Database Information Management:

A Taxonomy of Practices

September, 2009

Principal Author:

Michael Peterson Chief Strategy Advocate for the SNIA's Data Management Forum, CEO, Strategic Research Corporation and TechNexxus

Supporting Authors:

Gary Zasman Chair of the Data Management Forum's Long-Term Archive and Compliant Storage Initiative, and WW Practice Director, Business Applications and Database Integration, NetApp

Peter Mojica Chair, DMF-LTACSI Reference Guide Committee, Risk Management and Compliance Consultant

Paul Field Independent Consultant, DMF Treasurer

And DMF member companies:

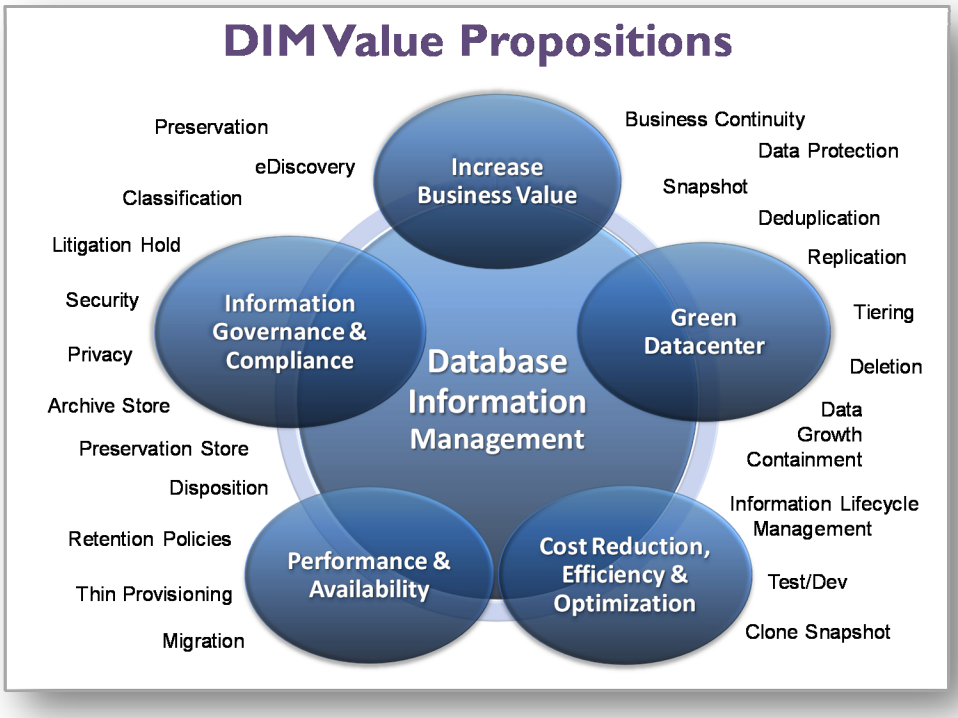
Produced and supported by the SNIA Data Management Forum. The Data Management Forum, DMF, operates a public website at www.snia.org/forums/dmf Comments to this document can be discussed online at the DMF's community site as well. http://community.snia-dmf.org

The database and enterprise application world tends to overlook the practices of enterprise information management and instead focuses on data quality and data use, referring to these as a database field called Data Management. Data Management, in the context used by the database community, is about the data itself, not its governance or business requirements. The problem is that the large body of practices used to manage and govern information and data assets remains ad-hoc, unorganized, and unstructured in the database realm. Change is needed to satisfy the overarching requirements for consistent enterprise-wide information management and governance that includes information and data under database control. In response to this need, the SNIA's Data Management Forum is proposing a structured practice area it is calling Database Information Management (DIM)[1].

What Is DIM?

DIM is a service management style methodology for managing information and data controlled in a database or enterprise application domain in accordance with its business requirements and policies. By properly applying many practices and services such as those identified in Figure 1, DIM's value propositions span information governance, litigation and compliance, improved application performance and availability. The benefits of this approach range beyond better control and management of information assets to include cost reduction, improved operating efficiencies, reduced energy consumption, controlled growth, and overall improved business agility and value.

Figure 1



Whether you look at DIM from the perspective of its value propositions or from the lengthy list of practices that support it, DIM is more than just a bunch of discrete and disconnected practices.

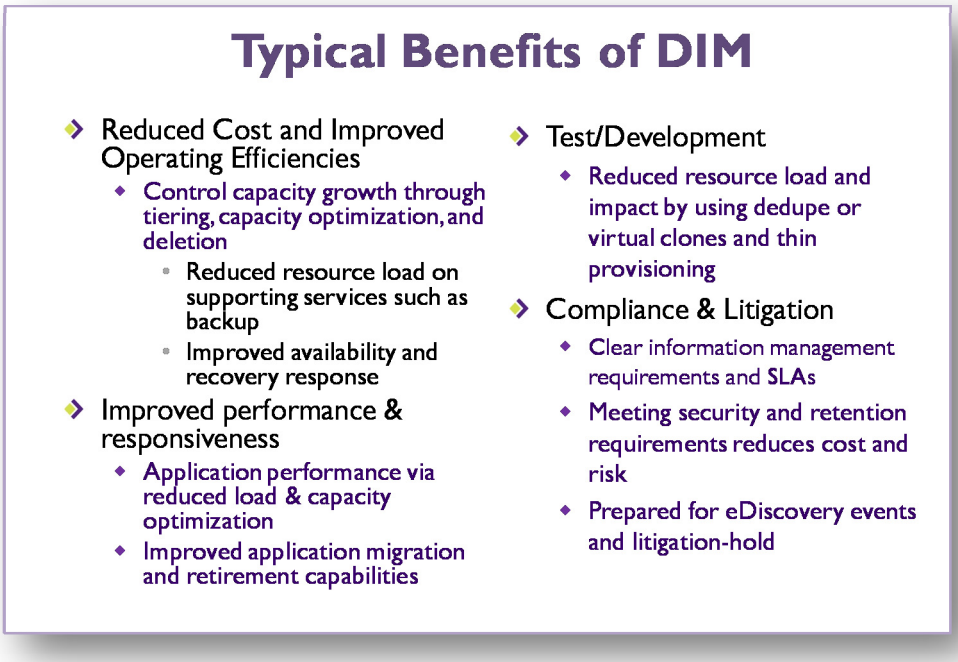
[1] See the white paper: "*DIM-Essential-Business-Practices*," April 2009

The premise of the SNIA's Data Management Forum's view of DIM is that the underlying management framework is based on Information Lifecycle Management, or ILM2.0, style service management methods[2]. The foundation of ILM2.0 is the call for a collaborative effort owned by the organization's information governance committee to identify, classify, and define the business requirements for the organization's information and data assets and to set the business policies governing operations. With proper service level requirements, database information management practices can be designed and implemented by records management, IT, database administration, application owners, and security to operate cost effectively in accordance with the organization's policies.

How to Manage Information and Data

Some people have argued that the simple goal of IT is nothing more than to store vast amounts of data for the lowest possible cost. Others will admit that they also have to meet other business needs such as meeting compliance requirements for retention, disposition, and protection. Unfortunately, this is too simple a picture. The typical datacenter is far more complex, and its business and operations requirements have to be included in the equation. Information is both the chief asset and principal liability of any organization in today's risk filled litigious and regulated world. Enterprise Information Management is a critical practice that has to be applied to information and data that are under the control of databases. And, ILM-based service management is the best methodology available for translating business requirements into infrastructure practices. Just as ILM-based practices are applied to unstructured information; it is time to apply ILM to databases and the information and data under their control.

Figure 2



In an ILM2.0 methodology, DIM practices are the service

[2] ILM2.0: Source: DMF white paper, "*ILM2.0: the Next Phase,*" 2009. ILM2.0 is a service management style framework for cost-effectively aligning datacenter storage, security, services, applications, and infrastructure with the business requirements for the organization's information

catalog, the available pool of services and practices utilized to achieve the business requirements. When multiple services are implemented in concert, the results are the many potential benefits outlined in Figure 2, ranging from cost savings and improved operating efficiencies to improved operations compliance. The potential to reduce cost and improve efficiency alone is worth the effort.

How to Get Started with DIM?

It is always best if the business requirements including governance policies and service level agreements are defined first. The second step is to always begin in the areas that have the greatest payback. Examples might be the cost containment and resource reduction available through capacity management processes such as tiering and deletion.[3] Reducing risk by achieving regulatory compliance or being ready for litigation and eDiscovery is an easy call. Another example is to contain primary capacity growth through thin provisioning and tiering and test/development capacity growth through thin provisioning and use of virtual clones[4] instead of full clones. These are a just a few examples. DIM practices can be segregated and implemented in silos or as part of a larger strategy as needed. Cost reducing test/dev is an example of a business objective. Use the business objectives to drive what needs to be implemented and the priority. Here lies the solution to understanding how to put DIM practices to work. The business objectives frame the practice disciplines and define a taxonomy.

Taxonomy of DIM Practices:

DIM consists of many practices that have overlapping benefits depending on the way they are deployed and how you look at them. This requires a taxonomy viewpoint to see the relationships. The following table is designed to be a start at explaining the relationships between the business objectives and typical DIM practices with an ILM2.0 style service management framework as the context. Please use it recognizing that it is not all-inclusive because of the breadth of the field and its dynamics. [5]

[3] Note: the term "database archiving" when used to mean a form of migration to the lowest level tiers where data is moved out of the database and stored 'offline' or 'nearline' is deprecated in the context of ILM-based practices and replaced with migration to a lower-level tier. Don't confuse the two practices. An archive is a 'specialized preservation-class repository', not just any storage system or a tier of a storage system. Copying or migrating certain classes of information and data into an archive for long-term, compliance, or litigation support preservation is still valid, but for these purposes preservation needs to begin upon creation not at end of life. Consequently, the ILM-based practices of retention policies, deletion, tiering, migration, and preservation speak more precisely to the way in which information and data are managed over their lifecycle. Should you choose to still use database archive as a term related to offline tiering or hierarchical storage management, please define it in the context of tiers of storage and specify the preservation aspects of it, if any. But, please remember that a 'Tier-4' bit-bucket is not an archive unless it has the required set of specific preservation services. Otherwise, it is just a tier of storage.

[4] Virtual clone: A secondary database clone of reduced capacity or virtual capacity created by a process such as a snapshot of a primary clone that can be used for test and development purposes.

[5] Note: the definitional discussion of many of these terms is sourced to "*Building a Terminology Bridge, for Digital Information Retention and Preservation in the Datacenter*" published by the SNIA's Data Management Forum, May 2009

DIM PRACTICE TAXONOMY MODEL

Objective	DIM Practices	Benefits
Align Business and IT	ILM2.0-based service management framework, Classification, Requirements Setting, Policies, SLAs	Provides the methodology for information management and supporting data and information services and practices
Improve Application Performance	Tiering, Thin Provisioning, Snapshot-based data protection, Virtual cloning for test/dev[1]	Reduce resource load and impact from data and information services, contain capacity growth, dynamically shrink and expand capacity as needed, dynamically store information and data on the lowest cost systems that can support the requirements while still maintaining access, reduce cost and power consumption
Control Storage and Network Costs	Deletion, Thin Provisioning, Data Deduplication, Compression, Tiering, Snapshot, Virtual Cloning, Optimize data protection	Reduce capacity growth and demand, dynamically shrink and expand capacity as needed. Reduce size of data protection and disaster recovery pools, test/dev clone capacity demand, WAN traffic, and power consumption
Info Gov/Compliance	Classification, Set Requirements and Policies, Retention Policies, Deletion, Preservation Repository, Masking, Information Assurance	Reduces security, compliance, and litigation risk and cost, meet mandated regulatory requirements, reduce litigation risk
eDiscovery & Litigation Support	Classification, Retention Policies, Deletion, Litigation Hold, eDiscovery Practices, Preservation/Archive Repository	Become 'discovery enabled" reducing cost of discovery and risk, establish practices that avoid spoliation and allow litigation holds and deletion to operate

[1] See the next page for definitions of terms.

Objective	DIM Practices	Benefits
Cost Reduce Test/Development	Virtual Clones, Thin Provisioning, Data Deduplicaiton, Masking for Privacy/Confidentiality,*	Reduce capacity consumption, time to create clones, and resource loads, maintain confidentiality in the test and development environment, dynamically allocate and reclaim space reducing cost and complexity
Reduce Cost and Pain of Application Migration	Virtualization, Snapshot, Tiering, Thin provisioning, Retention Policies, Deletion, Data Deduplication, Information Assurance	Reduce the amount of material to be migrated and the cost of the hosting platforms and storage, protect confidentiality
Improve Security Practices	Classification, Requirements Setting, Access Controls, Information Assurance, Masking, Retention Policies, Deletion, Tiering	Integrating security with information management gives a holistic and more uniformly applied practice at lower cost
Cost Reduce Business Continuity	Application Recovery, Snapshot, Data Protection, Disaster Recovery, Data Deduplication/Compression, Remote Replication, Deletion	Reduces operational risk by employing more granular processes to ensure service consistency and recoverability. Reduce cost due to a reduced load on services and improve recovery time.
Reduce Power Consumption	Thin Provision, Snapshot, Tier, Migrate, Data Deduplication/Compression, Deletion	Become more 'green' by reducing and controlling capacity and resource demand

* **Snapshot**: A fully usable copy of a defined collection of data that contains an image of the data as it appeared at the point in time at which the copy was initiated. A snapshot may be either a duplicate or a replicate of the data it represents. (SNIA Dictionary)
Virtual Clone: A secondary database clone created by a process such as a snapshot of a primary clone. Virtual clones can be used for test/dev instead of using full clones to reduce storage consumption, time, resource loads, etc.. (Source: SNIA-DMF)
Thin Provisioning: A technology that allocates the physical capacity of a volume or file system as applications write data, rather than pre-allocating all the physical capacity at the time of provisioning. (SNIA Dictionary)

Business Objectives Driving DIM

This table of taxonomies and practice relationships includes a column called business objectives that provides an important perspective to the value proposition for DIM. What follows are further explanations of each objective and the opportunities DIM presents:

Align Business and IT around Information Governance: Business and information governance efforts have created a new opportunity to establish this important alignment and a resurgence of effort. Service management methods are now being applied to governance creating an intersection with enterprise information management which opens the door for the application of ILM-based methods to database information management. Providing IT with clear requirements, policies, and service agreements really does work at motivating collaborative efforts to reduce cost and improve efficiency. The alternative data management methods or siloed information and data services in use today are just that, siloed processes. A holistic ILM-based approach to DIM has a much greater probability of long term success.

Improve Application Performance: This objective focuses on reducing resource loads on servers and storage systems and in the process improving responsiveness as well as dynamic scaling capabilities. Reduction in load comes about from tiering and migrating inactive, reference, and expired information off primary storage, from deleting expired information, from thin provisioning storage resources so that they can dynamically respond to changing demand, to using snapshot technologies to create virtual test/dev clones so that creating clones does not take up valuable production time and storage resources, and to use snapshots for data protection to eliminate loads and quiescence periods on the primary application. All these are examples of DIM practices orchestrated in this capacity. Existing data management practices do not address reducing resource utilization and optimizing performance this way.

Control Storage Costs: Many DIM practices center on improved storage and capacity optimization processes. The result can be profound reduction and control of storage growth, the number-one storage related management pain in the IT datacenter. The list of benefits is long, but it is important to point out that these practices need to be aligned with the business requirements and policies so that in the enthusiasm to control storage growth, no other damage is caused.

Information Governance and Compliance: Improving governance and compliance preparedness and conformance is effectively a risk management activity. The costs of failures in this domain are legendary. Success does not come by leaving IT in a vacuum. ILM-based methods utilizing DIM practices are very valuable in focusing IT practices on how to meet the business requirements cost effectively.

eDiscovery and Litigation Support: Like governance and compliance, ILM-based DIM practices can be optimized to cost effectively address eDiscovery and litigation support needs and enable preparation and responses. For example, having practices in place based on business requirements and policies to properly and consistently execute retention and deletion policies,

including controlling litigation holds is crucial. The only safe-harbor in operating automated deletion practices requires this approach.

Cost Reduce Test and Development: It is not unusual to find 10-20 copies of the production database sitting in full clones being used for test and development. The additional cost and risk that this imposes is so huge that fixing it is a 'no-brainer' in the area of cost reduction. DIM practices include methods such as data deduplication of clones or using virtual clones instead of full clones to eliminate the capacity consumption and time factors involved in creating and managing lots of clones in a typical test environment. Thin provisioning is another good methodology used to optimize and provide agility to capacity utilization that can be applied to clones. Then, there is the privacy problem of clones being shipped to development sites with no privacy or confidentiality controls in place. At a minimum, masking of personally identifiable information and confidential information are required. This is pointed out to illustrate the need for a holistic approach to DIM. Services must be implemented in concert with the business requirements and policies for the information and data being managed.

Reduce the Cost and Pain of Application Migration: Systems and applications hitting end of life or retirement must be migrated to newer platforms and integrated into current applications if their information and data assets are to remain useable. DIM practices can be applied to this problem and make significant contribution to simplifying, expediting, and cost reducing the process. Technologies such as virtualization, thin provisioning, and snapshots can make the work of migration much simpler. Tiering and migrating inactive, reference, and expired information prior to the migration job allows it to be tested and phased first with the active data (a much smaller data set) reducing potential disruption. Preserving business critical information or compliance information in a preservation store consistent with business policies is also important.

Improve Security Practices: Integrating security and confidentiality requirements for information and data assets as a business requirement and a component of ILM-based DIM methods is a logical approach to improving information assurance compliance. Security and confidentially are requirements and policies on par with performance, availability, retention period, or compliance requirements.

Cost Reduce Business Continuity: Business continuity practices are among the more expensive and complex practices of the IT datacenter. An example of how DIM practices can be applied is the potential benefit of utilizing capacity optimization technologies[6] to eliminate duplication and redundancy at various levels, reducing the traffic being synchronized through the WAN and the size of the supporting repositories. Smaller loads and more responsive smaller repositories are important improvements at reducing cost and improving efficiency and agility.

Reduce Power Consumption: Going 'green' is the mantra of the day. Reducing the server and storage requirements is a direct corollary to reducing power consumption. No other database practices are oriented towards a holistic approach at achieving these goals.

[6] Capacity Optimization: Methods that reduce the consumption of space required to store a data set such as data deduplication, compression, and thin provisioning. (Source: SNIA Dictionary)

Recommendations

This is a long list. Cutting through the fog of what to do and how to do it, can be a challenge. The SNIA's Data Management Forum recommends that ILM2.0-based practices (as a requirements oriented, service management methodology) should be applied to managing information and data within database control. As previously detailed, ILM2.0-based practices provide a holistic methodology for implementing database information management and will have great cost and operational benefits to your organization.

Our recommendation is to start with the greatest pain or risk and the greatest payback. Through some high-level group such as an Information Governance Committee, we recommend that your organization begin the process of classification and setting the requirements for information assets. Then, by mapping these requirements to policies and SLAs, the organization will have a holistic picture of its information management challenges and priorities. IT can take the SLAs and design supporting service catalogs and practices and implement them appropriately utilizing this taxonomy as a starting point. Remember, this does not have to be done all at once and it should be adjusted for your organization's specific needs. Again, start with the greatest payback, if possible.

The Data Management Forum is operating a special interest group on database information management. The DIM-SIG is developing these and many more deliverables and best practices for the industry. You can engage, access the online resources, and discuss issues you have with our experts through the DMF's website at http://www.snia.org/forums/dmf and the DMF Community site, http://community.snia-dmf.org. Your feedback is welcomed, encouraged and expected.

About the Data Management Forum:

The SNIA Data Management Forum is a cooperative initiative of IT professionals, vendors, integrators, and service providers working together to conduct market education, develop best practices and promote standardization activities that help organizations become Information-Centric Enterprises. Areas of focus include the technologies and services that support information lifecycle management, data protection, capacity optimization, database information management, and long-term retention and preservation. For more information, visit www.snia.org/forums/dmf or participate in our open online community http://community.snia-dmf.org

About the SNIA

The Storage Networking Industry Association (SNIA) is a not-for-profit global organization, made up of some 400 member companies spanning virtually the entire storage industry. SNIA's mission is to lead the storage industry worldwide in developing and promoting standards, technologies, and educational services to empower organizations in the management of information. To this end, the SNIA is uniquely committed to delivering standards, education, and services that will propel open storage networking solutions into the broader market. For additional information, visit the SNIA web site at www.snia.org.