CARINGO SWARM: A MAINTENANCE-FREE, EASY-TO-SCALE STORAGE PLATFORM FOR IQ MEDIA

iQ media

DATACORE - has - ACQUIRED CARINGO

By Deni Connor, founding analyst
SSG-NOW
April 2017

Lakshmi Venkataswamy, CTO and co-founder of iQ Media, had some storage challenges to overcome—he needed a new storage platform for his company's 6PB of video files (primarily TV broadcast clips)—dating back to 2011 and growing at a rate of 6 TBs per day. The cost curve of traditional storage technologies was straining his budget and he was looking for a storage solution that was appropriately priced, high-capacity, low-maintenance and that would extract every cent of value from the underlying hardware by enabling it to operate well beyond its warranty coverage without increasing the risk of data loss.

In addition, he needed to store this data and protect it with a minimum of human intervention. With just two IT staff members available, Venkataswamy needed a storage solution that could be easily deployed and configured and could grow drive by drive at the same rate as his storage—6TB a day, 365 days a year.

"iQ Media is a cloud-based media intelligence platform," says Venkataswamy. "The company's product, cliQ, lets marketers and media analysts quickly find and measure the impact of various campaigns, branded, or topic-related content over an extended period."

The company offers a SaaS-based media intelligence platform that monitors, collects and stores video clips containing logos, brand names and keyword mentions from live and historical TV programming in real-time.

The company's clients access a Web-based dashboard to search for and retrieve TV clips that include their desired content—clips that include their name, logo or those of their competitors. (Among iQ Media's 400 clients are 3M, Domino's, DreamWorks, the National Hockey League and SONIC Drive-In.)

"We can do brand recognition and also identify metadata for video objects—audiences and demographic data," says Venkataswamy.

Venkataswamy needed a storage system that could store, retain and provide access to everything captured. He also wanted to deliver an operational competitive advantage by providing historical data, as well as real-time data.

"We don't erase anything on our platform," says Venkataswamy, alluding to the fact that clients need to be able to view both real-time and historical data to make their assessments of their brand's effectiveness.

The search begins

In his search for a storage solution that would meet his needs, Venkataswamy and his staff first installed Windows Storage Server, an installation that consisted of 15–16 clustered Windows Storage Nodes.

"Our application was written to handle Storage Nodes and we thought of it as a poor man's object-oriented system," says Venkataswamy. "We knew where the video clip was, we had a pointer to it, we could move clips around and update our database for the asset."

Working with Window Storage Nodes, however, created problems. As the size of hard disk drives increased from 2TB and 3TB capacities, RAID rebuilds of failed drives increased in time.

"We were using 2TB and 3TB drives configured with RAID 5," says Venkataswamy. "We'd have six drive chassis hanging off a node, each chassis with two drives or 12 Windows shares mounted."

This was practical, Venkatswamy asserts, when you had smaller drive capacities. However, as drive capacities grew, so did the amount of time required to recover from a failure.

"The maintenance time was also high on 15 or 16 Windows Storage nodes and we didn't want to put our data all on one node if the failure of a single node would impact the entire pool."

When he realized that RAID 5 wasn't working for data protection, Venkataswamy chose and installed a Dell PowerVault MD3460 array with a PowerVault MD3060e expansion chassis. This installation, however, was limited to a total of 180 drives, far short of the number of drives Venkataswamy needed to handle the real-time and historical video clips.

That configuration lasted about a year before Venkatswamy was again searching for a storage solution that could not only provide built-in data protection, but was relatively maintenance-free. He looked at SwiftStack and the Hitachi Content Platform (HCP), but rejected them because they were either not maintenance-free enough or because of cost.

"We are a small shop and don't want to be in the business of storage," says Venkatswamy. "That is why we did not go with something like OpenStack Swift, where we would have to babysit the project. We wanted a plug-and-play storage environment."

Caringo arrives

In September 2015, Venkataswamy chose Caringo Swarm and its FileFly migration utility, which enables the migration of data from Windows Server to a secondary pool of Caringo Swarm Storage.

Caringo met a lot of iQ Media's objectives:

- It was maintenance free. Nodes could be added one-by-one as growth occurred.

- Drives could be replaced and automatically be recognized by the cluster and load-balanced without much human intervention.

- Caringo Swarm was storage-agnostic, letting Venkatswamy use storage nodes from any vendor at prices and capacities he chose.

- Files stored in the Caringo Swarm cluster could be automatically protected in the event of drive failure by employing Caringo's 5:2 erasure coding, and avoiding the cost and complexity of replication.

- FileFly automatically moved files from filers to Swarm and gave users the ability to consolidate, access and dynamically organize files at massive scale. Swarm optimized storage operations and freed up resources to allow iQ Media to focus on continuing to deliver value to customers instead of managing the storage infrastructure.

Venkataswamy and his team started with three Caringo nodes, migrating data to them from their Windows Servers. "We could just plug a new node in and it would boot itself up automatically," says Venkataswamy. "There was no prep work. We had no OS installs. We could migrate disk groups between nodes and Swarm would decide if the disk was part of the node or not and move it around to the correct node."

His goal is to have a landing point on a Windows Storage Server where, after 30 days, all data will migrate using Caringo FileFly to a Caringo Swarm node. And, he's most of the way there; with only 1.5PB to migrate, Caringo Swarm has kept up with data growth and done so incrementally.

iQ Media now has 12 nodes in Newark, DE and ten in Columbus, OH

The present, the future

The future for Caringo Swarm at iQ Media is an extension and expansion of the present.

One challenge is the advent of NFS v4 and how iQ Media handles its integration.

"Our application is a blend of open source software," says Venkataswamy. "The main website runs on Windows. Apart from that, everything else is open source. So, we need a mix of NFS and SMB mounts."

Venkataswamy is presently using Swarm and FileFly to migrate data from Windows Servers to Caringo Swarm for long-term storage. In the future, he plans to integrate Caringo's SwarmNFS for its support of NFS v4, which takes advantage of clustered server deployments, including the ability to provide scalable parallel access to files distributed among multiple servers.

"Getting all the processes cut-over from SMB to NFS is a big job," says Venkataswamy. "If we can cut-over to NFS access for a lot of these things and see the throughput of what's available on NFS, we'll run something parallel for a while and then we'll cut-over."

The other goal in Venkataswamy's plan for the future is to "sustain a no-maintenance approach."

"I want to use the hardware until it dies," he says. "Most drives come with five-year warranties. When they die, we just dispose of them which creates a maintenance cycle that is low effort from a human perspective."

Caringo Swarm makes this simple. "The deployment of new nodes is easy—you rack it, you plug it in, you point at it and 15 minutes later your node is up and you are ready to go."

When asked for his overall assessment of Caringo Swarm and FileFly, Venkataswamy said, "FileFly automatically moves files from filers to Swarm and gives us the ability to consolidate, access and dynamically organize files at massive scale, optimizes our storage operations, and frees up resources to focus on continuing to deliver value to our customers instead of managing the storage infrastructure."

Learn more about DataCore Swarm

About SSG-NOW™

SSG-NOW is an industry analyst firm focused on storage, server, cloud and virtualization technologies. Our goal is to convey the business value of adopting these technologies to corporate stakeholders in a concise and easy-to-understand manner.