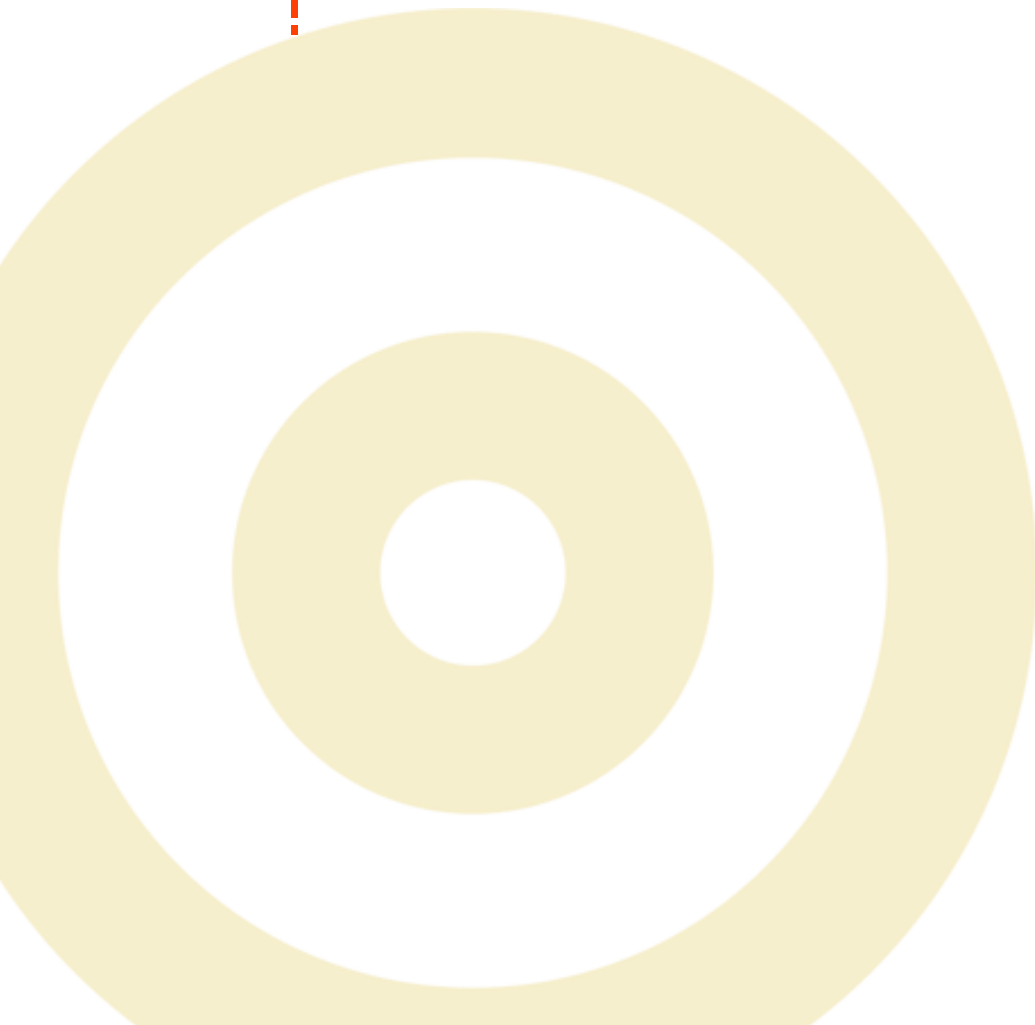




## A Blueprint for Tiered Storage



*Sponsored By:*



# A Blueprint for Tiered Storage

By Steve Duplessie

© 2004 TechTarget

BIOS

**Steve Duplessie**—Steve is the Founder and Senior Analyst of the Enterprise Storage Group, recognized worldwide as the leading independent authority on enterprise storage. Steve has also consistently been ranked as one of the most influential IT analysts. Prior to founding ESG, Steve was the Founder and CEO of Invincible Technologies Corp., a manufacturer of fault-tolerant NAS Systems. Prior to ITC, Steve held positions at Clearpoint Research and EMC Corporation.

This *IT Briefing* is based on an Apple Computer/TechTarget webcast, “[A Blueprint for Tiered Storage](#).” To view this webcast online, please click the link.

This TechTarget *IT Briefing* covers the following topics:

- What is Information Lifecycle Management (ILM)? . . . . . 1
- Data Lifecycle Management (DLM). . . . . 1
- Decision criteria for DLM. . . . . 4
- Selling DLM inside the organization. . . . . 5
- Building a DLM Infrastructure. . . . . 6
- DLM value proposition . . . . . 6

Copyright ©Apple Computer Inc. 2004. All rights reserved. Reproduction, adaptation, or translation without prior written permission is prohibited, except as allowed under the copyright laws.

## About TechTarget *IT Briefings*

TechTarget *IT Briefings* provide the pertinent information that senior-level IT executives and managers need to make educated purchasing decisions. Originating from our industry-leading Vendor Connection and Expert Webcasts, TechTarget-produced *IT Briefings* turn webcasts into easy-to-follow technical briefs, similar to a white paper.

Design Copyright ©2003 TechTarget. All rights reserved.

For inquiries and additional information, contact:

Tina Hills

Director of Product Marketing, Webcasts, TechTarget

[thills@techtarget.com](mailto:thills@techtarget.com)

# A Blueprint for Tiered Storage

## What is Information Lifecycle Management (ILM)?

Information Lifecycle Management (ILM) is getting a lot of attention in the press today. More than anything, ILM is a process. Today, and in the past, that process consists of Information Technology staff making conscious or unconscious decisions about what data should reside where in the physical infrastructure and how to optimize for use by the appropriate application at the right time. ILM at present is a concept more than a product. Several products, tools, and services encompass a company's total ILM strategy. ILM is a process in which people, processes, and technologies deal with data from cradle to grave. Data has different value at inception than it does at death, so it shouldn't be treated the same way during each phase of its lifetime.

Today most solutions being sold as Information Lifecycle Management focus on data migration or data movement, as well as the retention and protection aspects of where data resides. This concerns the distinction between Data Lifecycle Management (DLM) vs. ILM. DLM is the infrastructural element and ILM is the macro-level element that assigns both subjective and objective value. ILM relates not only to an application, but also to where the information itself resides in the company as a whole.

DLM—the infrastructural piece of what goes where and when—is very similar to the traditional Hierarchical Storage Management (HSM) that has existed in the mainframe world since inception. The primary difference is that HSM was a single-server, single-storage stack methodology for moving things down (usually) that physical stack in order to lower the cost of primary storage. HSM is still used today in the mainframe world and had limited success in some of the open and traditional world. DLM is HSM in the network world where there are lots of ways into and out of a storage environment from a multitude of potential servers, and each server has a multitude of different ways in and out. HSM was more of a single stovepipe. With DLM you can think of today's network storage world.

The HSM data migration technologies of the past were predicated on purely objective measures. For example, you might have a policy that if you have not accessed a file or type of data in a certain period of time, the data should be moved from expensive primary storage to less expensive tape storage. That works well if all the value is predicated on those objective measures. Unfortunately, today that is not true in most applications. Value is not only objective but it is subjective. Just because access might be limited to a certain data set does not necessarily mean that data set lacks great value to the organization. Conversely, data that you access all the time (such as PowerPoint, for example) may not be the most important application, in spite of the frequency of access.

ILM aims to use the combination of both subjective and objective measures—the data itself combined with overall business policies, (which typically don't exist today) to ascertain where things are located at a particular point in time. Since ILM is for the most part a concept rather than embodied in concrete projects, this balance of this white paper will be confined to DLM.

## Data Lifecycle Management (DLM)

DLM architectures almost exclusively comprise two physical tiers (see Figure 1). There is the physical tier of actual storage where data resides, including disk arrays of all sizes and tapes, and then there is the process tier for moving data between those elements. Today, by and large, the physical tier involves manual processes.

Tiered storage for DLM architectures are not one-size-fits-all. The economic downturn of the last few years has resulted in a sort of technological renaissance. We now have inexpensive arrays and other products that can be just as effective, if not more effective, based on certain applications, than the higher-end products. There are many different attributes: performance, availability, cost, replication, disaster recovery, retention—to name a few. All of



### A DLM Architecture is comprised of Tiered Physical Infrastructure and Migration/Control Software (or manual migration processes)

© 2002 Enterprise Storage Group, Inc.

**Figure 1**

these elements have to be taken into account at various points in time to help you determine the right physical asset that data should reside on at a particular moment.

Figure 2 lists forces driving the DLM phenomenon:

- **Recoverability** has become an important driver for a tiered physical infrastructure as of late. Disk-to-disk back-up, tape-to-tape back-up, and virtual tape libraries are predicated on disk technologies, which are subject to high failure rates. Most companies need to enhance recoverability. The average failure rate on recoverability is 40%. Part of the problem is just the sheer volume of data that organizations have and putting enormous volumes of data exclusively on tape. This doesn't lead to improved recoverability, only to more failed recoverability issues.
- **Reference data growth.**
- **Transaction-oriented data** has been the "lion share" of all the digital data that we have created today. Over the next five years that will be cut in half from a growth perspective, meaning half of all the new data generated is not transactional in nature. It is flat-file data, it does not change. That doesn't mean we need to get rid of it, it is just that the nature of its value from creation to stored facility has changed and as such it does not need to be stored on the primary form of media on which it was written.
- **Economic factors.** Only a few years ago, if you wanted a high-quality storage device, you had to pay a fortune. Now, it is difficult to buy an inferior storage product. If you are a smart shopper, you can find very high-quality products at an inexpensive price point. If your enterprise has \$100 million invested in storage, for example, you probably run core data center technologies like Symmetrix, Lightning, and Shark (see Figure 3). But around that core center, you undoubtedly also have reams of data. Probably the faster growing areas of data reside in the mid-tier (e.g., NetApp, LSI, and Clarion devices). Then you also have very inexpensive disk array technologies that still carry higher-end features and functions, just at a significantly lower price point. Traditionally, if an application was



### DLM Drivers -

- \* Economic - Multiple Classes of Storage
- \* Regulation - Increased On-Line Requirements
- \* Reference Data Growth (150% annually)
- \* Recoverability

© 2002 Enterprise Storage Group, Inc.

Figure 2

- \* Economic - Multiple Classes of Storage

### Example

- Tier 1: Platinum - Symmetrix, Lightning, Shark
- Tier 2: Gold (Fibre) - Netapp, LSI, Clariion, FAsT, HP
- Tier 3: Silver (ATA) - Nexsan, Apple, Dell
- Tier 4: Tape

© 2002 Enterprise Storage Group, Inc.

Figure 3



originally sized and built on an Oracle database with SAP on EMC boxes with storage tech libraries, it would have stayed that way forever. That clearly has an economic impact if all the data residing there only has limited value based on where the company itself is today.

## Decision criteria for DLM

Figure 4 highlights key DLM decision criteria. First and foremost, you shouldn't do DLM because it is the latest cool thing. Does it actually lower the cost of storing your data? If you have depreciated your big high-end systems, or they are in process of being depreciated, and data growth is not occurring, then leave the data alone, as long as you have the capacity. If you need to add capacity, you should evaluate what kind of capacity would make the most sense to add. In the past, companies would make their purchasing decisions once a year and shop for the least expensive storage devices that met their criteria—and the criteria tended to stay stable. Now, the better approach is to ask what category of storage the data resides on and what are the usage patterns in today's infrastructure. As you start to add capacity, you have the opportunity to move data from A to B to C. The

capacity addition should begin the process of evaluating what to store and where.

It is important to ensure you are not creating a bigger problem by adding tiered storage. If adding multiple vendors to multiple different tiers is going to increase the overall complexity of your shop it won't be worth the money you saved. The least expensive storage is not always the best. As with all things, you tend to get what you pay for. Make sure you balance the potential value vs. the management headache it may create.

Incorporating common services does not mean it is required to have one snapshot across everything. What is important is to create business policies such as "we are going to snapshot the data every four hours, we are going to guarantee that we are going to back it up everyday, and we are going to keep seven days online and then we are going to do one day and then we are going pull it to the second tier." Whatever your policies are, those services and processes should be consistent regardless of the underlying infrastructure tier. You don't want to have different ways of doing things for each different tier because you want the same individual that manages tier one to also be able to manage tier five, for example.



### Decision Criteria For DLM

- Lower Cost Of Storing Data?
- Can I Add Tiers Without Creating A Bigger Management Headache?
- Pick The Right Vendor For The Right Tier (avoid fly by night players)
- Incorporate Common Services (Snap, Replication, Etc.) Across all Tiers

© 2002 Enterprise Storage Group, Inc.

Figure 4



You also want to be able to gauge the downstream effects of making changes. If you move this data over here, are there negative implications anywhere else? Does the primary production database have negative implications on that application performance if you do X to Y? Much of the time there will be some relatively obvious examples of things that you can move and lose nothing in the process, meaning no performance implications and no availability. You stand to gain improved availability and lower cost by not only putting some data on less expensive physical storage but you also free up expensive storage media, perhaps enabling a new application to come online.

The trick is not to keep anyone from buying continued capacity—that is a given. It is just to make sure that you are buying the appropriate tier of capacity based on what your needs are at any given point in time. And then, even though this is manual by and large today, there is no reason you can't have ad-hoc discussions about value every 6 or 12 months. Ultimately, we want this to become an automated process where every Monday morning we get a report listing subjective policies and objective policies along

with recommendations for moving data from point A to point B. Someday we will click “yes” and it will happen automatically. This Nirvana does not yet exist. In the interim, you still have to do a lot of manual work.

Figure 5 represents the XYZ company, showing that though we treat data as if it is worth the same value from birth to death, there are different values over any slice of time you might choose to look at. There is no set answer for any of this, only that data value changes over time. Things become more valuable infrequently, they predominantly become less valuable, but they become less valuable at different rates and they become more valuable at different rates. What's important is to remember that things change.

## Selling DLM inside the organization

The single easiest way to sell the DLM concept inside the organization is to sell as capacity is needed, because it is easy for management to understand. Managers understand they have to pay to buy more

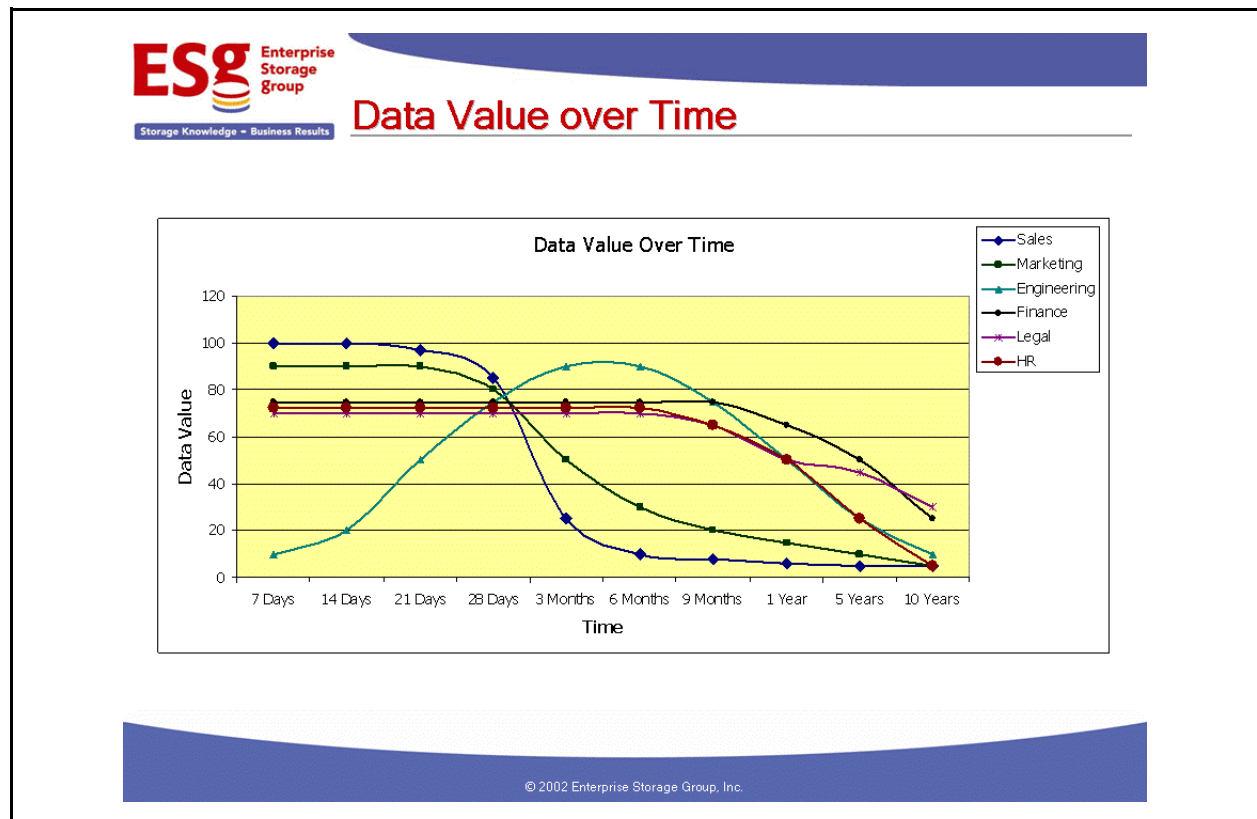


Figure 5



stuff. The question of what stuff is bought and where data is moved becomes more in your purview and it becomes a little easier. The concept of Information Lifecycle Management is a very political one and it is a difficult one to sell in any organization simply because whose stuff is more valuable to the overall organization? Directors of Marketing rightfully believe PowerPoint is the most valuable application in the organization. The CEO may not agree. He might believe the finance system is the most valuable. Sooner or later these decisions have to move up the food chain. Someone up high has to start making the decision that some things are more valuable than others. That becomes difficult with ILM.

With DLM, it is much easier because we are purely relegated to the infrastructure discussion when we start to talk about becoming internal utilities and providing various service levels to the business units. Everyone might want to keep all data on platinum-level storage, back it up every 15 minutes, with a recovery rate of not more than one hour. But you have to be able to associate an exact cost to doing that and be able to provide a menu of services to the business line managers. This level of availability performance redundancy costs \$100 a day, and this level costs \$9 a day. You will also need to educate business managers as to the changes in the market such that many storage devices are more cost-effective while the functionality is much greater. We don't have to treat all data as equal, and we don't necessarily have to charge the same price to everyone. When business managers see a bill and make a comparison, not everybody will buy the Rolls Royce.

Classifying data is another political element. In the DLM world, it's more about the administrative abilities as opposed to the political abilities of the organization—meaning how well you can invoke these objective measures. You can demonstrate that the users have not accessed a file for a certain amount of time, so you will keep it on line for X period of time and they can access it at any time. But since the file is taking up a large footprint on the primary storage system, you will move it to a second tier of storage.

Automation will ultimately be the tool that facilitates the policies and procedures that you create. Automation is the overlay that puts policies into effect automatically. We are quite far from doing that now, though we are starting to automate more and more.

## Building a DLM Infrastructure

Figure 6 outlines the basic steps toward building a DLM infrastructure. First, classify your physical assets. You may find you already have multiple asset tiers. They may not be in the same environment, they may not be in the same physical stand, but you have multiple tiers of assets out there. Figure out which new ones you need to add. If you have 50 terabytes of platinum, 100 terabytes of gold, 100 terabytes on silver, and 10 terabytes of bronze, what do you have in terms of capacity and capabilities vs. what types of data do you have?

Evaluate the low-cost alternatives. A penny is pretty cheap for a megabyte today for high-performance, high-function spinning media. Discuss with business managers the service levels and policies that are available at what price points. No single DLM vendor offers all the related hardware and software. The only approach is to look for best in class. At every tier look for the best in class that meets your requirements within that tier.

## DLM value proposition

At the end of the day, having tiers will save you money (see Figure 7). If values are not constant, being able to change with them from an infrastructure prospective opens up opportunities of moving data from your most expensive type of storage to a less expensive alternative. Another benefit is that DLM positions you for ILM (when it arrives) and utility grid computing environments. Improved recovery is an immediate tactical benefit. The more stuff you have on spinning media, the better your access will be. Recovery times will in fact improve and that in itself can be a reason to justify adding a bronze tier of storage to your environment.

An inexpensive, low-risk way to bring new storage vendors into the mix is to bring in some startup technologies, to lower the cost of keeping data that you have decided is not platinum in value. This gets you into the gain at a low risk and you may find that in five years these technologies become mainstream and replace platinum storage. In the meantime, you don't have any risk and don't carry that burden. This gives you a reasonably inexpensive way to get experience with some of these future technologies with no downside risk.



## Building A DLM Infrastructure

- 1. Figure Out What You Have - Classify the ASSETS
- 2. Figure Out What New Assets You Need To Add
- 3. Buy Cheap - But Not Junk (you can buy rock solid, fast RAID for PENNIES - don't go too far)
- Decide ON PAPER What Your Migration Policies Between Tiers Will Be - Then Find Software That Fits

© 2002 Enterprise Storage Group, Inc.

Figure 6

## What Do I Get From DLM?

- \$\$ - Having Tiers Will Save You Money
- DLM Positions You For ILM and Utility/Grid Computing Environments
- Improved Recovery (If you keep backup data on disk)
- Inexpensive, Low Risk Way To Bring New Vendors Into The Mix - Keep Established Vendors Honest

© 2002 Enterprise Storage Group, Inc.

Figure 7





### **About TechTarget**

We deliver the information IT pros need to be successful.

TechTarget publishes targeted media that address your need for information and resources. Our network of industry-specific Web sites give enterprise IT professionals access to experts and peers, original content and link to relevant information from across the Internet. Our conferences give you access to vendor-neutral, expert commentary and advice on the issues and challenges you face daily. Practical technical advice and expert insights are distributed via more than 100 specialized e-mail newsletters, and our webcasts allow IT pros to ask questions of technical experts in real time.

### **What makes us unique**

TechTarget is squarely focused on the enterprise IT space. Our team of editors and network of industry experts provide the richest, most relevant content to IT professionals. We leverage the immediacy of the Web, the networking and face-to-face opportunities of conferences, the expert interaction of Webcasts and Web radio, the laser-targeting of e-mail newsletters and the richness and depth of our print media to create compelling and actionable information for enterprise IT professionals. For more information, visit [www.techtarget.com](http://www.techtarget.com).

APPLE\_0001\_06/2004

