Craig S. Mullins

Return to Home Page

August 3 - 9, 2007

HOUSTON BUSINESS JOURNAL

Financial Institutions' Sensitive Data Requires Careful Archiving

By Craig S. Mullins

Scanned PDF **HBJ web site**

Data retention is a significant aspect of regulatory compliance. The need to retain data is impacted not just by highly visible regulations like Sarbanes-Oxley and Basel II, but also by over 150 international, federal and local laws that govern data retention. Financial institutions are particularly impacted by these laws, due to the sensitive nature of the data with which they are entrusted.

Given this regulatory landscape, financial institutions need to develop plans for archiving data from operational databases as their data retention requirements expand. Many financial institutions have already begun working with experts to implement database archiving that will protect their institutions from potentially crippling litigation and settlement costs.

Litigious Society

A recent survey of corporate counsel found that the typical U.S. company faces an average of 305 lawsuits and spends $12 million a year on litigation alone, not including settlements or judgments.

In this technological age, many businesses have migrated to storing corporate data purely electronically. As such, there has been a strong focus on e-discovery, led by the recent changes in the Federal Rules of Civil Procedure. Companies understand the importance of this data and the need to preserve it, however, many are not properly retaining and securing structured data for business and legal protection.

There are typically three main requirements lawyers hope to meet when providing legal protection for their clients:

- Guaranteeing all data is secured in its authentic form.

- Ensuring easy access to data.

- Allowing the application of a "litigation hold" to any data -- so it can't be deleted before the trial is complete.

Drowning in data

Overcoming the challenges posed by these requirements can be difficult because organizations are generating and keeping more data now than at any time in history. And, according to industry analysts, enterprise data is more than doubling every year. Complicating matters, as much as 80 percent of that data is not actively used to conduct business.

Why are so much data being produced? Advances in technology have better enabled the ability to capture and store data, but technology alone cannot sufficiently account for the current rate of data growth.

Data is retained for both internal and external reasons. Today, organizations are storing more data for longer periods of time, in many cases to enable more in-depth data analytics. But external reasons, driven by the mandate to comply with legal and governmental regulations also compel businesses to store additional data.

Maintaining data in operational databases over long periods of time creates problems -- transactional performance

problems as data volumes expand and authenticity problems for data that must be maintained for legal purposes. As such, data must be periodically archived from operational databases.

Database Archiving

But what is database archiving? Database archiving is the process of removing selected data records from operational databases when they are not expected to be referenced again and placing them into an archive data store where they can be retrieved if needed.

As simple as this may sound, there are many significant challenges and requirements posed by database archiving. Perhaps the most important consideration is that archived data must be hardware and software independent. When data retention requirements span decades, the production system from which the data was archived may no longer exist -- at least not in the same form, and perhaps not at all. For example, listening to music on an 8-track is virtually impossible today.

The archive also must be able to store a large amount of data. As more data is stored, more data is archived. Combining this with long mandated data retention periods provides an explosive combination.

The archive must be able to manage data for very long time periods. Many data retention requirements are stated in decades, so the archived data may outlive the systems and the programmers that generated them. And because data structures change over time, the archive must be able to support multiple variations of the data structure as it changes.

To support regulatory compliance, data must remain unchanged once it is archived. So the archive must be able to protect against data modification. Only "read" access should be available to the archived data, and the data must be guaranteed to be authentic. Mechanisms to prevent surreptitious modification are necessary, too.

Finally the archive requires metadata to be useful: Metadata tags define the archived data, the archive processes and what is archived when. The archive must be able to store multiple versions of the first type of metadata. As the operational schema changes, the archive must track and function across these variations in schema. The second type of metadata also logs when data is actually archived. The final type of metadata controls which data is archived, when, and from where it was archived. Multiple types of metadata are needed for the archive to operate properly.

Taking all of these considerations into account, a secure, durable archive data store must be used to retain data that is no longer needed for operational purposes, and it must enable query retrieval of the archived data in a meaningful format until it is discarded.

CRAIG S. MULLINS is corporate technologist for Sugar Land-based NEON Enterprise Software Inc. (www.neonsoft.com), which provides enterprise data availability software and services.

From *Houston Business Journal*, August 3 - 9, 2007.

Home.