Information Life-Cycle Management in a Virtual Data Center

Growing acceptance of commodity, standards-based hardware and server virtualization is leading many organizations to consider dramatic changes in their IT infrastructure. These changes are expected to transform the enterprise data center into well-defined building blocks of virtualized compute, storage, and network resources. In the resulting IT environment, business unit managers should be able to establish performance, availability, compliance, and data classification requirements for enterprise applications. The intelligent IT infrastructure will be designed to automatically provision the necessary hardware and software resources once business requirements have been defined, without further administrative intervention.

BY NIK SIMPSON AND MATTHEW BRISSE

Despite advances in many areas of data center technology, IT organizations are still grappling with growing complexity and cost. At the same time, administrators are under pressure to control total cost of ownership for the IT infrastructure. One promising solution is a data center infrastructure that is designed to adapt to changing business and application requirements by automatically provisioning resources—both hardware and software—as demand requires. The goal is to develop a data center that is built upon well-defined, standards-based building blocks of virtualized compute, storage, and network resources.

In such a virtual data center, when business unit managers roll out new applications they will establish business requirements for performance, availability, compliance, and data retention. In response, the virtual data center will have the capability to provision the necessary hardware and software resources to meet business requirements without further administrative intervention once those business requirements have been defined. The virtual data center will be designed to monitor each application to ensure that it meets the established business requirements (or updated requirements caused by changing business conditions)—dynamically adjusting data center resources as required. The result is envisioned as a closed-looped management system that will have the capability to become increasingly self-managed over time.¹

The concept of the virtual data center is powerful, but to exploit its full capability, enterprises require a strategy for information life-cycle management (ILM). The need for ILM derives from a basic tenet of the virtual

¹For more information about the virtual data center, see “Progressive Degrees of Automation Toward the Virtual Data Center” by Jimmy Pike and Tim Abels in Dell Power Solutions, February 2005; www.dell.com/downloads/global/power/ps1q05-20040228-Pike.pdf.
data center: installing an application is a logical task for which the administrator enters basic requirements—or service-level objectives (SLOs)—and the infrastructure management layer provisions the resources to meet those requirements. In the virtual data center context, ILM requirements must be reflected in the SLOs when an application is deployed. Examples of ILM requirements that might be specified include:

- **Downtime**: The amount of downtime that is acceptable for the application. This metric helps the infrastructure layer determine which disaster recovery options are available.
- **Time to zero value**: The length of time for the value of the information to reach zero. For example, if the system is producing weather forecasts, then the time to zero value is one day—yesterday’s forecast has historical significance but little value. This metric can be used to determine when to purge old information or move old information to less-expensive storage as its value declines.
- **Regulatory compliance**: Regulations, if any, that affect the information created by the application. For example, the information may need to meet the retention requirements of the Sarbanes-Oxley Act. This metric could be used to set up off-site file archives.

Providing ILM requirements as part of the application deployment process helps align information management actions with the value of the application from the moment it is installed.

### Understanding the information management problem

Although the basic concept of ILM has been a standard operating procedure in the mainframe world for several years, outside the mainframe world ILM is far less mature. In the open systems world, most IT organizations have some procedures in place (often in the form of custom scripts and manual administration tasks) to manage the information produced by applications and users. Unfortunately, these procedures typically suffer from the following problems:

- **Lack of integration**: Few IT organizations have a holistic approach to information management; procedures tend to be application- and operation-specific, using different tools with different management interfaces for each task. This heterogeneity can make it difficult to document procedures and ensure that the right procedures are in place for every application.
- **Insufficient distinction between information types**: Not all information is of equal value, and the procedures for tasks such as backup and disaster recovery should be driven by the value of the information. Today’s data centers typically implement one of two classes of ILM: a combination of off-site disaster recovery, extensive archiving, and comprehensive backup; or a bare-minimum tape backup. This binary approach can lead to excessive IT costs because administrators tend to err on the side of caution and overprotect information that may have little residual value.
- **Inability to scale**: An ad hoc approach to information management is feasible as long the number of servers and quantity of information remains small. However, even in midsize companies, existing ILM approaches can be difficult to scale as the number of servers and the quantity of stored information continue to increase.

Until recently, such problems have been marginalized while efforts were focused on physical infrastructure management. Today, management tools for physical infrastructure have matured and the time is fast approaching for organizations to address information management in a consistent, flexible, and scalable way. Reasons for a renewed focus on ILM include the following considerations:

- **Growing complexity in data and information management**: As organizations have come to rely on computing infrastructures for a growing number of day-to-day functions, effective information management has become increasingly important.
- **Regulatory compliance**: The problem of complexity is compounded by the increasing number of government regulations—such as the Health Insurance Portability and Accountability Act (HIPAA) and the Sarbanes-Oxley Act in the United States, and the Data Privacy Directive in the European Union—that affect the way organizations manage data.
- **Growth of digital information storage**: New applications, new regulations, and increased use of existing applications are leading to rapid growth in capacity requirements.
- **Performance problems caused by information overload**: As the amount of information stored by an application grows, the performance of the application typically degrades and administrative operations such as backups and disaster recovery also take longer.
- **Increasing maturity in infrastructure management**: In the past few years, manufacturers and industry standards bodies have cooperated to produce standards, such as the Storage Network Industry Association (SNIA) initiatives to simplify the management of storage infrastructure.

Altogether, these issues have caused information management to become a high priority for many organizations. Centralized
control of how information is managed can be a key first step toward more ambitious goals such as creating a virtualized data center. Perhaps the biggest roadblock for many organizations is a lack of standards.

The need for open standards
A standards-based approach to ILM procedures will require broad cooperation from the entire IT industry. To that end, the SNIA is currently focusing on information management and the underlying storage infrastructure through two projects:

• Storage Management Initiative (SMI): The SMI aims to create a standard set of technologies for the management of storage—whether in a storage area network (SAN), network attached storage (NAS), or a locally attached disk—and its associated infrastructure. With support from major vendors in the storage industry, the SMI has made great strides in creating a set of interfaces that are designed to manage storage infrastructure components regardless of vendor.

• Data Management Forum (DMF): The DMF is charged with two tasks: defining a common language and a set of SLOs for information management; and working with the SMI to define a road map for the development of ILM technologies and how those ILM technologies will interact with other management interfaces like the SMI specification.

The goal of the SMI and DMF efforts is to create the set of service layers shown in Figure 1. These service layers will be designed to communicate through open interfaces so that high-level SLOs at the ILM layer can be translated into low-level data management and storage management services that have the capability to provision and manage the storage infrastructure.

For example, an ILM SLO could require certain application data to be available 24/7 with a two-hour maximum acceptable recovery time. This would cause the data management layer to request a locally mirrored volume (for maximum reliability under normal operation) as well as synchronous mirroring to a remote data center for two-hour recovery in the event of catastrophic failures. As with other elements of the virtual data center, this capability is far from being realized. However, the DMF has defined a set of steps that can help organizations to develop the fully interoperable tool sets required for broad adoption of ILM technologies.

Employing the DMF ILM model
Understanding the DMF’s model for ILM (see Figure 2) is critical to understanding how ILM may be implemented in the future. The ILM model reflects a pragmatic approach to data management, in which management and control of the IT infrastructure is based on needs defined by the business framework. Business requirements coming from the business framework drive the management of applications and information. Within the ILM framework, the goals management layer transforms business requirements into policies that are enforced in the service and infrastructure layers. In turn, the goals management layer provides feedback to the business framework regarding cost, risk, and status.

Based on this ILM model, the system administrator would register an application with the mapping and control layer of the virtual data center management framework, which is detailed in Figure 3. The administrator would then specify a level of service and performance or a compliance requirement such as Sarbanes-Oxley. The virtual data center model would reference the ILM requirements to deploy the application and provision the necessary

---

**Figure 1.** A layered approach to the virtual data center

**Figure 2.** DMF framework for ILM in the virtual data center

**Figure 3.** The administrator would then specify a level of service and performance or a compliance requirement such as Sarbanes-Oxley. The virtual data center model would reference the ILM requirements to deploy the application and provision the necessary
resources and methodologies to meet the SLO—without requiring the administrator to provision resources, tune the application, or define a backup and recovery strategy.

Implementing intelligent, flexible data management
Because the information management needs of an application change over time, managed information should be continually reevaluated to help ensure that it is managed appropriately. Achieving this goal requires a flexible and intelligent data classification system capable of analyzing the information and applying data management rules that match the needs of the application and the value of the information. However, intelligent classification is only half of the solution. Once information has been classified, intelligent data placement is required to direct that information to the appropriate tier of storage—or to arrange operations such as backup and replication to help ensure that the services provided to the application meet the SLO.

The requirement for dynamic information management is spurring a new generation of software designed to automate data classification and use the results from classification to drive intelligent data placement. The combination of intelligent data classification and placement can allow software to execute simple tasks such as ensuring that adequate capacity is provided for the data created by an application. In addition, intelligent data classification and placement can enable the development of software designed to execute complex tasks that have the capability to meet the high-level requirements of Sarbanes-Oxley regulations. Creating such software will require the development of standard interfaces for every aspect of an application’s environment—allowing data to be moved, copied, purged, archived, or otherwise manipulated to meet the stipulated data management requirements.

Such comprehensive ILM requirements indicate that next-generation systems must be designed to support the following features and capabilities:

- **All data types:** To meet complex information management requirements, systems should be capable of working with all types of data—regardless of which application produces the data. Platforms that support a single data type or application will not suffice.
- **Any platform:** Corporate information is stored on many types of platforms, including UNIX®, Microsoft® Windows®, and Linux® operating systems. A comprehensive system should be agnostic about OS platforms and infrastructure choices such as NAS versus SAN.
- **Flexible classification:** The system must be able to combine a variety of techniques for classifying information to determine value. Conventional methods, such as age or size, should be combined with emerging methods, such as enterprise search engines, to provide the flexibility needed for truly intelligent data placement.

![Figure 3. Virtual data center management framework](image)

**Evolving into the virtual data center**
Well-defined, standards-based data center components for compute, storage, and network resources can enable enterprises to transition to a virtual data center infrastructure that has the potential to automatically provision hardware and software resources in response to changing business and application requirements. The key advantage of information life-cycle management is its ability to help solve the growing problem of data center complexity. By streamlining business processes, ILM can help reduce complexity and enable substantial cost reductions for IT infrastructure. Given its focus on business requirements and emerging standards, the ILM approach explored in this article is well suited to help automate the data center of the future.

**Nik Simpson** is director of marketing at Scentric Inc., an independent software vendor that designs next-generation ILM technologies. Nik also co-chairs the DMF’s Information Lifecycle Management Initiative Technical Liaison Group and co-authored the SNIA’s Guide to Storage Virtualization.

**Matthew Brisse** is a technology strategist in the office of the Dell CTO and is vice chair of the SNIA board of directors. At Dell, Matthew chairs the Standards Review Board and is a member of the Systems Management Architecture team.

**Storage Network Industry Association:**
www.snia.org