Product Spotlight
Realizing the Promise of HSM - Komprise

Storage Switzerland, LLC

Joseph Ortiz | Senior Analyst

Organizations have always struggled with the cost and complexity of maintaining sufficient storage for all their data. Instead of getting better, the problem just gets worse. Those of us who have been in IT for longer than a decade probably remember when enterprising vendors of the day decided to borrow a page from the mainframe systems playbook and started coming out with hierarchical storage management (HSM) software designed for distributed server environments. HSM software was supposed to automate the process of identifying cold data sets and automatically migrating them from primary disk to less expensive optical and tape storage devices of the day. It was also supposed to handle file recall requests automatically whenever a user clicked on a stub file.

The HSM Letdown

Unfortunately, these early HSM products suffered from a number of deficiencies such as:

- They were custom designed for specific proprietary storage systems, which limited hardware choices and resulted in vendor lock-in.
- Many required file server agents that required substantial memory and compute resources, and operated in the direct data path, impacting performance.
- They used static stub files left in place of the moved data. These static stub files could be corrupted, deleted, and orphaned making it difficult if not impossible to locate the original source file.
- The early HSM solutions did not scale well. As file counts increased, HSM performance deteriorated significantly since they were traditional database-driven architectures.
- The solutions would disrupt storage system performance, interrupting active usage.
- File recalls could take a long time, especially if the requested file was stored on tape.

So bad were these deficiencies, that HSM became a "bad word" amongst IT professionals. Many of those IT pros believed that the only viable way to manage storage was to just keep adding more capacity to the primary tier.

The Deteriorating Data Condition

Over the last few years, the data center landscape changed significantly with organizations having a wide range of storage options available. Flash memory devices have replaced high performance physical disk drives as Tier-1 storage. High performance and commodity physical hard disks now function as secondary and tertiary storage tiers. Cloud storage and object storage options are available to handle large bulk, long-term storage requirements. All of these options are needed to combat the data onslaught that most organizations are facing.

However, the main problem remains; how to automatically detect "warm" and "cold" data sets then migrate them to the most cost effective storage tier while also managing the entire file life cycle. In short, we have more storage options than ever but less intelligence about how and when to move our increasing data to which storage platform.
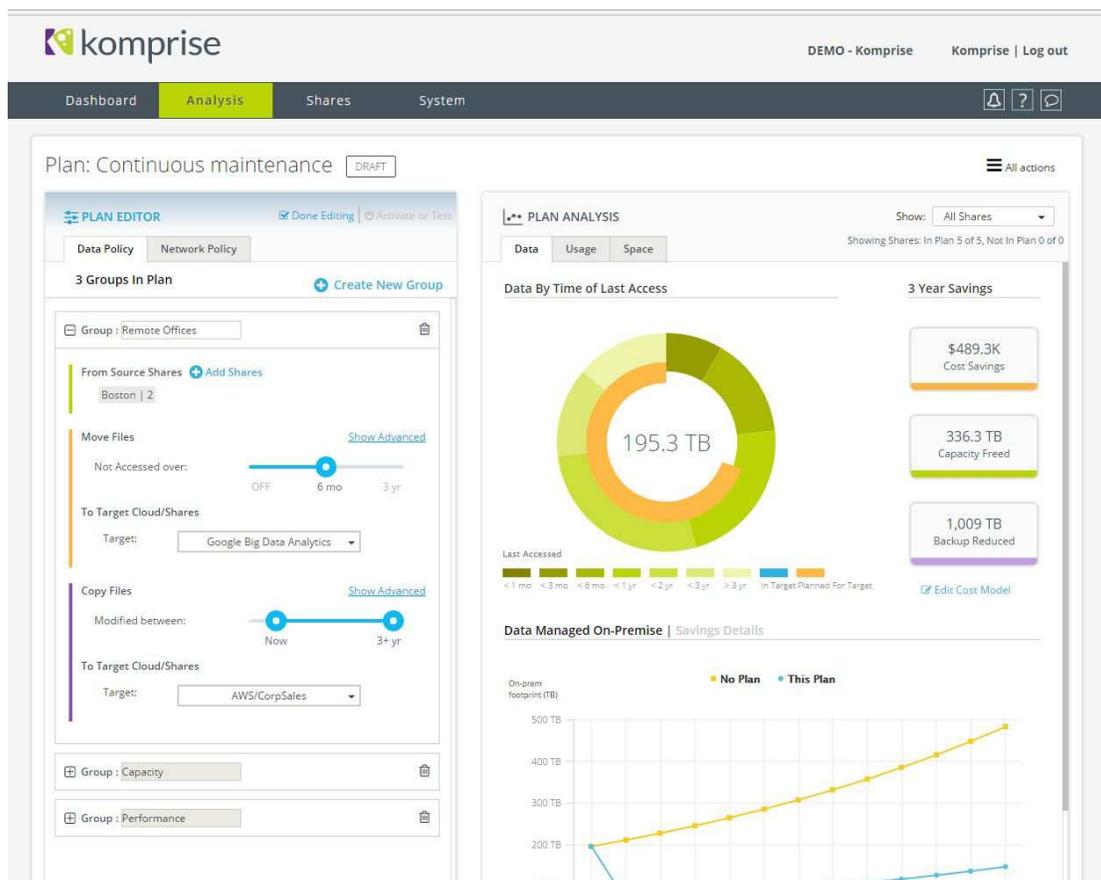
Komprise Delivers the HSM Promise

Komprise is a company that addresses these challenges with its new data management software that uses a unique analytics-driven, adaptive automation solution which can transparently manage massive amounts of data across all of an organization's storage silos whether on premises or in the cloud. With a modern architecture built from the ground up to handle massive scale of data with intelligent automation, the simple to deploy solution is hardware agnostic and consists of two main components, which are a "virtual Observer" and a "Director."

The Director can run as either a cloud service or on-premises. The virtual Observer piece runs in a virtual machine (VM) on premises. As more data needs management, you simply add more virtual observers to meet the load. Komprise works seamlessly across any on premises NFS, SMB, and object/cloud storage without requiring any storage agents. It runs in the background out of the active data path. This is an ongoing background process rather than a real time process and adaptively throttles back when the file system or the network is in use.

Once active, Komprise creates a control plane that sits to the side of the active data path and manages data and storage efficiently while eliminating the need to add new silos to an existing environment. The virtual observer works by "crawling" the file system and tags both on-premises and cloud based data. The information gathering process has a minimal impact on CPU and I/O operations and throttles back as needed so it is invisible to active users. The analytics engine provides extensive metrics on all data such as data growth rates, data locations and storage types, as well as which data users regularly access and which they do not, and which data has protection or not. When files are migrated, Komprise uses symbolic links that are dynamic and always point back to a virtual access address that is disintermediated by one of the Komprise virtual machines, with the metadata stored at the target. If a symbolic link is deleted, Komprise detects this and replaces the link based on policy. This provides resiliency and enables ongoing lifecycle management of the data so it can be moved again without any changes to the link.

Komprise also manages data recalls transparently by first placing the recalled file in its own cache. If users repeatedly call for the recovered data, it will automatically move the data back to its original location. From the user's perspective, it always appears as if the data never left its original location. Recalls back to the source file system are also policy controlled.

Komprise has an easy to use single pane graphical user interface that lets users create complex, custom policies governing file migration, copies, retention periods, target locations and a wide range of other features and options to meet almost any need.

StorageSwiss Take

Komprise is a powerful and flexible solution that provides all the features that were the goal of early HSM solutions. It provides automated and effective management of data throughout its entire life cycle. With an analytics engine providing extensive metrics on all data along with the ability for the user to run "what if" simulations to gauge the effects of various data policies before they are implemented as well as recommendations on data placement, Komprise is a solution any organization today should seriously consider.

About Komprise

Komprise, the first intelligent data management service, empowers businesses to efficiently manage today's massive scale of data growth while unlocking its value. The Komprise mission is to radically simplify data management through intelligent automation. Komprise is used by enterprises across industries to get visibility into their data and cut storage costs by intelligently managing data at scale. For more information, go to www.komprise.com.

About Storage Switzerland