ActualTech Media

DATACORE

Attain Efficiency and Cost Savings Through Automatic Load Balancing of Your Distributed Files

Trevor Pott

CONTENTS

IN THIS PAPER

Explore the benefits of file storage virtualization with vFilO. This paper shows how transparent file-level load balancing allows previously unattainable storage system efficiency without compromise.

Profitability is a careful balance between controlling costs and driving growth, and this applies to every aspect of your business. Focusing too much on managing costs can hinder growth, but it's also all too easy for costs to spiral out of control. This is particularly true with regard to data. Every organization runs on data, and what's needed is a solution that threads the needle between data storage agility and prudent frugality.

IT departments often have multiple file storage systems (such as NAS devices, file servers, Windows/Linux shares, etc.), some nearly at capacity and others with capacity to spare. The same holds true of performance: While one storage system might be under intense load, others may be almost entirely idle.

Stranded capacity and stranded performance lead to inefficiency, but solving either isn't easy. It's no simple task to move data to where it's most useful. Organizations may have filers from multiple vendors, and even where only a single vendor is in use, the majority of today's filers are simply not designed to relocate folders/directories between them without impacting end users and applications.

What's needed is a virtual, multi-site file management alternative. DataCore vFilO is a software-defined storage (SDS) solution that intelligently migrates files based on administrator-defined policies in order to optimize an organization's entire storage fleet for cost, space, performance, governance, or other preferences. To accomplish this, vFilO takes advantage of a file's own metadata in such a way that's completely transparent to end users and applications.

The future is unpredictable, and this presents something of a problem for those who are tasked to plan for it. Different types of data are used in different ways, and most, but not all, data has a shelf life—a date past which it's likely never accessed again.

The access patterns of both people and machines not only change and evolve, they move physically among storage systems, and even geographically among data centers. Managing change is an essential part of managing an organization's storage, and this includes taking advantage of new capabilities, such as the ability to archive cold and inactive data to public cloud storage or other lower-cost object storage options on-premises.

How Storage Gets Used

The access patterns of files generated by humans are relatively predictable. Most are accessed a given number of times before never being touched again, except by backups and (possibly) audits. When you think about it, this makes sense: Humans work on an initial draft of something, circulate it for review, make revisions, and then make it generally available in some fashion.

> The future is unpredictable, and this presents something of a problem for those who are tasked to plan for it.

For the overwhelming majority of human-generated data, the useful lifespan of any given file is less than six months, and many files end their useful lifespans far earlier. But though this is a useful rule of thumb to understand data lifecycles, it's not an absolute.

Complicating matters is that file access patterns are often wildly different for machine-generated data, especially log files, such as those generated by Internet of Things (IoT) sensors and other automated sources of data. This makes traditional file archival approaches difficult: Administrators can't simply write a script that says: "Move any file older than six months off primary storage and archive." And if you did need to restore one of them in the future, good luck finding it.

Files—regardless of origin—may have to remain accessible to humans and machines indefinitely. Others can likely be managed automatically. With no concrete rules for administrators to follow in their attempts at automation, each file has to be considered independently.

Unfortunately, even small businesses generate far too many files for humans to manage manually. This leads administrators to triage their archival efforts, automating the trimming of predictable or rapidly expanding files (like log files). It's also common to allow generous time buffers for archival scripts, often leaving stale files on expensive primary storage for the maximum legal hold time, such as seven years.

All of this has an economic impact. Not all storage costs the same—either to acquire or to operate—nor do all storage systems provide the same level of performance, durability, or availability. Stale files occupying expensive primary storage are simply wasting money, but most organizations have little choice but to allow this.

Change and Other Nightmares That Keep Sysadmins Awake

The ability to respond to change can determine whether an organization succeeds or fails. Change isn't predictable. Years can pass with nothing fundamentally changing at all, and then, all of a sudden, all the rules go out the door overnight. When change impacts an organization, changes in data access patterns are likely to follow.

It's fairly common, for example, for one department's needs to grow faster than predicted, leading to them "borrowing" space on the other departments' filers. Similarly, a need to stand up new services can lead to the acquisition of new NAS that hadn't been budgeted for.

The capacity and performance to support the new demand may already exist, but not be logically separated in a way that's useful. Or the capacity and performance may be available, but distributing the load appropriately among multiple systems might be difficult or impossible today.

The bigger your organization (or at least the number of data sets), the bigger the problem. Complicated access patterns make it more complicated to achieve storage system efficiency, and automation becomes an absolute necessity in order to control costs. But file storage automation hasn't been easy, especially across diverse filers. Nevertheless, past a certain size or complexity point, automation is non-negotiable: Adding more systems administrators doesn't allow you to manage more systems/data/and so on, and automation is the only thing that makes operating at scale possible.

But automation has been constrained based on files always being where various applications and users expect them to be, which can lead to voluntary inefficiencies in an organization's storage systems because, as previously noted, automation and orchestration of archival data is difficult.

Preparing for the Unknown by Solving Today's Problems

For most organizations, meeting today's unstructured data storage needs means overbuying storage. You know growth and change will happen before the end of a given storage system's life, but the details of what kinds of change, the timing of the change, and the velocity of change are unpredictable. Today's storage problems are partly the result of not making the right guess when overbuying during the last cycle, and unless something changes, tomorrow's will be the same.

> Complicated access patterns make it more complicated to achieve storage system efficiency, and automation becomes an absolute necessity in order to control costs.

How organizations attempt to solve these problems varies considerably. Some organizations choose to invest in storage analytics with an eye toward buying new storage systems as needed. Others may decide to alter how or where data is stored to make the best of what they have available.

Buying only what you need as you need may seem prudent, but it can be problematic if global supply chain disruptions occur. This can result in paying a premium for storage, or even make it difficult to physically obtain enough new storage to meet demand for your most critical workloads.

Carefully curating existing storage won't solve data preservation or legal hold requirements, and ultimately requires a place to archive data to. Regardless of your approach, sometimes you just need more storage than you have, which makes public cloud storage and/or on-premises object storage an important consideration for today's storage teams.

Public cloud storage and on-premises object storage are most frequently deployed as low-cost, but also low-performing storage. For this reason, they're an excellent target for unused and infrequently accessed files. Because the files stored on these storage classes by vFilO are so infrequently accessed, the low performance of the storage systems themselves is less impactful; fewer simultaneous accesses means less demand, allowing files stored on these systems to be accessed with reasonable rapidity in the rare cases this is required.

This demand profile also means that the files stored on these systems can be deduplicated and compressed, further reducing the amount of storage space consumed. In the case of public cloud storage this is especially useful because that storage is often rented on an as-needed basis. This can dramatically reduce file storage costs, especially when compared to storing all files on expensive fast file storage systems.

This is another place where DataCore vFilO comes in. vFilO supplies organizations with automation that meets today's data storage management needs while providing the agility needed to adapt to change. With vFilO, organizations can break free of the storage overbuying cycle today, and in the process build a more resilient, responsive, and adaptable file storage virtualization solution that will keep costs down tomorrow.

The Solution Is Software-Defined Storage

vFilO software-defined storage services solve all of these file storage problems. vFilO automatically moves files between existing filers and cloud/object storage based on various data placement criteria determined by the organization. Individual files can be moved to different storage systems based on how data is accessed, as well as its relative value at the time. Most importantly, the movement of these files is completely transparent to both end users and machines: A file still appears to be at its original location, even if it's been physically moved halfway around the world.

vFilO allows administrators the flexibility to define policies based on business requirements for autonomic data placement. Providing file-level and volume-level granularity, vFilO can automatically move designated inactive data to lower-cost storage, such as object store on-premises or any elastic cloud.

- Move data based on data access temperatures, i.e., when files were last accessed
- Move a specific type of file based on its format (e.g., snapshots) or size criteria (e.g., video files), age (e.g., files older than a four years) to pre-defined storage tiers
- Make copies of data for backup and move it to cold storage

This is transformational because, until now, the filer where you stored data constrained the directory/folder hierarchy as well as the level of service you could expect for the entire share. All data gets the same expensive treatment whether it needs it or not. It also anchors workloads to specific hardware, which in turn can affect organization-level IT architecture, such as WAN topology. And this can place limits on how data can realistically be used, ultimately impacting an organization's ability to respond to change.

DataCore vFilO allows customizing data placement policies based on its value. The dynamic choice of location is influenced by an organization's availability, performance, resilience, and compliance goals. Combined with the capabilities to reduce storage space via deduplication and compression techniques especially for cloud-migrated data, vFilO helps reduce storage costs over time, make IT operations more efficient, and deliver maximum value for money for where you decide to store your data.

See how DataCore vFilO can help you take control over your file and object data storage and automate load balancing for greater operational efficiency and cost savings. Check out this self-paced hands-on lab to put vFilO to the test: www.datacore.com/vfilo-lab.