



Database Information Management: An Essential Business Practice



By Gary Zasman (left), WW Practice Director for Business Applications and Database Integration, NetApp, and Chair of the DMF's Long-Term Archive and Compliance Storage Initiative and Michael Peterson (right), CEO Strategic Research Corp. and TechNexus, and Chief Strategy Advocate for the SNIA's Data Management Forum.

The conversation always seems to be about how the datacenter is drowning in the growth of unstructured (file-type) information, but not about the challenges of managing the content owned by the enterprise applications and databases that actually run the business? Rather than focus on what type of information and data are most common, let's ask, "Which processes are being used to manage and control all that information and data?" Control is what is at issue and databases are running away with the game. Databases of one form or another control 50-75% of

information and data in the datacenter and are predicted to be the growth path.,

Why Database Information Management?

It is true that enterprises are seeking to cope with the deluge and complexity of information and data that they own, and most are putting them under some form of structured organizational control. Yet the irony is that the more database influence grows, the greater the information management problems within them become. The database administration community has long focused on what they call the "data management" challenges associated with databases: data quality, cleansing, modeling, master data, mining, etc. Now is the time to address the "information management" challenges as well. Applying a management framework such as Information Lifecycle Management (ILM) is a viable approach to help optimize the performance and reduce the operating costs of databases and the enterprise applications that rely upon them.

Database Information Management Challenges

How important is it that we focus on improving database information management? Consider the answers to these questions.

"Are the database IT practices in your organization in alignment with the business needs and requirements?"
"Have the requirements for your information assets been defined by all functional units including Legal, Security, and Compliance?"

What problems occur if the full range of information management requirements is not considered or if one department dominates what should be a collaborative process?

Clarifying Database Management

db Data Management

- Data Quality, Cleansing
- Modeling, Architecture
- Warehousing, Mining
- Metadata Mgmt
- Master Data Mgmt

db Information Management

- ILM-based practices
- Classification and Policies
- Archive
- Retention & Preservation
- Tiering and Storage Pooling
- Dynamic Provisioning
- Capacity Optimization
- Test/Dev Data Management
- eDiscovery
- Deletion

It is not that the industry is starting out on this quest from scratch. Many good practices are already in place. Some databases and enterprise applications even have Information-Lifecycle Management (ILM) style classification and requirement setting provisions to establish retention policies and to tier or archive inactive information and data, and remove it from the active data-set

To illustrate, here is an example:

In this example, ERP, CRM and custom applications support mission critical business processes and carry out business transactions that create, record, and update active data. After a time period each transaction is completed. For example, a customer order is fulfilled, an airline ticket reserved, or a car insurance claim has been serviced. As time goes by, the associated process data is no longer updated since either the transaction is closed or the data just becomes inactive. Each of these transactions has personally identifiable information (PII), requiring they be managed and retained securely in accordance with legal and compliance requirements. (potential for sidebar #2 here – see associated document)

As operations continue, the business applications dynamically evolve requiring ongoing maintenance and tuning to match changing operations needs. Making test and development clones of the production dataset are a normal and routine practice. With each clone, each test dataset is propagated onto new storage, driving up storage demand and associated provisioning and maintenance costs. Some of these datasets are sent to remote and relatively uncontrolled development locations.

This leads to some questions: Does the company

realize that confidential and personal information is included in these clones? Are there proper safeguards in effect or is this a hidden risk? When a test or development project is finished what happens to the data and the allocated storage space? What retention and deletion policies are in effect and are they properly followed? Here is an entire domain of information management challenges and control issues that on one hand can expose critical and confidential information and on the other may result in uncontrolled data growth and excess cost.

More questions about the organization's information management practices need to be asked. Going back to the transactions we were following, what is the retention period for this class of transaction, when and how should they be deleted? If they must be retained long-term, then where and how do they get preserved? Will these records ever be deleted? What is the risk to the company for keeping them? Are old transaction records clogging up the database reducing its responsiveness and increasing storage and operating costs? Are inactive or expired records being kept on expensive Tier-1 storage? What happens to the database when applications are retired but the information must still be retained? How will this data be read as information without the application 10, 20, 50 or 100 years from now?



This example is just the ‘tip of the iceberg’ of the many challenges that exist in the domain of database information management. They range from excess cost, clutter, and performance degradation to the unique requirements of databases relative to deletion, retention, and preservation practices and to business concerns such as legal discovery or regulatory and compliance mandates that encompass all Electronically Stored Information, ESI. Issues like these translate into opportunities for organizations to reduce operating costs and capital costs by implementing improved practices. Some examples include reducing capacity bloat by archiving inactive data, deleting expired data, or using tiering, capacity optimization, and thin provisioning to reduce storage operating costs.

It is not that the industry is starting out on this quest from scratch. Many good practices are already in place. Some databases and enterprise applications even have Information-Lifecycle Management (ILM) style classification and requirement setting provisions to

establish retention policies and to tier or archive inactive information and data, and remove it from the active data-set. A place to start is to investigate the practices your own company has incorporated and how it copes with its own database information management challenges.

Call to Action

Database information management is not a new field, but it is time to focus on it. The Data Management Forum’s Long-Term Archive and Compliant Storage Initiative has formed a Database Information Management Special Interest Group, known as the DIM-SIG, to address the need for cooperative market education and best practices in this field delivered world-wide through SNIA’s broad resources. This group is open to all SNIA members interested in fostering industry-wide programs for database information management. For more information regarding this new SIG, visit http://www.snia.org/forums/dmf/programs/ltacsi/dim_sig

Database Information Management Challenges

Compliance & Litigation

- Weak long term retention preservation practices put information at risk of being lost
- Lack of declassification practices
- eDiscovery readiness
- Maintaining security, confidentiality and privacy on production and test
- Performance degradation over time as loads increase

Storage Practices Performance

- Tiering and capacity optimization practices
- Inadequate service agreements and clear requirements
- A cost-efficient plan for migration and preservation

Cost & Efficiency

- High percentage inactive and expired information retained within the database
- Proper deletion practices
- Improving storage provisioning and utilization efficiencies
- Reducing operating costs

Operations

- Development and test practices including management of the test data and storage
 - Application retirement and migration