Life Sciences

Truly Visionary: Research Institute Cuts Image Processing from Months to Days with Qumulo's Modern Scale-Out Storage

Storage I/O bottlenecks were dramatically slowing imaging projects for the University of Utah's SCI Institute, while lack of insight into data usage hampered effective capacity management. Qumulo's modern scale-out storage software changed all that, cutting processing times from months to days, and delivering the analytics necessary to make smarter decisions about data management, ultimately turning black-box storage into a highly visible resource.

Massive Data Files Deliver Massive I/O Bottlenecks

The University of Utah's Scientific Computing and Imaging (SCI) Institute is an internationally recognized facility specializing in visualization, scientific computing and imaging analysis. And like any research organization that deals with scientific imaging, it's confronted by massive data files – and equally massive processing and capacity challenges.

"Our projects are much too large to fit in memory, so I/O performance is critical to our ability to quickly process data sets," observes Nick Rathke, Assistant Director, Information Technology for the SCI Institute.

Take one recent project as an example: The SCI Institute, which focuses heavily on the medical field, conducted a cutting-edge analysis of rabbit retinal structure as an analogue to human retinas. Using electronmicroscopy, thin slices of the retina were scanned, and then assembled into a high-resolution, 3D image for researchers. Yet image processing on an original 20TB data set like this typically took upwards of three months or more.

"We have massive parallel supercomputers that are starved for data because we can't get anywhere near saturation on our disk I/O," notes Rathke. "We've purchased scale-out NAS systems that expand capacity, but not performance, so we're stuck with architectural decisions made years ago based on much smaller data sets."

At the same time, as the size of imaging data sets have grown, so too has the impact on storage capacity.

"When we run out of capacity, the direction from higher up is inevitably 'just delete old data'," Rathke says. "But which old data? There's a big distinction between data that's old and data that's important, and I can't tell which is which without running lengthy manual reports."

> ***"Other vendors claim to 'scale up' storage, but there's a fundamental difference with Qumulo: the ability to linearly scale capacity and performance. Every time I add a node I'm adding network bandwidth and I/O operations, not just a tray of disks, and that's critical for us."***
>
> — **Nick Rathke,**
> Assistant Director,
> Information Technology
> SCI Institute

And given this lack of visibility, there's also no way for Rathke's team to work directly with users on storage management. "I can't easily tell them how much they're using, I can't dispute the importance of a file that hasn't been touched in years, I can't track allocations – it's an extremely painful process."

SCI
www.sci.utah.edu

Solution Overview

- 4 Qumulo QC24 Hybrid Storage Appliances
- NFS and REST protocols
- Qumulo Care enterprise support

Key Benefits for SCI

- Slashes image processing times from months to days.
- Eliminates throughput and IOPS performance bottlenecks
- Scales performance linearly along with capacity
- Delivers real-time visibility into data usage for intelligent archiving
- Allows pay as you grow, modularly expanding capacity and performance through non-disruptive addition of new nodes

In short, the SCI Institute's massive data files and high performance compute cluster demanded an equally high performance storage solution, delivering large, shared single pool capacity, multiple parallel processing paths and the visibility necessary to understand what was happening with all that data.

"Cheap and cheerful storage, like the kind we were previously relied on, just doesn't cut it anymore in our data-rich environment," Rathke explains.

Scale Means More Than Just Capacity

To meet its needs, the SCI Institute implemented the QC24 hybrid storage solution from Qumulo – 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.

"Other vendors claim to 'scale up' storage, but there's a fundamental difference with Qumulo: the ability to linearly scale capacity and performance at the same time," says Rathke. "Every time I add a node, I'm adding network bandwidth and I/O operations, not just a tray of disks, and that's critical for us."

As importantly, this 'pay as you grow' strategy is built into the design of Qumulo's appliances, so new nodes automatically and non-disruptively join the cluster to expand the SCI Institute's single pool of storage – adding scale without adding downtime. And, that's a real benefit with grant-based research funding, giving Rathke's team the ability to incrementally fund and expand storage capacity and performance as needed.

Turning Storage from Black Box to Precious Resource

The move to Qumulo's modern scale-out storage paid immediate dividends, slashing processing of image data like the rabbit retinal scans from 11-14 weeks down to only nine days. And, software enhancements delivered through Qumulo's agile two-week development and release methodology are shortening those processing times even further, as is the performance improvement each time a node is added to the Qumulo cluster. This ability to scale I/O performance is a critical enhancement to the SCI Institute's productivity, allowing the central storage cluster to finally keep pace with the performance of the HPC server cluster.

> ***"It's impossible to make smart decisions about data when your storage is a figurative black box. Qumulo lets me tell in an instant how the data is used, who touches it and how often, so storage is no longer a technical issue but a management decision."***
>
> — **Nick Rathke,**
> Assistant Director,
> Information Technology
> SCI Institute

But the real eye-opener for Rathke was Qumulo's real-time data analytics, which deliver the instant visibility needed to bring users into a conversation about storage resources. "Everyone thinks their data is critical, but now when someone disputes the importance of a project that's due to be archived, I can quickly pull up the dashboard to show that it hasn't been touched in years. That adds clarity into an otherwise murky storage decision."

Conversely, the real-time visibility enables Rathke to also show that sometimes, even though a file is old, it represents a data set that still gets used regularly.

"It's impossible to make smart decisions about data when your storage is a figurative black box," notes Rathke. "Qumulo lets me know in an instant how the data is used, who touches it and how often – so storage is no longer a technical issue, but a management decision."

And managing that storage has never been simpler. When another group considering Qumulo within the University asked about its ease of management, Rathke had a ready answer: "I told them to have senior staff figure out the architecture and configuration, then just throw the daily management over the fence to the student help desk, because from a management standpoint it's that easy – a true 'fire and forget' solution."

Throughout the implementation process, the Qumulo Care dedicated customer support team worked in lock-step with Rathke's group to ensure smooth deployment and growth. "Their flexibility and willingness to help us make the right decisions – not just up front, but going forward – has been key to keeping us on track," he says.

In the future, the SCI Institute plans to add an additional QC24 node to its cluster every month or two, and looks forward to future software feature enhancements, such as hard user quotas, that promise to make management of a large shared resource even easier. In particular, Rathke's team will be exploring Qumulo's REST-API to give users direct insight into storage usage through a customized user dashboard.

Ultimately Qumulo helped the SCI Institute rethink how it approaches data management. "We used to say 'just add more capacity', but what you're doing these days with storage is so much more complicated and mission critical than it was 10 years ago – and the volumes are so much larger – that for us to effectively grow we need to consider storage a finite and precious resource, and manage it accordingly. That's what Qumulo gives us."

The SCI Institute is an internationally recognized leader in visualization, scientific computer and image analysis. Its overarching research objective is to create new scientific computing techniques, tools and systems that enable solutions to problems affecting various aspects of human life.

Qumulo

Q129 0417