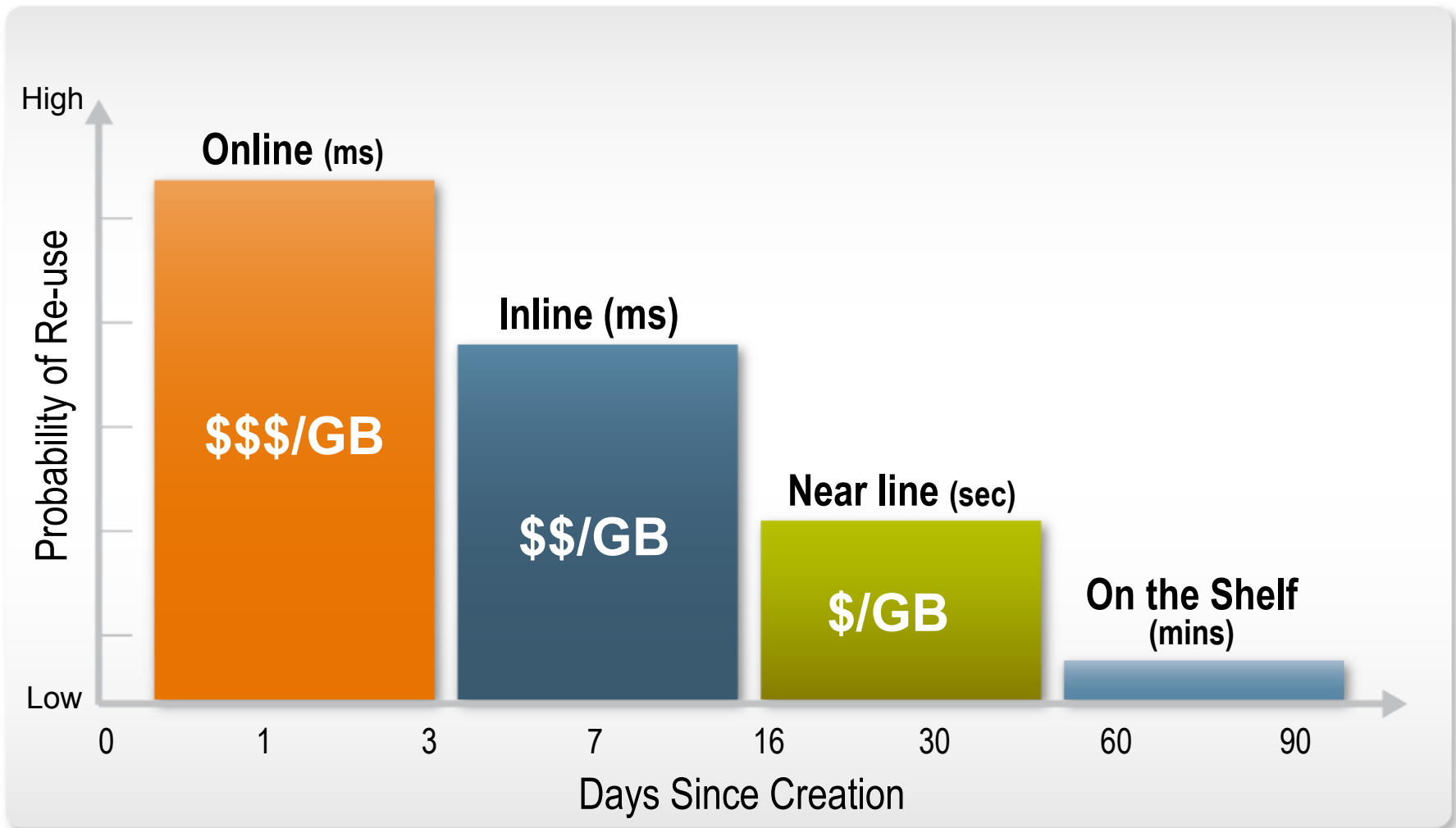
Challenge: Cost Efficient Data Management and Fast Access to Large Volumes of Data

- Budgets remain flat while data growth is exploding
 - > Increasing management costs
- Compliance requirements
- Users require timely access to information throughout its lifecycle

SAM Customer Benefits

- Leverage existing hardware – defer high cost disk purchase
 - > Add (cheaper) storage tiers transparently
- Provide timely protection and timely access to information throughout its lifecycle
- Meet compliance requirements with WORM support
- Eliminate backup window problem & complexity
 - > Reduce operational costs
- Quick disaster recovery for business continuance

Matching Data Value to Storage Tier



Sun Storage Archive Manager (SAM)

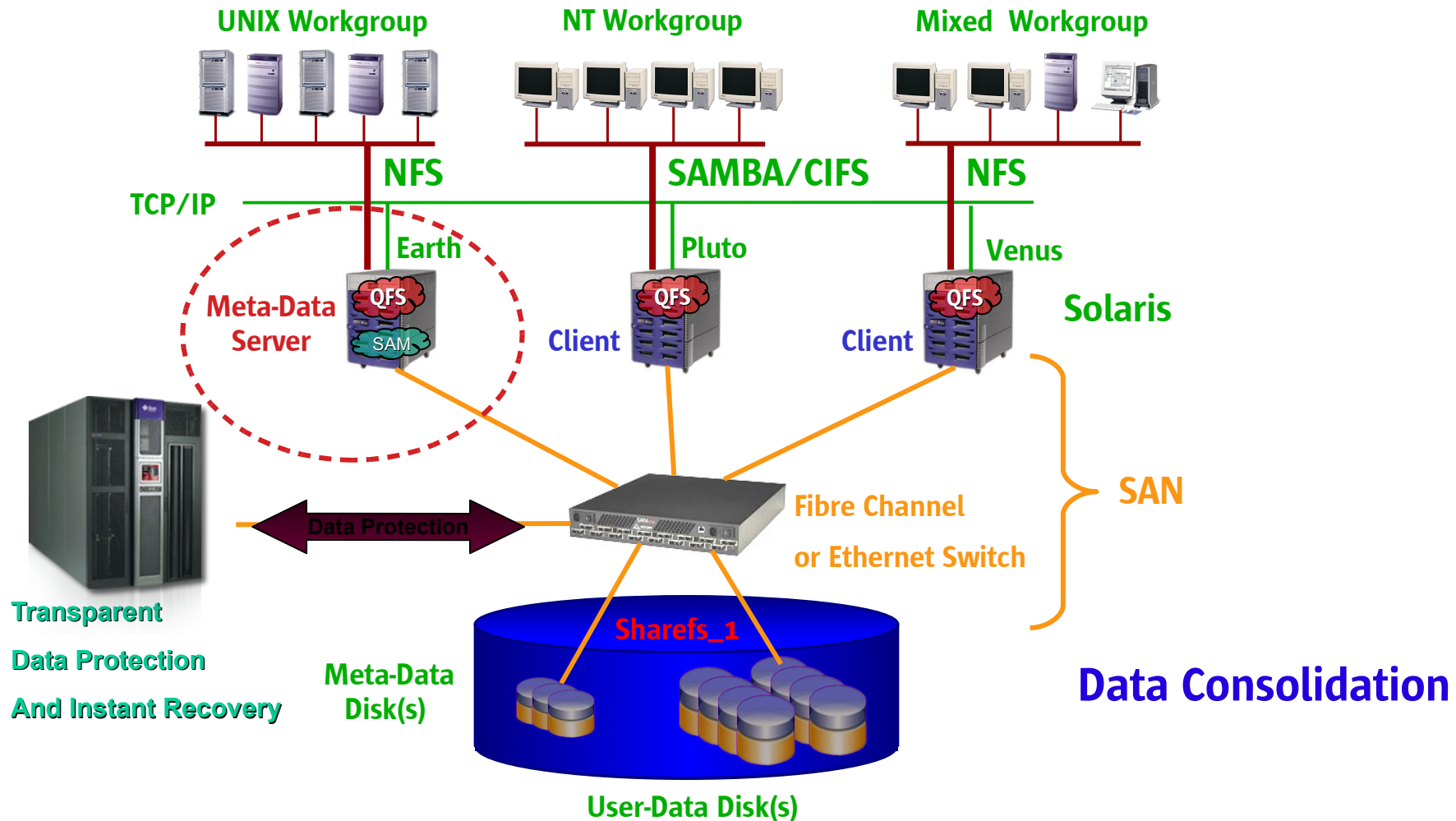
- Policy based archiving
 - > Media can be disk, tape, or optical
 - > Local and remote copies
 - > Classification – path, owner, group, size, wildcard, age, date
- Media format is tar – open format
 - > Small files are put into a tar container so data is streamed at device speeds out to the tape
- Keeps all data available, but not on high cost storage
 - > Archives data across the tiers according to access patterns
- On-demand, transparent file retrieval
- Continuous data protection – no waiting until midnight

Policy Driven SAM Processes

- Transparently **Archive** from disk cache to removable media without operator intervention based on policies
 - Time based archiving
- Manage disk space and **Release** archived files from disk cache based on policies
- Automatically **Stage** released files back to disk cache when accessed
 - Option to pre-stage and option to bypass disk cache
- **Recycle** removable media by repacking media

SAM-QFS Data Consolidation

Integrated Data Management



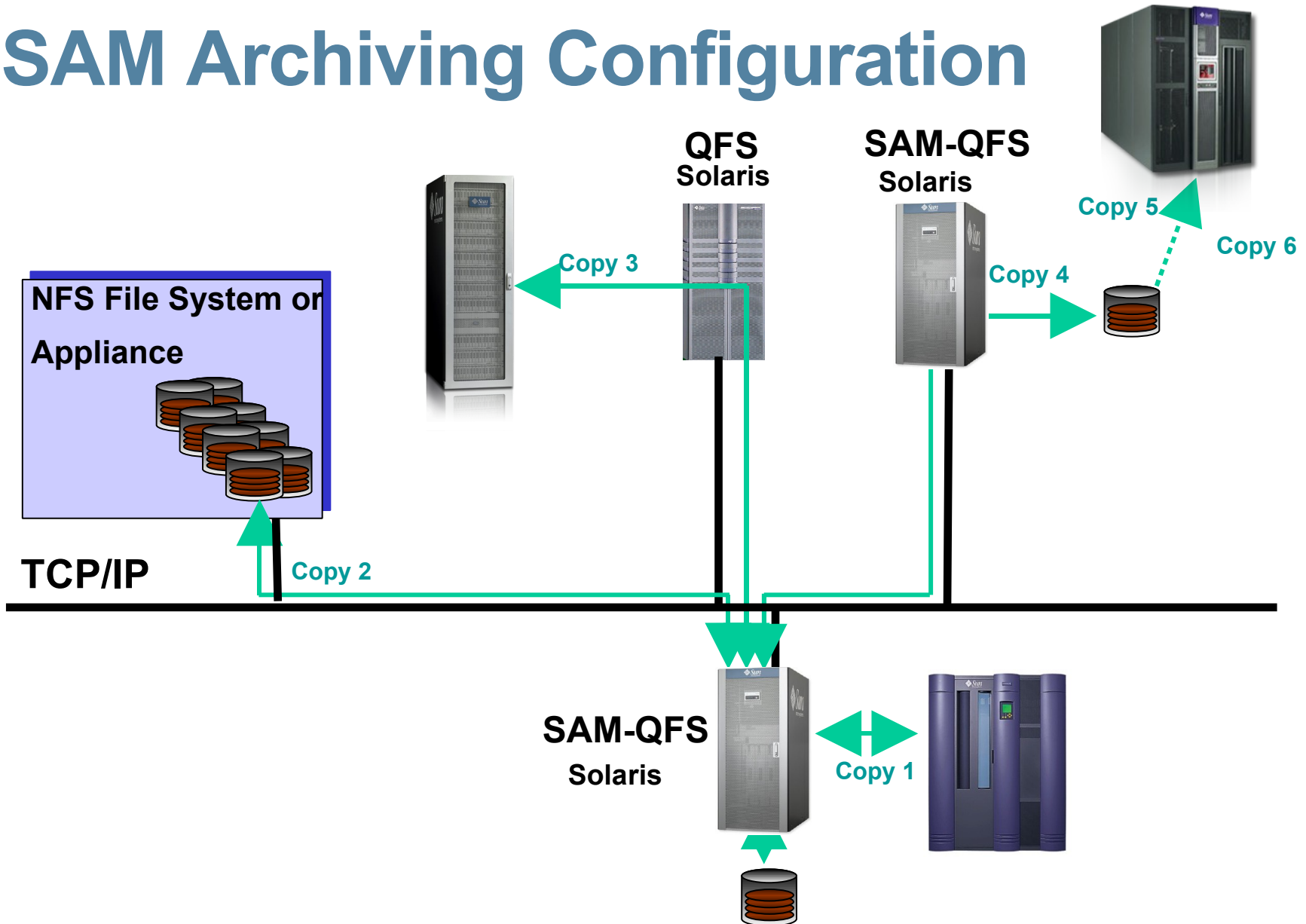
SAM Features Include

- Read behind stage
- Fast access to the first byte
 - > Position to the block and offset to the tar header
- Partial option – leave beginning of file on-line
- Media vaulting; Scratch pools; Shared drive support
- Checksum for archive files and data verification
- LibSAM API interface for commands: archive, stage, release, ssum, etc.
 - > LibSAMRPC for remote procedure calls to LibSAM

SAM Features for Verticals

- Associative Staging – Medical vertical. Doctor views one picture while all related pictures automatically stage
- Associative Archiving – Satellite vertical. Small descriptor files and large data files are co-located on the same media
- Wildcard attribute settings and pre-staging – Pre-press vertical. Load up disk cache and automatically keep the resource forks on disk
- Volume overflow and interchange – Seismic vertical. Facility to move tapes from the ships to the datacenter; read tapes not written by SAM
- Virtually unlimited disk – Video virtual. Disk cache backed by massive nearline and offline tiers of storage

SAM Archiving Configuration



SAM Migration Facility

- Move from foreign HSM to SAM (PS Engagement)
 - > Import metadata into a SAM-QFS file system
 - > Copy foreign HSM data to SAM in the background
 - > Production **up and active** during the migration process
- Migrated German Weather from AMASS to SAM
 - > Moved 10 million files into a SAM-QFS in 8 hours
 - > Successfully migrated 700+ TB of data in 161 days
- Migrated Boeing IT from Veritas to SAM, 81 million files in 6 months
- Migrations for DMF, UniTree, AMASS, and Veritas

SAM's Archives are OPEN

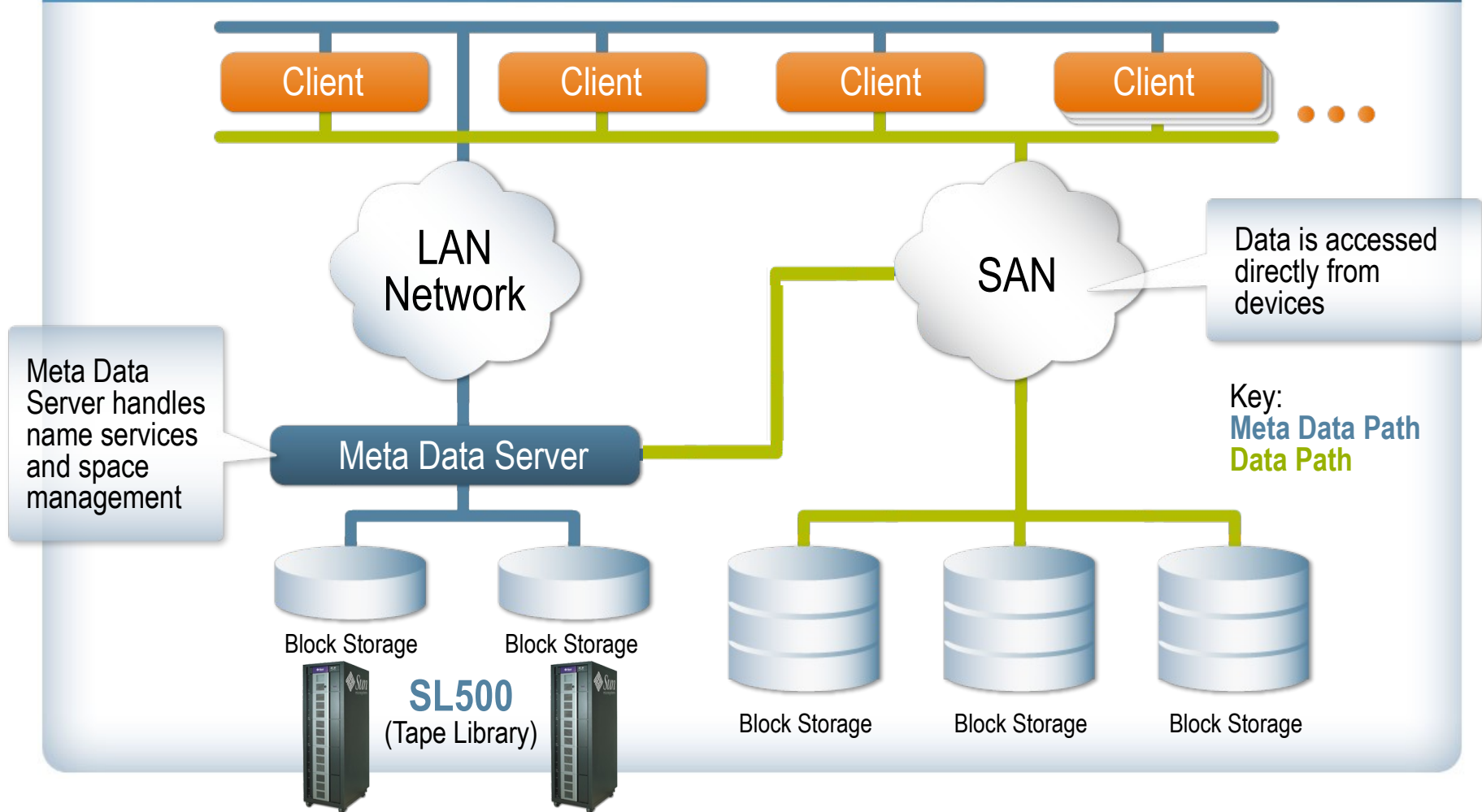
- Media format is open, not proprietary – tar format
 - > Files can be recovered with or without SAM – our media format is open, NOT proprietary
 - > No vendor lock-in
- Metadata about the data is on the archives
 - > If file system metadata is lost, the archives can be recovered with a procedure we call the “Ultimate Disaster Recovery”
- Move data to newer media over time, transparently

Shared QFS File System

- sQFS – SAN file system
 - > Solves data consolidation with multi-host r/w access
 - > Delivers a performance edge over NFS for large files
 - > Takes advantage of the SAN with multiple data paths in contrast to NFS where all data is transferred OTW
- Targets large enterprises, Web, and HPC
 - > Clients run on Solaris (SPARC, Intel, & AMD) & Linux
 - > Linux RedHat and SuSE clients
 - > Metadata server runs on Solaris (SPARC, Intel, & AMD)
 - > A certified shared file system for Oracle RAC
- sQFS configuration supports up to 512 nodes

Shared QFS with SAM

Integrated Data Management



Shared QFS Features

- POSIX file system
- Two file system types, both true 64-bit file systems
 - > Metadata separated from data for streaming I/O
 - > Metadata interspersed with data for random I/O
- Variable block size (DAU) ranging from 8k up to 32MB
 - > Stripe groups – option to group LUNs (parallel I/O)
- Metadata cached on clients
- POSIX ACLs and quotas (user, group, & path)
- Flock fully supported
- Optional WORM functionality for business compliance

QFS Shared File System Benefits

- Data consolidation with SAN file sharing
 - > HBO – 5000 hours of programming to manage
 - > “Provided the scalability to store and manage large files created by program-length video with the performance necessary to meet HBO's demanding throughput goals”
- <http://www.sun.com/customers/storage/hbo.xml>
- Performance and scalability
 - > Near raw I/O performance for streaming I/O and transactional I/O
 - > File system I/O performance scales linearly with the hardware
 - Parallel processing w/ multi-node read/write access
 - Built in automatic & continuous data protection w/SAM

QFS Certified with Solaris Cluster

- Solaris Cluster HA failover support
 - > Voluntary and Involuntary failover
- Solaris Cluster Advanced Edition for Oracle RAC relies on Shared QFS with Solaris Cluster for HA
 - > Oracle certified on 9i, 10g, and 11g

http://www.oracle.com/technology/products/database/clustering/certify/tech_generic_unix_new.html

- Oracle Billing and Revenue Management uses sQFS
 - > “The best known transaction rate seen on the RAC-database can be achieved when the data files are on the QFS file system .”

<http://www.oracle.com/industries/communications/pdfs/oracle-sun-performance-benchmark-wp.pdf>

SAS World Record Performance

- Sun worked with SAS to achieve a second, world-record ETL performance benchmark
 - > SunFire E25k and Solaris 10 OS
- Using 20 Sun StorageTek 6140 arrays using the Sun StorageTek QFS shared file system
- Benchmark handled throughput of 1.7 terabytes in 17 minutes, representing 5.97 TB per hour – a world record based on publish benchmarks in the industry

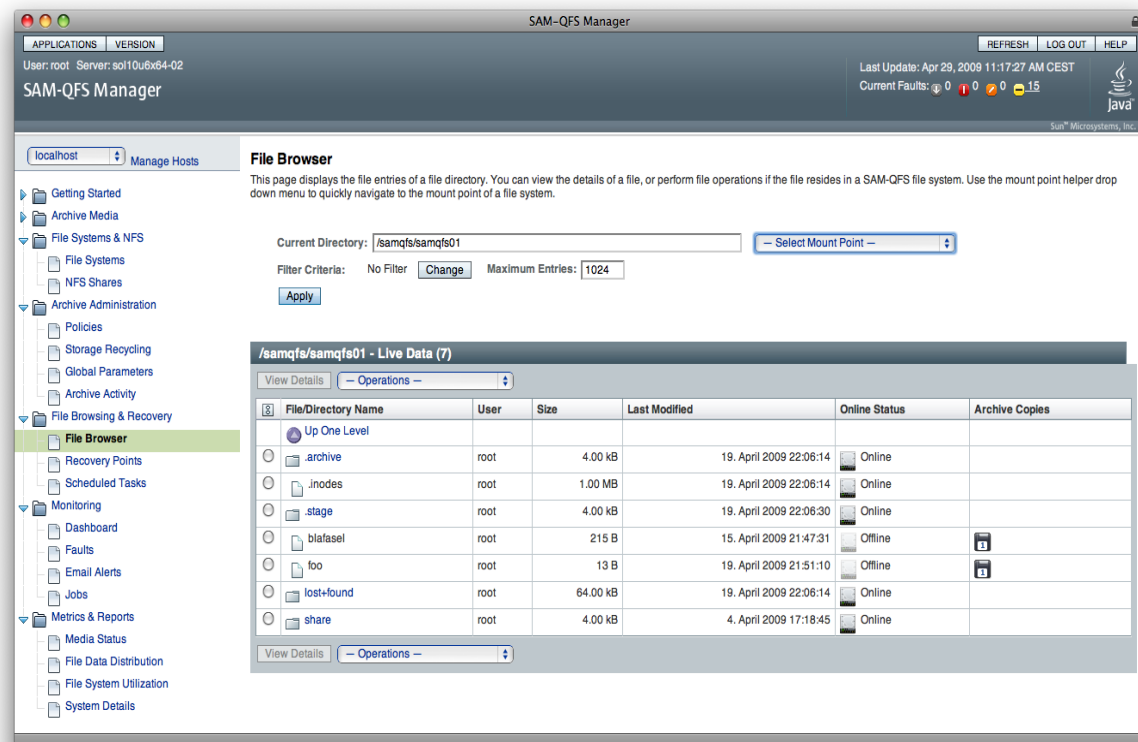
<http://www.sas.com/news/preleases/041707/news9SASSUN.html>

http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9129920&source=NLT_ES

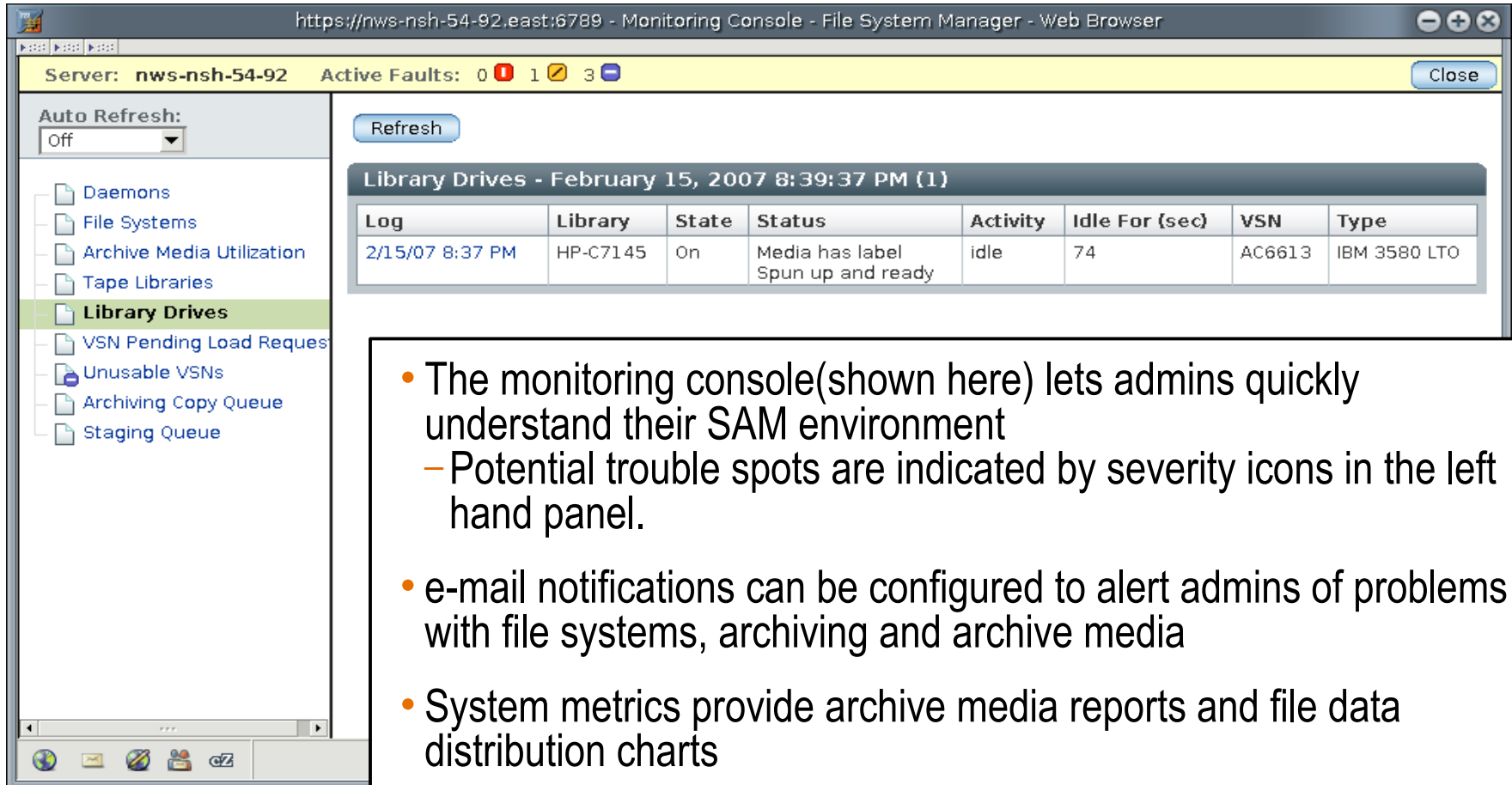
SAM-QFS Management Simplified

Wizard Guided Set-up and Browser-Based User Interface

- Centralized browser-based configuration and management of SAM and QFS on multiple hosts
- Configuration Wizards for
 - > Archive policy
 - > File system creation
 - > Adding tape libraries
- Archive media operations and reporting
- Support for multiple levels of privilege using roles
- Versioning support



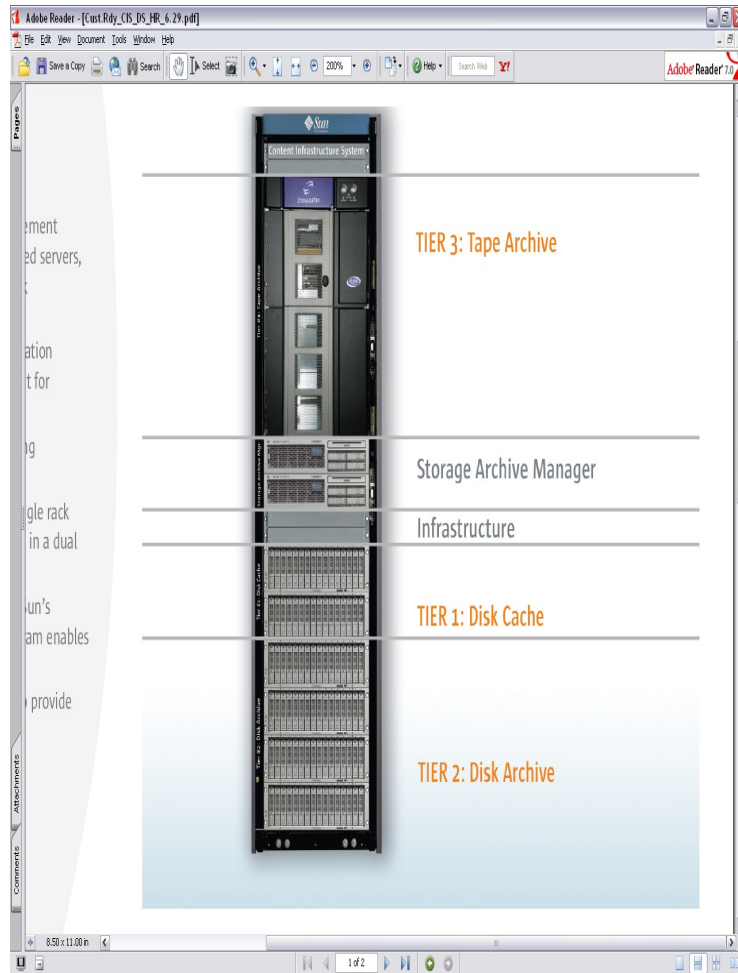
Support for Monitoring SAM and QFS



Log	Library	State	Status	Activity	Idle For {sec}	VSN	Type
2/15/07 8:37 PM	HP-C7145	On	Media has label Spun up and ready	idle	74	AC6613	IBM 3580 LTO

- The monitoring console(shown here) lets admins quickly understand their SAM environment
 - Potential trouble spots are indicated by severity icons in the left hand panel.
- e-mail notifications can be configured to alert admins of problems with file systems, archiving and archive media
- System metrics provide archive media reports and file data distribution charts
- Faults provide a record of adverse conditions that have occurred in the system (including tape alerts)

Infinite Archive System (IAS)



Core Features & Functions

- > Complete archive solution in a rack
 - > Software, Servers, Storage, Services
 - > Simple ordering process
- > Works with existing tape libraries
- > Automated data migration
- > Continuous backup protection

Key Benefits

- > Lower Cap-Ex
 - > **20% - 33% savings***
- > Lower Op-Ex
 - > Simple administration
- > Easy installation

** over list price of separate components*

Sun Solution for SAP Archiving

Customer Challenge

- Proliferation of SAP application data
- Meet regulatory compliance and discovery mandates

Sun's Solution



Results

- Save up to 75% in energy costs*
- Save up to 45% on TCO*
- Improve system performance and reduce time for backup and data access

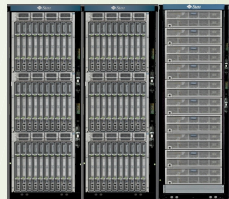
* 3 yr view of 4000 user infrastructure vs. status quo, includes maintenance costs

HPC Storage Solutions

High Bandwidth
Scalable Storage Cluster
with Lustre

Compute Cluster

Metadata
Servers



IB Network



Object Storage Farm

Load



Data Movers

Long-Term
Data Retention
with SAM-QFS

Near Line Archive



SAN

Archive



Home Directories
Tier 1 Archive

TACC Supercomputer Storage

HPC Storage Solutions

Compute Engine Data Cache

- Will scale to over
 - > 72 GB/sec. sustained bandwidth
 - > 1.728 Petabytes of raw capacity
- Configuration includes
 - > 72 SunFire x4500s
 - > Over 3,000 500GB drives
 - > 8 racks



Long-term Retention and Archive

- Will scale to over
 - > 200 Petabytes of near-line
 - > 3.1 Petabytes of on-line
- > Configuration includes
 - > 5 StorageTek SL8500s
 - > 48 StorageTek T10000Bs
 - > 10 StorageTek 6540s
 - > 6 SunFire Metadata servers with SAM-QFS



Seamless Transfer

SAM-QFS Customer Snapshot: Education

KISTI

The Korean Institute of Science and Technology Information (KISTI) is a government-funded institute promoting national competitiveness and provides cutting-edge research in a variety of disciplines.

- **Business Results**

- Achieved rank of 130 in the global Top500 list
- Targeted HPC cluster performance of 300 TFLOPS
- Increased HPC cluster performance by approximately eight times

- **Customer Challenges**

- Maintain its leading position in high performance computing in Asia
- Replace its current supercomputer to improve computational efficiencies
- Increase available computing power for academic research
- Improve price-performance ratio

- **Solution**

- KISTI built its fourth supercomputer “Tachyon” based on the Sun Constellation System. Built on Sun technologies its the first open Petascale computing environment combining ultra-dense high performance computing clusters, networking, storage and software into an integrated system to deliver highly available computational power to the Korean research community.
- SAM-QFS migrates and recalls huge user data to and from tape storage. To accelerate the transfer of stored data on the cluster, KISTI uses the Lustre file system.



Version 5.0
April 2009

5.0 Features

- Solaris 10 released April, 2009
 - > OpenSolaris version to follow with U1
- Online grow of a file system by adding a new LUN – meta, data, or stripe group
- Online shrink of a file system (with & without SAM)
- Rolling upgrades for Shared QFS
- Directory Lookup Performance Improvements

5.0 Features

continued...

- Solaris Feature Integration
 - > Zones Support
 - > Solaris Service Management Facility (SMF)
 - > VPM virtual memory performance improvement for X64
- GUI Enhancements:
 - > Usability study enhancements (including first time configuration checklist)
 - > Online Grow/Shrink, Shared Client On/Off, WORM

5.0 Features

SAM

- Archiver improves performance by changing the worklist from a list of directories to an actual list of modified files
- Stager improves performance by aligning writes to the disk cache
 - > Read/modify/write overhead is eliminated
- SAM sideband MySQL database
 - > Improves samfsdump performance
 - > Enables fast queries, i.e., all files on a VSN

Directory Performance Improvements

- Trust the directory caching
 - > Trust the cache when the entry exists and do not read the entry from disk to verify
- Performance measurements
 - > Rewrite of 500,000 existing files in the same directory is 33% faster; Removal is 800% faster
 - > Postmark test of 1 million files and 100 directories (10,000 files per directory) is 10% faster

SAM MySQL Sideband Database

- File system metadata can be indexed
 - > File system automatically collects metadata (door interface provides fast event-based filesystem notifications)
 - > Events generated by QFS for create, modify, rename, remove, archive, and release/online
- Enables fast criteria-based searching
 - > Date ranges
 - > Media contents
- Reduces administrative costs
 - > Option to generate SAM metadata dump (samfsdump)

5.0 Performance

• PostMark file system benchmark

- ma-mm-mr file system type with 8 meta data devices (mm) and 8 data (mr) devices.
- ms-md file system type with same 8 devices as for the ma-mm-mr runs
- There were 200,000 files in one directory and also did 200,000 transactions.

Times (sec)	SAM-QFS 4.6	SAM-QFS 5.0	Improvement
Creation	62	42	1.5 x
Transaction	4,256	1,060	4.0 x
Deletion	132	83	1.6 x
Overall	4,450	1,185	3.8 x

Times (sec)	SAM-QFS 4.6	SAM-QFS 5.0	Improvement
Creation	62	43	1.4 x
Transaction	3,407	319	10.7 x
Deletion	92	88	1.0 x
Overall	3,561	450	7.9 x

• Additional performance tests

Test	Times (sec)	SAM-QFS 4.6	SAM-QFS 5.0	Improvement
Run with 4 large files and stager set so that if direct I/O is used, it is properly aligned with direct I/O.	Stager wall time	845	528	1.6 x
	Stager CPU time	655	2	330 x
samfs dump after archiving 6 million files (effect of MySQL side band)	wall time	1315	125	11 x
100 million file archive test (4K file size)	wall time	20.97 hr	10.66 hr	1.97 x

Release Strategy

- SAM-QFS 5.0 released on April 28, 2009
 - > Supported on Solaris 10 Update 6
 - > Linux clients
 - > RHEL 4/U4, 4.5; SLES 9/SP2, 10, 10/SP2
- Update releases are planned at ~6 month intervals:
 - > Customer driven feature enhancement
 - > Support for strategic internal projects
 - > Additional device qualifications & bug fixes

Documentation & Training

- Documentation is wiki-based
<http://wikis.sun.com/display/SAMQFSDocs/Home>
- External Training
 - > NWS-4510: Sun (TM) Storage Archive Manager and Sun(TM) QFS 5.0 Administration

5.0u1 Planned Features

- Release of 5.0u1 planned for fall, 2009
- Solaris 10 & Open Solaris (Sparc/AMD/Intel)
- Red Hat Linux 5.3 client support added
- Increased HA support with Solaris Cluster
- Device Qualifications



Customers

SAMQFS Success Stories (Europe)

<http://www.sun.com/customers/index.xml>

- Konrad-Zuse-Zentrum für Informationstechnik (D)
- Deutsches Klimarechenzentrum (D)
- Universitätsklinik Magdeburg (D)
- Kinderspital Zürich (CH)
- Deutsche Rentenversicherung ZR West (D)
- Bundesministerium der Finanzen (D)
- GGP Media (D)
- British Board of Film Classification (UK)

SAM-QFS Customer Snapshot: Government Federal Ministry of Finance, Germany

- **Customer Challenges**

- Develop and implement automated tariff and local customs handling systems (ATLAS) for customs processing
- Provide a secure, scalable, and highly available infrastructure with mirroring capabilities



- **Solution**

- A new three-tier mirroring system architecture with Web browsers leveling the first tier
- Data stored on Sun systems 10 km apart with Sun StorageTek 6540 arrays for disaster recovery
- **Sun QFS file sharing software**

- **Business Results**

- Customs clearance is now completed more accurately and much faster than before
- Exceeded server expectations and delivered solution six months ahead of schedule
- ***“We have a zero failure rate*”**

SAM-QFS Customer Snapshot: Media & Entertainment

British Board of Film Classification

Business Results

- No longer need to pay for costly storage of bulky analog tapes.
- Expects to cut archive costs by more than 90% — from £100,000 to £8,000 a year.
- The British Board of Film Classification (BBFC) is an independent regulator of the film and video industry in the United Kingdom. BBFC is legally required to archive all DVDs and videos classified in the U.K.
 -
- Converted approximately 7,000 analog tapes to digital storage and processing approximately 700 more each week. Valuable recordings are protected and management made easier and more efficient with Sun SAM.
- Before, when another vendor managed the BBFC's external storage vaults, the video tapes had to be manually labeled and retrieved. Now, the BBFC can manage its own library, and automated storage and retrieval helps ease the management burden and reduce error.

Customer Success Story

Digital Content Archiving

- Industry leader in video post-production
- Locations in US and EAME
- Digital Media Environment
 - > Managing massive amounts of shared digital content
 - > View, edit, store uncompressed data between global facilities
- Implemented tiered storage solution from Sun
 - > SAM-QFS, 6540, X4500, SL8500, T10000
- Streamlined digital file-based workflows
 - > Archiving content cost effectively
 - > Generating new revenue streams



SAM-QFS Customer Snapshot: Healthcare

Kinderspital Zürich (Zurich Children's Hospital)

A multidisciplinary medical institution devoted to the treatment of illness in children and infants and to the study of developmental growth of children. It supports a wide range of research facilities.

- **Business Results**

- Increased scalability for future growth
- Enlargement of the main memory
- Flexibility for building new solutions
- Maintain data with constant number of IT employees despite a tenfold increase in the amount of data
- Sun Storage Archive Manager Solution Manages Data and Keeps Hospital Competitive
- The Children's Hospital solution includes Sun StorageTek Storage Archive Manager (SAM) software to classify and manage data over its entire lifecycle.

- **Customer Challenges**

- Deal with flood of data from implementation of new PACS solution
- Safely protect statistical data
- Replace existing SAN solution with scalable storage to support growth

- **Solution**

- Information lifecycle management and archiving solution based on Sun Storage technology for quick access to data generated by Picture Archiving & Communication System (PACS).

-

SAM-QFS Customer Snapshot: Technology

EURIDISS

EURIDISS, formerly GSR, is based in Spain. The company provides security and surveillance services for companies in numerous business sectors such as construction, transportation, and retail.

- **Business Results**

- Expanded business opportunities
- Centralized security operations

- **Customer Challenges**

- Provide customers with alternative surveillance options
- Automate and consolidate security services
- Build a telesurveillance solution that is highly available and secure
- Manage and store an exponential number of surveillance videos

- **Solution**

- EURIDISS uses a telesurveillance system built on Sun servers to provide uninterrupted monitoring and security management from a central location. The videos, which provide a record of events, reside on Sun StorageTek devices.



harriet.coverston@sun.com

