Network Computing's storagepipeline Questionnaire

ENTER
CATEGORY
Comparative Review

Dear Data/Information Lifecycle Management Solution Provider,

The following is a questionnaire for our use in understanding and evaluating the numerous solutions that are collectively called data or information lifecycle management solutions by vendors. Our working definition of these types of solutions is “a software and/or hardware solution for migrating data through the storage infrastructure using automated policies that match data content and/or access characteristics to storage platform capabilities and costs.” If you would like to suggest amendments or revisions to this definition, please make your suggestions as part of this questionnaire.

Please respond as per the cover letter attached to this questionnaire. You can edit this document directly to provide your information.

Usage Scenarios

Since the capabilities of these solutions may vary, we are going to assess their effectiveness in the context of different usage scenarios.

Capacity Utilization Efficiency. These users mostly use a D/ILM solution to place data on platforms or media that are least costly, most capable and best suited to their access characteristics and “inherited” requirements in terms of retention, security, and criticality. Automated data migration may also be seen as a strategy for reducing TCO by reducing or capping administrative staff requirements.

Capacity Allocation Efficiency. These users are primarily interested in D/ILM as a mechanism for sharing capacity in the most efficient way and for eliminating unnecessary replication or junk data – often to defer additional hardware purchases. Automated capacity allocation may also be seen as a strategy for reducing TCO by reducing or capping administrative staff requirements.

Disaster Recovery and Business Continuity. These users seek to use a D/ILM solution to segregate data by its restoration priority in the wake of a disaster and to route data to the appropriate protection process (mirroring or continuous data protection, snapshot or disk-to-disk replication processes or tape backup). These users may also seek to leverage D/ILM to provide input to change management processes designed to keep plans up to date with burgeoning data.

Questions

Development Status and Objectives. Describe your vision of D/ILM and identify what features or functions your platform currently provides and what features or functions are still in development.
Sun’s vision of ILM is not a single box or piece of hardware that users implement in order to manage data. Rather, ILM is an entire set of packages (hardware, software and practices) that represent the aggregate of technology enabling data management. These run from simple media maintenance and backup through complete system managed data placement and enterprise continuity offerings.

Some of the key cornerstones of Sun’s ILM strategy are the StorEdge SAM-FS and QFS software.

- *Sun StorEdge SAM-FS software simplifies data lifecycle management by enabling business policies that automatically protect, migrate, backup and restore data while providing centralized administration across all forms of storage media and locations.*
- *Sun StorEdge QFS software enables the consolidation of data into <u>one</u> shared file system that may be accessed concurrently by many hosts for ease of management and better return on storage investments. Multiple users can then share the same files, reducing overall storage space consumption. The software also eliminates the need to mount and re-synchronize file systems when an environment needs to scale to multiple terabytes, simplifying administration. Additionally, Sun StorEdge QFS software helps dramatically increase the span of control over a growing data environment.*

SAM-FS and QFS may be combined to offer a complete shared, space managed data environment.

Categorizing Sun's ILM offerings in terms of the usage scenarios you have outlined above:

- *<u>Capacity Utilization Efficiency</u>- With the understanding that organizations are currently storing huge amounts of data, of which only about five percent is typically accessed on a frequent basis, there is a need for the ability to move data to different types of medium (not always tape) with little/no human interaction. Sun's SAM-FS software provides just this data prioritization by managing and placing data according to its business value. The software enables dynamic archiving, reduced backup windows, and fast recovery to help enhance productivity and improve resource utilization. Sun StorEdge SAM-FS software consolidates the best archiving and backup methodologies in a high-performance file system with unmatched scalability. The software can replace (or augment) traditional backups to improve storage resource utilization for applications where data needs to be available continuously and quickly restored in the event of a business disruption. Users can set automatic archiving and migration policies to determine when, where, and how data is stored, enabling the management of large volumes of data cost-effectively. Metadata archiving and read-behind features help recovery from business disruptions in minutes or hours, as opposed to days, and let users begin reading files even before they are fully restored. Enhanced policy-based administration and security features let users set quotas and define access control lists (ACLs) to control space consumption and data access.*

- *<u>Capacity Allocation Efficiency</u>- Sun StorEdge QFS software provides maximum scalability, performance, and throughput for the most data-intensive applications. Complex datasets no longer need to be spread across multiple file systems, enabling them to be kept in a single, scalable file system, dramatically reducing administrative overhead with no performance degradation. Multiple servers, applications and users can share the same files and volumes on the network, which dramatically improves productivity and scalability from the workgroup through the data center. This high-performance file system can help optimize the data path for end-to-end infrastructures, enabling the convergence of network-attached storage (NAS) and storage area networks (SANs).*

- *<u>Disaster Recovery and Business Continuity</u>- As disaster recovery and business continuity have been a corporate mandate for many years, Sun does not see this as a specific ILM solution, but acknowledges that it falls under the ILM umbrella of technologies. Sun currently provides a comprehensive set of solutions that address this need from simple back-up to the Sun Enterprise Continuity solution that provides for geographically dispersed, active/active clusters that can reduce the disaster restoral window from days and hours to minutes.*

To further bring to life its ILM strategy, Sun is currently developing Integrated Object Based Storage technology as well as applications that provide secure, guaranteed data deletions (especially important for tape archiving environments) and tamper-proof data.

Development Partners. List the vendors with whom you are working to deliver your D/ILM solution functionality.
Currently Sun works with the following ISVs/data management partners, including: Asigra, Imagemark, Veritas, IXOS, Altana, Legato/Veritas, Magicstore (Siemens), Butler (Net&Storage solutions), Documentum, and TimeNavigator (aTempo). Please note: this is not a complete list, rather a sampling of the sorts of key ISV's with whom Sun partners.

Data Naming Scheme. What is your method for identifying data storage requirements or characteristics and for using those requirements or characteristics in building policies?

Sun StorEdge SAM-FS and QFS file systems allow users to determine and set business rules that allow data to automatically migrate to the appropriate storage medium, based on importance and access sensitivity. Some of the business parameters that may be set include the following: Size, Name, Extent, Directory Location (or subdirectory), Create Date, Last Modified Date, Last Touched Date, etc. Based on these parameters, multiple sets of actions can be defined for SAM-FS to take. For example a policy rule might look like the following: "Any data in a users home directory that has not been accessed over 30 days should be migrated to tier2 (Less costly) storage, however exclude any executables (programs) and/or system created data, and any migrated data that has not been accessed in over 1 year should be archived to tape."

Access Frequency. Identify the mechanism by which your solution identifies how frequently specific files or datasets are being accessed and whether and how access frequency is used in migration policies.
Sun's SAM-FS software keeps track of the access to and manages placement of data based on pre-set business rules that are established by the user (see above example). All information is logged and available through published API's to anyone. Additionally, access frequency is used in migration policies to determine the type of storage media that best suits the data. As shown above, SAM/FS can implement far more sophisticated migration policies compared to just looking at access times and frequencies.

Storage Platform Characterization. Explain the mechanism that your solution provides for characterizing the performance capabilities and costs of specific hardware platforms for data storage so that this information can be used to target the appropriate storage platforms as destinations for automatically migrated data.
Sun's SAM-FS software allows users to architect and set their own data migration policies that characterize data and send it to the appropriate destinations. Costs and specific hardware platforms are determined by the users specific needs. SAM-FS provides the ability to define "Storage Tiers" that represent different types and classes of storage. Tier-1 typically would be very high performance, highly available storage. Tier-2 typically would be enterprise class storage; perhaps with SATA drives to reduce cost, and Tier-3 might be a near-line tape library. SAM-FS transparently migrates data between these three tiers based on the rule-sets implemented by the administrator. The actual location of the data is transparent to the using application, so no code or procedure changes are required on the part of the user.

End of Useful Life. Explain how your platform facilitates the automated removal and clean up of data that has outlived its useful life and restoration of freed capacity for use by applications.
Based on the users pre-set requirements data can be assessed, and if appropriate removed, based on age, last read, last accessed and last modified. This can entail optionally creating an offline archive copy, deleting all references to the specific data within the managed data domain, or other policies that may be appropriate to a specific installation. See the rules set above. The "secure delete function," which is currently under development, will render data stored on near-line libraries and data stored on offline, archived tape to be rendered non-accessible when it's "life clock" expires. For offline data, this can be accomplished without retrieving the media and rewriting it.

Policy Articulation. Describe how policies are created and how they are applied to existing data.

All policies, as described above, can be applied to (and modified) to both existing data and new data. For instance, at a major automotive company, SAM-FS archiving policies were applied to a 60,000 user email system and were able to migrate over 80 percent of the stored email (with attachments) to more cost effective storage devices over a period of just a few weeks.

SAM-FS provides a simple user interface where the user can define the business rules, define the data pools to which they apply and define the frequency of data inspection. Once the rule is defined, it becomes effective immediately and data migration will begin.

Device Support. What storage devices does your product support? Are their any proprietary devices (controllers, arrays, HBAs, switches, SAN topology, virtualization products, etc.) that are required for your solution to work?

The following products and features are supported with Sun's SAM-FS and QFS software: SPARC architecture servers and workstations, all Sun supported SCSI and FC HBA's, all Sun supported FC switches, all Sun RAID arrays, all Sun hard disk drives (HDDs), Solaris 7, 8, and 9 (Shared QFS requires Solaris 8 Update 4 or later software), Solstice DiskSuite, Veritas Volume Manager (VxVM), Solstice Backup to a SAM-QFS file system, Veritas NetBackup to a SAM-QFS file system, IBM Tivoli SANergy, Sun Cluster (via Sun Professional Services), Veritas Cluster Disk quotas on a QFS file system, and shared QFS metadata server switch-over (samsharefs-s), plus additional tape libraries (including most popular third-party libraries), removable media devices, jukeboxes, and optical devices.

Resource Consumption. How much bandwidth and server CPU "overhead" is introduced by your solution (e.g., to support polling processes, migration processes, agent processing, access frequency counting, etc.).

Resource consumption for this solution is totally based on application and usage and varies per user. SAM-FS allows the user to control the scheduling frequency of SAM-FS's various components to fine-tune the system to meet the specific business needs of the enterprise.

Flexibility. Does your solution have multi-vendor support? Can data, once integrated into your D/ILM scheme be migrated readily between your solution and other solutions in this space? Is you solution interoperable with other solutions in this space?

Data is stored in an open format "tarball" and can be read by anyone who can "untar" a file. This includes all UNIX and LINUX based platforms. Additionally, a NFC and/or CIFS interface is provided so virtually ANY platform can access SAM-FS managed data. One of the more common uses of SAM-FS is to backup large WIN2K server farms.

Speeds and Feeds. How should a prospective customer compare the relative performance of competitive solutions in this space? What are the appropriate performance measures or metrics to use in evaluating competitive solutions?

Sun's solution can be scaled to meet customer requirements for bandwidth, IOPS and capacity. A small installation might require 2 V240 servers to operate SAM-FS, and deliver a moderate amount of performance. Fixed content document archiving might be a good candidate for this. At the other extreme, a major oil company has over 80,000 shotlines under SAM-FS management and uses dual SUN E10K processors that can deliver over 1 GB of sustained data a second. The San Diego Super Computer center uses multiple Sun F15K processors to deliver in excess of 3.5 GB/Sec. of bandwidth.

Interconnect Support. List the network or fabric interconnects that you support. Is there an optimal interconnect for data movement for your solution? What about for management?

Any IP-based wire (Ethernet, Gig-E, ATM, SONET, Fibre Channel, etc. can be used for interconnect. Management should be separated (physically and logically) from the data transfer fabric, and is supported via Ethernet.)

Protocol Support. List the storage networking technologies supported, including FCP, Ethernet, iSCSI, FCIP, iFCP, Parallel SCSI, SAS, and network file system protocols (NFS, CIFS, HTTP, DAFS).
The only protocols NOT currently supported by the solution are iSCSI and DAFS.

Data Type Support. List the data types supported by your solution, including file types and file systems, databases, and hybrids such as email. Comment on the granularity of your D/ILM solution: will it support the migration of database components or subsets, subsets of email files, etc.?
The only data management or data ingest systems not supported by this solutions are those that require raw disk I/O (e.g. no file system)

Cost. List the cost of your solution or provide some means for calculating cost for a specific environment.
There is no formula for calculating the cost of an ILM environment, as cost is dependent on the size of the environment implemented, the performance levels required, etc. Cost will vary per user. However a pre-configured solution could range from 1.5C/MB for the smallest environment to approximately .5C/MG or less for the largest environments. Many many variables go into designing a cost effective ILM infrastructure, so the final cost points will, by definition, vary.

Standards. List any relevant open standards upon which your product is built.
Sun's ILM strategy is built on the following standards: NFS, CIFS, SMI-S, HTTP, FTP, RPC, MPI, SNMP, FCIP and FCP.

Other features. List other pertinent aspects of your solution.
Suns solution is extremely scalable, from a few TBs to 252 TB per addressable file system. Note that up to 255 file systems may be defined, each of which could be 252TB in size.

Miscellany: Please note any additional information that you think would be worthwhile for prospective customers to consider about your solution or other solutions in this space.

Sun has over 2000 licensed installations of SAM-FS worldwide and has probably more experience in delivering managed data solutions than any other systems vendor. It is a core expertise of Sun's vs. certain other offerings that appear to be "marketing bundles" put together to address the recent interest in ILM type solutions.