

**March 2008**

## **Structured Data Is At Risk for Long-Term Readability**

By Julie Lockner, Solix Technologies

It is safe to say that organizations of all sizes, across all verticals, store massive volumes of business critical data in a database or multiple databases. In order for these companies to operate on a daily basis, this information needs to be available to the end user whenever necessary. As the data ages and time passes, keeping the application that created the data available to read the data becomes more challenging since technology becomes obsolete over time. Continually upgrading the technology can be extremely costly and sometimes not possible, but is required to maintain database readability.

Each component in the technology stack is required to be online and operational in order to ensure the data can be read in the same context as when it was written. Upgrading the application to maintain readability involves many dependencies on the application vendor, the database vendor, the O/S and server vendor, the network vendor and the storage vendor. Standards exist to improve the ability to swap out and replace old technology with new, but the effort is huge and costly as data volumes continue to grow. The challenges are manageable over a 3 to 5 year span. But consider how much technology and standards have changed in the past 20 years. Then look into the future and consider the changes 20 years from now. In order to keep the data available and online, IT is required to continually upgrade the entire technology stack in order to keep current. This is time consuming and costly, but necessary.

Additional challenges associated with upgrading an application include:

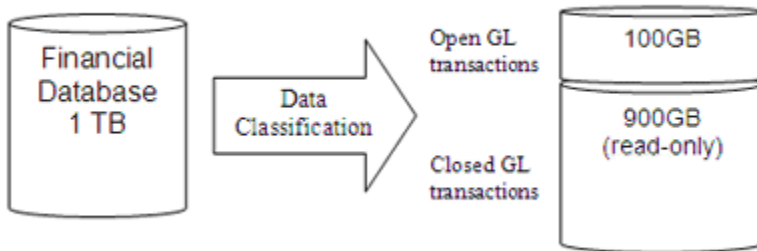
- Upgrading every 2-4 years is expensive and in many cases may not offer a ROI
- Companies choosing not to upgrade on a regular basis may face larger costs associated with re-implementation projects
- Upgrades require downtime which can impact the business
- Upgrading a production application also requires upgrading all test, development, standby, and disaster recovery copies as well

SNIA's Data Management Forum promotes using an Information Centric approach to data preservation to address some of these challenges associated with upgrading business critical applications. The information-centric approach begins with IT collaborating with the business users to classify business information based on the value to the business and any data retention policies that need to be adhered to. As the value of the information becomes less critical, or ages beyond a data retention period, the data can be moved to an infrastructure with different service-levels or deleted/purged. Without applying an information centric strategy to the data, all information is classified into a single category. The technology upgrade process must be applied to large volumes of data, consistently, without consideration to the actual value of the data.

By applying an information-centric approach, companies can realize lower TCO, improved production application performance, improved operational efficiencies and a significant reduction in upgrade costs and time. Here is an example to help illustrate the benefits:

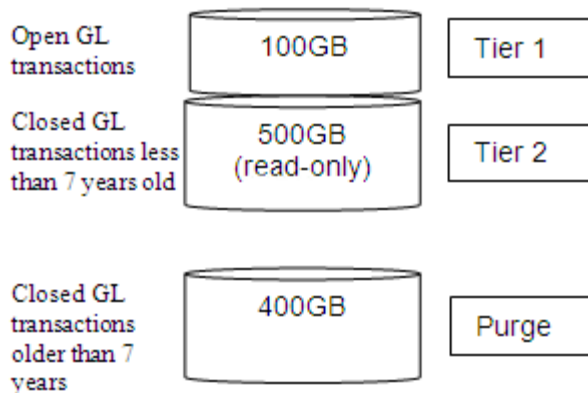
Upgrading a database application with a multi-terabyte data store requires upgrading the application and database stack to a current version. It is not uncommon for IT data centers to take advantage of this upgrade opportunity to also upgrade the application and database server, operating system and storage infrastructure to take advantage of new features. The upgrade process also requires making additional copies of the production environment to simulate the upgrade for testing prior to upgrading the production environment. This entire process could take as long as 6 months to 1 year to accomplish, resulting in an extremely high cost. Consider the same application now in 10 years, assuming the data store has grown to petabytes. As the size of the data store grows, along with the number of copies required to achieve the upgrade, the result is an exponential growth in the costs and time involved in the upgrade process.

By using an information-centric, data classification approach, and implementing information lifecycle management policies, the size of the production volumes can be significantly reduced, which drastically improves an upgrade process. When it comes to upgrading the technology stack, smaller production data volumes can be upgraded with significantly less effort, time and cost. Classifying database data involves mapping the business process to the data stored in the database. For example, if the database application stores financial data such as general ledgers when the booking period closes, the general ledger data in the related tables and rows becomes read-only.



In the US, general ledger data needs to be made available to the business for reporting purposes for up to seven years. Then it is no longer required. At this point, the data should be purged if the business policy and data retention policy allows.

Applying a Data Retention Policy  
Reduces Data Volumes



Using this classification example, the data volume that is required to be online has been reduced from 1 Terabyte to 600 Gigabytes (100GB open data and 500GB closed data), a 40% reduction. The benefits this type of approach offers include:

- Improved application performance due to smaller production tables
- Lower production management costs due to less burden on application servers and storage
- Faster backup and restore of production data
- Lower cost if the read-only data is stored on lower cost storage using a server with less CPU requirements, potentially lowering database license cost

The technology upgrade process is also significantly improved because the production database size is much smaller. During the upgrade process, copies of production data are made to test the process before upgrading. Because the production environment is smaller, so are all the copies, lowering storage costs and server requirements. The read-only data set also needs to be upgraded, but can be done after the active data set is upgraded, reducing production down-time during the upgrade process. The purged data doesn't need to be upgraded at all.

If data cannot be purged, but needs to be retained for longer periods of time, another option is to export the data out of the database to an application and database independent format, such as XML or a character delimited ASCII text archive file. When data is exported from the database, in many cases, accessing the data from the native application becomes difficult if not possible, depending on the application. In this case, if data needs to be accessed, it can either be reloaded back into a database of a newer version or different platform, or a reporting application can be used to view the data directly from the file. When data is archived from the database to a file, the only technology that would need to be upgraded in this scenario is the archive media where the file is stored.

SNIA is focusing on developing standards to address data migration and technology upgrade challenges. For example, the XAM specification (eXtensible Access Method) aims to provide ISVs, storage vendors and end users with a standard interface for unstructured content to address needs such as interoperability, information assurance (security), storage transparency, long-term records retention and automation for Information Lifecycle Management-based practices. This could be applied to database data archived to an application independent archive file. SNIA's Data Management Forum (DMF) launched a task force specifically focused on researching the long-term archive and digital information retention requirements. One of the primary goals of this task force aims to assist end users and practitioners in understanding how to implement best practices associated with long-term digital information retention, archiving and compliance. More information and resources are at [www.snia-dmf.org](http://www.snia-dmf.org).

#### **About the Author**

Julie Lockner was the Treasurer for SNIA's Data Management Forum (DMF) in 2007. The SNIA DMF is a cooperative initiative of IT professionals, vendors, integrators, and service providers working to address customer information management issues related to data protection, compliance, cost, and complexity. Ms. Lockner has over 12 years of experience architecting, marketing and managing database applications in the ERP, CRM and Marketing Analytics space. She has held various engineering, sales and marketing positions in companies such as EMC, Oracle, Verbind (acquired by SAS Institute) and Raytheon. As vice president of sales operations at Solix Technologies, she is responsible for defining and implementing sales and product strategies for Solix's data management suite.