

What is the Difference: HSM vs. Storage Management Software?

By Bob Cone, Spectra Product Evangelist

Introduction

Words matter... so do acronyms. We've seen a new category of storage management software developed which is being referred to as, you guessed it, "Storage Management Software." Some have questioned whether this is something new or if it is just another offering of HSM (Hierarchical Storage Management). There are differences between the two categories. This isn't just about product positioning or marketing; it's about what certain products do or don't do and whether or not they are a fit for your data center or a specific need. Both solution types are designed to keep frequently accessed data on the fastest storage (usually the most expensive and least dense) while moving inactive data to slower storage (usually much less expensive and significantly more dense). How the solutions accomplish this is very different.

Transparency Matters

A common vocabulary is critical to communication, so I wanted to lay out a few terms vital to this discussion: Stub Files, Symbolic Links and HTML Links. These are the "bread crumbs" left behind when a file is



moved. This is how the application or user is going to get the file back as transparently as possible. And as you will see, the level of transparency differs based on methodology and is a key differentiator between HSMs and Storage Management Software.

HSMs try to map nearline storage to appear to be online, including tape storage. This is tricky given the sometimes long latency of nearline storage. Applications will timeout if they don't get the data requested in the time they expect to get it. To pull this off, HSMs typically use stub files (along with filter drivers, but we don't need to go that deep). The stub file looks like the original file and often contains the beginning of the original file. The stub file can respond to the read request while the HSM gets the rest of the moved file and brings it back – no small feat! HSMs become part of the file system, usually have kernel code, are very operating system (OS) dependent, have to be upgraded with the OS, and offer no ability for the application or the user to see that a file has been moved.

For the above reasons, HSMs don't play well in all environments. They *do* play well in [High Performance Computing](#) (HPC) environments where the operating systems – like Lustre, GPFS, etc., are more "timeout tolerant" or "HSM aware." Successful HSMs like IBM's HPSS and HPE/SGI's DMF are extremely effective. They are also expensive, complex and require a lot of resources – but when you need them, nothing else will do. In contrast, HSMs introduced for the general IT market were *not* successful. Many HSM products were introduced from the late 1990's through the mid 2000's. Few if any of those HSMs exist today.

Generally speaking, the solutions we see today in the category of Storage Management. Software require less budget, headcount and infrastructure, and sit well outside of the file system. There are some exceptions, but this is a good categorization of [modern Storage Management Software](#). These packages are much less complex and much more compatible with a large range of applications and use cases. Symbolic links and/or HTML links are more likely to be used to find the moved data. These links work quite differently from stub files as well as from each other.

By leaving a symbolic link in place of the original file, a read can simply be redirected to where the file has been moved. This works great when moving infrequently accessed data off of primary storage (high speed disk or SSD) to a lower tier of storage like NAS disk. Most applications can tolerate the small increase in latency. However, this methodology does not work well with tape or low response level cloud. That's where the [HTML links](#) come in.

Solutions that support HTML links are created specifically to support storage mediums with longer latency, such as tape or low-response cloud levels. When an [HTML link](#) is left in place of the migrated file, the user is presented with an HTML page which states that the file has been archived, gives information about when, where and how this was done, and allows the user to start a restore from tape or a recall the file from cloud without having to contact IT. So transparency exists, but as noted earlier, it is a significantly different level of transparency (with HTML links) and offers distinct advantages over those derived from HSM solutions.

A Focus on Data Protection

The final differentiator between HSMs and Storage Management Software relates to [data protection](#). Storage Management Software is designed for both primary storage offload *and* data preservation. That has not been the focus of HSM solutions. Storage Management Software allows multiple copies of the original file to be created and stored during the data migration process. In addition to moving infrequently accessed files from expensive primary storage to lower cost NAS, savvy users can also send a copy to an on-



premise tape library, to offsite tape storage, and to cloud for disaster recovery.

In this way, Storage Management Software can decrease the size of the Primary Tier of storage (increasing performance and decreasing backup windows) while also assuring the preservation of moved/archived data via multiple copies to multiple mediums.

While the lines are sometimes blurred between HSM and Storage Management Software, as well as between individual solutions that fall into each category, there are numerous differences that are important to identify and consider when implementing a data tiering solution for the specific needs of a data center. Developed from the ground up to address the problem of managing and storing large data sets, Spectra's new storage

management software, [StorCycle™](#), includes features not found in other data management solutions. Project-based archiving, which allows users to migrate grouped data related to specific projects or machine created data; HTML linking, mentioned above; and pricing models which

keep the cost of software in line with lower hardware costs are all unique approaches which further differentiate StorCycle even within the category of Storage Management Software. Learn more at spectralogic.com/products/storcycle/.

About Spectra Logic

Spectra Logic develops data storage and data management solutions that solve the problem of digital preservation for organizations dealing with exponential data growth. Dedicated solely to storage innovation for 40 years, Spectra Logic's uncompromising customer focus is proven by the adoption of its solutions by industry leaders in multiple markets globally. Spectra enables affordable, multi-decade data storage and access by creating new methods of managing information in all forms of storage— including archive, backup, cold storage, private cloud and public cloud. To learn more, visit www.SpectraLogic.com.

Copyright ©2019 Spectra Logic Corporation. All rights reserved worldwide. Spectra and Spectra Logic are registered trademarks of Spectra Logic. All other trademarks and registered trademarks are property of their respective owners. This document is provided for information purposes only, and the contents hereof are subject to change without notice. All opinions in this white paper are those of Spectra Logic.



SPECTRALOGIC.COM