

READER - EXPERT I / O

# How ILM—and not HSM—improves backups

Q:

What is the best way to integrate ILM so that my backups run more efficiently?

■ BY NOEMI GREYZDORF

Customers often tell me that they have an initiative to implement information lifecycle management (ILM) to make their backups more efficient. When customers are asked how they think ILM will help them solve their backup issues, the response often describes what I have always known as hierarchical storage management (HSM). Backup systems are complicated; there are a lot of moving parts, and bottlenecks can exist anywhere along the data path, including the client server itself, the network, backup server, and the final destination, be it tape or disk. Understanding the bottlenecks is the first step in addressing the challenges associated with backup systems. Once the backup system has been tuned, there are other steps, such as the application of ILM, that organizations can take to make their backup system more efficient.

## What is ILM?

In the past year or so, ILM has become one of the hottest acronyms in the storage industry. The definition varies based on who is talking, so to have a meaningful discussion about ILM it is critical to start on common ground in terms of understanding what ILM means.

The biggest misconception is that ILM is a technology. Rather, ILM is knowing what data you own, what it means to the success of your business, and what its relevance is as it ages. In other words, if you created a file today and the information in the file is critical and relevant for the success of the business, you want to ensure the file can be easily accessed and is protected against corruption, deletion, or disasters. In a year, the same file will not be as critical to the organization and the content might be out of date. Although it is still important to the organization as a reference or a record of activities, the file might not require the fastest, safest storage because the impact on the business if the file is lost is not as great. Understanding the lifecycle of the file can help you make the decision to move the file to an archive. In effect, archiving static data can help free up higher-cost storage capacity and reduce the size of backups.



Unfortunately, the term “ILM” has been hijacked by HSM vendors to capitalize on the buzz around ILM. Users who have bought in to this definition have focused too much attention on the idea of migrating data from expensive drives to cheaper drives and

## How ILM helps

Implementing true ILM means understanding data, its value, how it is created and by whom, and how the organization uses the data over time. Going through this process enables organizations to categorize data by users and types and assign service levels for storage and backups. Understanding data helps administrators make decisions, such as

- Do we need archiving?
- What data falls under regulatory compliance requirements?
- Who is considered to be a business user?

## How HSM doesn't help

In turn, deploying ILM only as HSM can wreak havoc on your backup system. Here are some of the reasons why HSM can be a problem for backups:

- If a 40KB file is replaced with a stub of 2KB to 4KB, the space saved is significant, potentially more than 90%. A file system full of these files can thus be reduced, making your backups go faster and using less backup media. Unfortunately, the capacity and performance gains don't outweigh the costs of implementing HSM;
- A file system with a large number of files will not be improved by replacing larger files with small file stubs. Backing up small files is actually harder than backing up large files from a performance perspective. If you are backing up directly to tape, keeping the tape drives spinning will be a challenge with lots of small files. Backing up to disk

won't help since the bottleneck is most likely at the file system level on the client server; and

- Storage optimization is often listed a benefit of HSM. A file is moved to secondary, cheaper storage, thus freeing up space on the more-expensive storage device. The reality is that the costs associated with deploying HSM will offset the benefits of using cheaper disk. Besides, data moved to cheaper disk still has to be backed up; therefore, the efficiencies gained are more theoretical than real.

## Recommendation

Returning to the original question regarding the best way to integrate ILM to help with backup, note that ILM, if applied properly, can help streamline backups through better understanding of data, including how it is used, by whom, and for how long. Decisions can be made regarding what data is backed up and when, and what media is used to store the backup images.

Applying ILM doesn't mean deploying HSM. Before deploying HSM, consider the actual costs associated with the solution, confirm whether it will help with the backups (e.g., will the backups go faster?), and always keep in mind what it might take to extract yourself from this in a year or more. A file server with lots of small files will create a bottleneck on the file system, in which case HSM will not help at all but might actually hurt. If you are backing up stubs, how do you actually recover the data in case of a deletion or a corruption? Since the data is not actually on the file server, first a stub would be restored, which would then point to the place where the actual data is located.

In the end, always try to first identify what the issues are in your backup system that you are trying to address and what objectives you need to achieve; then understand your data and how it is used. Only then will you see true gains. □

.....  
**Noemi Greyzdorf** is a senior solutions consultant at Cambridge Computer ([www.cambridgecomputer.com](http://www.cambridgecomputer.com)) in Waltham, MA. She can be contacted at [noemi@cambridgecomputer.com](mailto:noemi@cambridgecomputer.com).