



Life Sciences

Qumulo's Modern Scale-Out Storage Helps Researchers Combat Global Disease

A leading global health research organization found itself with little visibility into its massive data store, relying on days-long manual reporting procedures to try to stay on top of its escalating storage usage. Qumulo's real-time analytics quickly showed them exactly what they were missing, providing the scalability and operational visibility necessary to efficiently conduct and store this life-saving research.

Graduating from "Terabytes" to "Petabytes"

Compilation and analysis of health statistics on a global basis is in many ways the very definition of 'big data' – with research teams often generating tens of millions of files in a single afternoon. Keeping up with that massive data growth, and more importantly, understanding its constant ebb and flow, is an equally massive challenge.

That's exactly the situation faced by the IT team at one major U.S. university's independent health research center, which tracks population health trends, including the global impact (or burden) of 400 different diseases in over 180 countries worldwide. In essence, the center collects data, by country and geography, to determine why people are dying, and uses that insight to recommend effective policies, treatment strategies, funding allocation and more.

Seven years ago when the IT director was hired, the storage capacity at the organization was 1.5 TB of disk-based storage, and growing fast. "It seems every year I'd estimate that we needed a certain amount for storage, and it inevitably ended up really being five times as much," he said. To help make that growth manageable, the organization implemented a first generation scale-out NAS system, creating a single shared "bucket" that was infinitely expandable by adding more nodes, without relying on multiple mount points or separate file shares.

The problem is that with essentially infinite space comes the creation of seemingly endless amounts of data. And as that data soon claimed petabytes of capacity, the ability to analyze usage became increasingly difficult.

"Understanding data usage was the thing that plagued us the most. With the real-time visibility I saw in Qumulo, that was clearly going to be a thing of the past."

— IT Manager, Independent Global Health Research Organization

"It's easy to see how much space is available, but when you have hundreds of millions of files in a given volume, it takes days or weeks to run a single report on storage usage. And god forbid you get the report and find you've overlooked some statistic, because you'll have to start all over," notes the IT director. "Sometimes it seems like running reports is all we do."

Complicating matters is the fact that different research teams are assigned specific storage budgets and IT needs to support all of these different groups. So determining 'who is using what' becomes a very real dollars and cents calculation – and conversation. A conversation that's hard to have with limited visibility into the data itself.

Solution Overview

- 11 Qumulo QC24 Hybrid Storage Appliances
- 6 Qumulo QC208 Hybrid Storage Appliances
- NFS and REST protocols
- Qumulo Care enterprise support

Key Benefits

- Delivers real-time visibility into data usage for better forecasting, increased efficiency and improved user communication
- Saves days and weeks of valuable IT time by eliminating manual reporting
- Scales to support massive petabyte-level data growth around new research initiatives
- Eliminates throughput and IOPS performance bottlenecks
- Simplifies IT infrastructure through elegant architecture and cost-effective use of commodity hardware
- Streamlines support for fast root-cause problem resolution
- Identifies low-priority files for easy archiving

And because this is critical data on global health trends over time, it's necessary to keep it all. But it doesn't all have to be on primary storage. Unfortunately, differentiating live research data from what can be readily archived is another time-consuming manual analysis and reporting process.

According to the IT director, "It was relatively easy to keep adding servers and storage, but that's not cost-effective and it was rapidly growing unmanageable. I kept wondering 'what are we missing?'"

With Qumulo, Seeing is Believing

Ironically, the answer came in an email. "I first learned about Qumulo through my inbox, and the email caught my eye because like any former database guy, I tend to obsess a bit about storage. After taking a look at the company's impressive pedigree, I thought 'okay, a quick demo can't hurt.'"

"We had reached the point where IOPS was our major hurdle, and until we installed the Qumulo solution I couldn't even have clearly said that."

— IT Manager, Independent Global Health Research Organization

Qumulo offers a modern approach to scale-out storage, delivering fast, flexible and highly scalable storage together with the real-time analytics necessary for visibility into data usage and performance at petabyte scale. And a demo of that system blew his socks off. "Qumulo was building and managing a massive fileshare live right in front of me, showing real-time analytics on everything that the storage was doing in the process: performance workload, real-time activity, who was generating what file, how much capacity it was using, what the storage was doing in terms of IOPS. Everything. I didn't even know that was possible at petabyte scale, but I immediately knew we needed it. And that it would have saved us an entire summer of running reports."

Moreover, he knew it would be a huge boon for the institute. "Understanding data usage was the thing that plagued us the most. With the real-time analytics I saw in Qumulo, that was clearly going to be a thing of the past," he says.

'Here's how much you're consuming, and it's not what we budgeted.'

The institute installed eleven Qumulo QC24 hybrid storage appliances in a scale-out storage cluster with a single file system, and can handle future capacity and performance expansion simply by adding more nodes. Not only does this provide the massive scalability needed, but the real-time insight to understand how the data and that storage is operating.

"We had reached the point where IOPS was our major hurdle, and until we installed the Qumulo solution I couldn't even have clearly said that. Now not only do we have performance and scalability, but we could go over 120,000 IOPS without any problems at all," he explains.

More importantly, the IT team has the visibility needed to ensure efficient operation. Now they can go to project groups, pull up the Qumulo dashboard and literally show users, 'Here's how much you're consuming and it's not what we had budgeted for your storage.' Using these real-time analytics, the organization can determine which teams are generating more data than anticipated, identify if something is awry, and adjust or adapt as necessary. It's even possible to streamline this process using the REST-API to create a customized user dashboard for easy access to the most common or relevant data queries.

This same visibility also makes managing the capacity much easier, as the team can quickly identify what data is being used – and therefore needs to stay readily available – and what can be safely archived. All of this saves the organization significant time, energy and money.

Managing file metadata has also become more efficient with Qumulo, "reducing complexity in our overall infrastructure," the IT director noted. And the plug-and-play nature of the Qumulo Core software further simplifies the center's architecture, running on cost-effective, commodity hardware.

Moreover, Qumulo exceeded all expectations in terms of customer support. The IT team was especially impressed with the streamlined, down-to-business nature of the Qumulo Care dedicated customer support managers.

"IT teams don't want to spend hours troubleshooting a problem by going through scripted Tier 1/Tier 2 minutiae like 'Is your computer on?'" explained the director. "Going through that rigmarole is really irritating, especially when you have talented IT staff that knows better." Fortunately, the Qumulo Care team always gets straight to the root of the issue.

The center's IT team has previously worked with many large vendors, but few have earned the same praise as Qumulo. "In fact, my team sometimes actively discouraged me from buying other vendors' solutions because customer support is so poor," he says.

By contrast, the experience with Qumulo has been so positive that the institute has purchased six additional QC208 hybrid storage appliances to support a critical new forecasting initiative – one that promises to generate multiple petabytes of data going forward. As the IT director notes, "This new initiative is ambitious, and from a data standpoint promises to consume tons of storage. Securing the necessary funding required showing not just scalability, but demonstrating an ability to efficiently handle that crucial data. And both were addressed by our Qumulo storage cluster."

With its new, modern scale-out storage solution in place, the organization is now positioned to carry on with its mission to provide the world with the most meaningful and actionable population health data that ultimately saves lives on a global basis. Qumulo was indeed the cure for data storage limitations in this research, and the only side effects are scalability, efficiency and vastly improved real-time analytical insight into their data and storage.

