Pund-IT

PUND-IT, INC.

Weekly Review

OCTOBER 19, 2005

Active Archiving Is Not an Oxymoron

By David G. Hill

With all the continuing talk about information lifecycle management (ILM), enterprises are coming to realize that they can manage their fixed content data differently than they have in the past. In the simplest sense, data should be put into two buckets. The first is for active data that is either changing or presumed to be changeable, while the second is for fixed data that is unlikely to change. Historically, many companies have stored both forms of data together on high performance disk arrays until they either delete data that is no longer useful or send it to the deep Gulag of archived information that is unlikely to ever see the light of day again.

Why should companies bother migrating data from the active changeable bucket to a fixed data bucket (which we can also call the active archiving bucket)? The active archiving bucket typically requires less expensive tiers of storage, but the falling cost of disk drives has lessened the importance of this issue. The real motivating factor for embracing active archiving is improved manageability and enhanced service levels. This can lead to benefits including, in the active changeable buckets, significantly reduced time for back-ups, faster restores, and reduction or elimination of out-of-space conditions, and, in the active archiving buckets, simpler data protection procedures, more effective use of disk space, streamlined data retention policies, and the ability to meet regulatory compliance requirements more efficiently.

Active Trade-Offs

Yet there are also tradeoffs. Many active archiving solutions require the purchase of new arrays, although an underutilized older array might provide an alternative. However, the biggest problem for most businesses is in how to plan for, acquire, and implement new policies, procedures, and technologies for tiered storage infrastructures. Service level requirements, user functionality requirements, and user training requirements all have to be taken into account for the active archive. In addition, both IT organizations and end users, have to be involved in the process. For example, data classification policies require end users and IT administrators to reach a consensus on what will work. Although IT has responsibility for data migration, end user buy-in is critical to the success of this process.

Why would an enterprise take on the coordination and implementation challenges of building an active archive? First, the head-in-the-sand status quo approach promises to become increasingly stifling as data growth continues. In addition, alternatives such as deep archiving, where production information is written to tape and then deleted from the original disk, can be difficult processes in that getting this data back for any ad hoc analysis is very difficult. So the critical issue concerning building an active archive is really not why, but when — and waiting will not make the problem get any easier.

CAS and Beyond

A term that is often associated with active archiving is content-addressable storage (CAS), which is often used in the same breath as object-based storage. Object-based storage, in this sense, is said to differ from traditional file-based or block-based storage in that files or data blocks are stored as individual objects, each with a unique digital signature. If any change is

made to an object, a unique new global identifier is created. The identification process eases data retention processes such as automatically deleting a document on a prescribed date as well as facilitating the compliance process. Moreover, unique identifiers facilitate single instancing since unnecessary copies can be easily detected (since they have the same identifier) and deleted to more efficiently utilize disk space. Finally, object-based storage often relies on an architecture called a redundant array of independent nodes (RAIN) which is designed to enable objects to move throughout a disk storage system in a location-independent manner without loss of data protection or reduction of service levels. This eases the burden of managing the storage as there are no zoning, LUN binding, or other traditional storage administration tasks to perform.

Yet CAS is only one approach to active archiving. Non-CAS vendors may offer different strategies for providing fixed content archiving functionality. And since an appliance is a black box in which both inputs and outputs are well-defined, an enterprise should hear out the full story before making a decision on the best approach. In truth, there are typically two classes of vendors — the big boys and smaller companies that would either like to emulate them or become partners in their success. The big boys in archiving are the major storage companies who are household words within the IT community — EMC, HP, IBM, NetApp, and Sun/StorageTek. Three smaller challengers are Archivas, Nexsan, and Permabit.

Table 1: A Sampler of Active Archiving Product Suppliers

Vendor	Product	Product Focus	Technology Foundation
Majors			
EMC	Centera	CAS-based appliance that stores all types of fixed content and has specific editions, such as a compliance edition	CAS-specific appliance for online access to fixed content that is highly scalable and can support a variety of enterprise data
HP	HP StorageWorks Reference Information Storage System (RISS)	Integration to provide all-in-one content storage. Application aware, including MS Exchange, Oracle, SAP, and others	Reference Information Building Blocks called “Smart Storage Cells” Grid computing active archiving
IBM	DR550 with DB2 Content Manager	Managing the retention of non-erasable and non-rewritable data in an archiving repository	The heart of the DR550 is IBM Tivoli Storage Manager for Data Retention. This software is embedded on an IBM eServer P5 520 using POWER5 processors
NetApp	NearStore	“Nearline” storage that can be used for both archiving and data protection (such as backup and restore)	Combines its well-proven Data ONTAP operating system with ATA drives
Sun/ StorageTek	Content Infrastructure System (CIS) (Sun) IntelliStore (STK)	CIS — fully assembled and configured system to ease adoption pains IntelliStore — management of multiple applications in the same box	CIS — object-based archiving IntelliStore — “intelligent archiving” with policy management and data classification software

Minors			
Archivas	Archivas Cluster (ArC)	Archiving focus is on software and partnering with complementary software providers	File focus both semi-structured data (such as e-mails) and unstructured data (such as audio and video)
Nexsan	Assureon	CAS-approach with time stamping to provide proof of creation, encryption, and high-availability clustered SATA-based architecture	Used the acquisition of Evertrust and its AEStore software for life-cycle management of fixed content data
Permabit	Permeon Compliance Store	Long-term retention of both fixed content and compliance data	Software focus with modules for retention policy, high availability, and replication.

Source: Mesabi Group October 2005

EMC and NetApp have maintained substantial shares in this emerging market for some time, but the other major vendors — HP, IBM, and Sun/StorageTek — are starting to take aim. Among the younger challenges, Nexsan can build a strong case for being able to build upon its success in selling SATA-based systems. Archivas and Permabit have to be able to leverage the fact that they have working solutions now. In addition, there are emerging specialists who focus on one aspect of the market. For example, ZANTAZ focuses primarily on the use of active archiving for compliance.

Mission Accomplished?

The move to active archiving offers potential benefits for both the array from which active data was removed and for the active archive that received the removed data. On an active production array, significantly reducing the storage also reduces the time needed for full backups and restore. In addition, businesses could also see performance improvements about through disk defragmentation. On the active archiving array, data retention policies (including those aimed at compliance) can be instituted more effectively than on an active changeable array. In addition, storage space can be saved through single instancing and data can be protected easier using one-time replication techniques instead of using traditional backup/restore software.

Overall, the active archiving market is likely to heat up in the coming year as existing suppliers enhance their product offerings and other contenders make an entry. In this complex and evolving market, IT organizations should consider all their options and, if they have not already done so, consider how the might best employ active archiving solutions to achieve their strategic business goals.

David G. Hill is principal of the Mesabi Group (www.mesabigroup.com). The Mesabi Group focuses on the revolutions in Storage Networking and Storage Management, and helps clients make the best and most efficient use of information for business value.