Get to the Core of
Virtualization

ZL Technologies: A Unified Approach To Active Archiving

Posted by David Hill on February 25, 2011

Even though the "cloud" occupies much of the dialogue about transformative processes in IT, the place of archiving in that transformation is still poorly understood by many IT organizations and needs to attract more attention. Therefore, we should welcome any light that can be shed on the subject, and ZL Technologies' unified archiving approach is particularly luminous. Let's examine what the company is doing, but first, we should discuss the need for active archiving in general.

Frankly, for many IT managers, archiving is not a top of mind issue. Operationally, IT tends to do today what it did yesterday. If a new problem arises, such as how to provide the information required for eDiscovery, the tendency is to hive off the necessary data to a separate copy that can be managed for that express purpose. That way, the IT operational production train can continue to run uninterrupted.

However, that train is headed down tracks that are more and more likely to break down (even though expensive fixes may keep it from total derailment). From an IT storage management perspective, explosive data growth creates issues from budget bloat to how to do weekly backups in the time allotted. And for businesses, a onetime quick fix is inappropriate for critical processes like eDiscovery.

Compliance is another complex issue that requires careful thinking, planning and management controls on the part of IT. On top of this, add the issue of data retention. Keeping information important to organizational and regulatory processes is obviously critical, but so is getting rid of data that has no further legal or business use, resulting in a reduction of fat storage and eliminating potential legal exposures. It must be done properly, however, as eliminating data that should not have been destroyed can have serious negative legal or business consequences. To work correctly, these issues all require effective information governance that manages the data through all the stages of its life-cycle.

And that is where active archiving comes in, providing an overarching software management platform in which information governance processes, such as single policy management, can do their thing. Yet achieving this requires IT to create a new operational train (active archiving) with its own separate tracks that connect to those for discreet processes like data migration and ongoing active production operations.

Yes, the "cloud", which so many love to talk about, is the key to transforming the computing infrastructure to get to true IT-as-a-Service and, of course, virtualization is a necessary ingredient. However, that transformation also requires changes in internal IT processes as well. Active archiving is an integral part of that change, because business as usual is not an option. Moreover, there don't seem to be any feasible alternatives.

So let's say that we create an active archive. What does that mean and what are some of the implications? By definition, an active archive contains the production copy of all fixed content data. While the application that created or originally ingested the information (such as an e-mail sent or received) may still have read-access, it

loses the right to delete or alter the information. Why? If that application is allowed to alter information, it erodes the consistency and validity of the active archive. If that were the case, from a legal perspective no chain of custody would exist and the data could not be used as evidence in litigation.

By definition, an active archive is data-centric instead of application-centric. It may contain many types of data, and, typically, there is no one application that created or originally used all of the information. Though these applications have lost their power to manage the lifecycle of the data any further, there is an "application" in the form of an overarching software platform that manages disparate data under its wing. Unlike a single-purpose application, no matter how sophisticated and functionally-rich, this umbrella software has to cover a wide range of requirements with deep functionality that is not only analytically sophisticated and broadly functional but must also meet system requirements, such as scalability and performance.

Even though an active archive is data-centric, the software package defines whether or not the active archive will achieve the full measure of benefits for which it was built. In addition, content-based information (typically filed-based) archiving and database archiving are usually separate because they are typically handled by different software packages. While database archiving is obviously very important to most companies and many business processes, fixed content file-based information tends to take up the bulk of storage today and is the most important source of information related to eDiscovery requests.

Unified Archive, ZL Technologies' overarching software for active archives, exemplifies what is needed as the software cement for an active archive. Unified Archive manages the capture and ingestion of a wide range of types of content-based information. Of course, e-mail oriented applications, such as Microsoft Exchange and Lotus Domino, are primary data sources but Unified Archive is much more than an e-mail archiving solution.

Basically, ZL Technologies' solution captures all (or nearly all) of the user-created data that an enterprise needs in an archive. That includes information from IT-managed servers, such as Microsoft's Office and SharePoint Server, and related files systems. Unified Archive also incorporates data from sources that an enterprise may be legally required to capture for possible litigation, but which IT typically does not manage, such as instant messages and Blackberry communications.

Note that we used the term "semi-structured" instead of "unstructured." Semi-structured data is often classified as unstructured data, but there is an important distinction: true unstructured data is bit-mapped information, such as photographs or videos. Without sophisticated analytic software, such as for-pattern recognition, unstructured data can only be perceived and analyzed by human senses, such as sight or hearing. Unstructured data can be electronically managed by its file metadata, but this offers very limited functionality. In any event, unstructured data is unlikely to be used in eDiscovery in very many cases.

The same is not true for semi-structured data. The key is that, in addition to its metadata, the content of semi-structured data can be searched with numerous tools. The ability to search (and hence to classify) such content-based data is mandatory for eDiscovery and other purposes. The bottom line is that to be fully effective and deliver on its promises, an archive has to capture as much relevant semi-structured information as possible.

Hidden here are two big requirements. The first is that the archive must scale to handle all the data needed for specific processes, and that it is able to perform any and all related tasks, for example, to classify and search information in a timely manner. Both these cases are areas where active archiving solutions may stumble.

To avoid these problems, ZL Technologies states that it has built Unified Archive at carrier-class, not just enterprise-class, strength. What does that mean? Enterprise-class infers robustness and reliability, which is fine for many purposes. But "carrier-class" relates to the requirements of telecommunications companies, whose humongous volumes of data and performance requirements go far beyond what is required generally of average enterprises to large scale multi-national enterprises that operate data centers on the level of communications carriers. The implication here is that if Unified Archiving is good enough for a telco; it is good enough for the largest enterprise.

For example, Unified Archive scales to handle billions of documents. So what, you say? After all, many storage arrays can now store petabytes of data. But volume is not the issue that breaks down a large archive. The problem, in addition to the daily ingestion rate, is how long it takes to index and, more importantly, to search all those documents. ZL Technologies claims a major differentiator through its use of massively parallel processing (MPP) technologies, which it claims reduce search times by three orders of magnitude (i.e. 1000 times) over that of non-MPP searches.

That makes it possible for companies to effectively manage and maintain one unified archive. Otherwise, the data for a particular eDiscovery litigation event (which would hopefully be much less than the total archive) would have to be copied separately. That is not as easy as it sounds and results in two copies of the data, since the original data has to be left in the archive. That also means, by definition, that there is no unified management mechanism. One archive means less space, one uniform set of policies. That avoids confusion and conflicts, and means there can be no conflicting displays or reports that purport to show two different versions of what the enterprise perceives to be the truth.

There are a number of benefits in a Unified Archive-managed environment. From an internal IT perspective, storage administrators should be happy. Active production systems have a new, slimmer look, which increases the performance of processes such as weekly back-ups. On the archive side, Unified Archive's support of deduplication and single instancing means that redundant data, such as many copies of the same e-mail attachment, need not be stored, leading to overall savings.
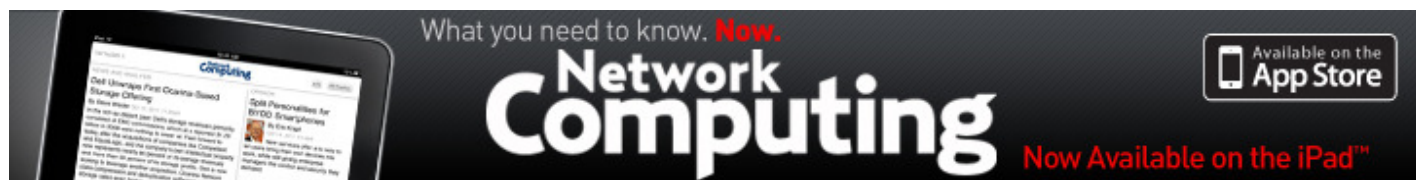
The end user also benefits. Individuals can find and use information without the need for the intervention of a systems administrator. But it is the larger business that derives the biggest benefits. An information governance process can be put in place that truly manages data retention. Data that should be purged can be, but data that needs to be retained will not be lost and create legal exposures.Compliance is another area that the Unified Archive can help address for both government agencies and corporate organizations.

However, the biggest enchilada is records management, which deals with the critical issue of data retention management. eDiscovery is another key area of concentration. Unified Archive provides a number of rich features and functions in addition to classification and search, such as legal hold capabilities, strong analytics, and conceptual searching, that are crucial to effectively managing eDiscovery processes.

By now, you should have gotten a sense of why active archiving in general is important and, by way of illustration, how ZL Technologies' Unified Archive functions and features support active archive requirements. Until recently, IT did not have to know "what" and "where" data was at a level of granularity that included the individual file or document. For semi-structured information, an active archive answers both questions. "Where" the data is located is obvious; it is in the archive and we can get to it. "What" can now be answered through indexing and search capabilities that allow the efficient exploration of each data item in as much detail as can possibly be extracted.

As pointed out, what defines the ability of an active archive to do its job is an overarching management software framework. For ZL Technologies, that framework is Unified Archive. Unified Archive overcomes scalability and performance obstacles that cause some other active archive solutions to stumble. It also provides a wide and deep set of features and functions that include comprehensive solutions for eDiscovery, compliance, and retention management. With Unified Archiving, ZL Technologies has made itself a player in data center transformation processes that require a move to active archiving.

At the time of publication, ZL Technologies is not a client of David Hill and the Mesabi Group.

Copyright © 2012 United Business Media LLC, All rights reserved.