Qumulo is latest to pitch 'data awareness' with new take on scale-out NAS

Analyst: Simon Robinson

19 Mar, 2015

Storage startup Qumulo revealed itself this week. The company – founded by the same team behind successful scale-out NAS specialist Isilon – believes its software offers a vast improvement on the model for effectively storing and managing unstructured file data. Qumulo says its unique spin with its Qumulo Core system is 'data awareness' that can drive new levels of performance and efficiency at significant levels of scale.

The 451 Take

The storage startup narrative has been so dominated recently by startups in the block world – from Pure Storage to Nutanix – that the file side of the discussion has been somewhat neglected. And even where the unstructured data debate has continued, the focus has been on object-based storage as an alternative to NAS. Qumulo believes that this is largely a distraction, and that despite the undoubted potential of object storage, what many organizations dealing with humongous volumes of file data really need is a better NAS platform. In this respect, Qumulo's launch brings the NAS discussion sharply back into focus. Does the world really need another scale-out NAS platform? Clearly, Qumulo believes so. And while the challenges in overcoming entrenched rivals EMC and NetApp are significant, it's also true that few other technology teams have a better understanding of the profit and pitfalls of this market.

Context

Seattle-based Qumulo was founded in 2012 by a trio of engineers who were instrumental in creating Isilon's OneFS, the scale-out NAS system that was acquired by EMC in in 2010 for $2.5bn. CEO Peter Godman was director of software engineering at Isilon, VP of engineering Neal Fachan was a distinguished engineer, and CTO Aaron Passey was chief architect. Among them, they are named on 55 of Isilon's 62 patents, so it's fair to say they have a reasonable handle on scale-out NAS. Other former senior Isiloners include VP of marketing Brett Goodwin and VP of operations Mary Godwin. Additionally, Isilon's founder, Sujal Patel, recently joined Qumulo's board. Qumulo has raised $67m in venture funding, including a recent $40m series B that was led by Kleiner Perkins Caufield & Byers.

Strategy

Qumulo says that before it started building anything, it spent time talking to 600 storage and IT decision-makers about their storage pain points. Dealing with unstructured data at scale consistently came out as the main challenge. This is especially the case for organizations that are built on such data – think rendering in media and entertainment, oil and gas, seismic data processing, genome sequencing, etc.

But weren't scale-out NAS systems such as Isilon created to address this very challenge? Yes, says Qumulo, and they helped to a great extent with the overall cost of management and improving performance in large file environments. But existing scale-out NAS systems still have their limitations, Qumulo says. For one, they are still geared for handling large files, so they struggle when it comes to dealing with the more transactional I/O that comes with handling small files. This is becoming a bigger issue as more applications deal with huge numbers of small files, which is typical output for genome sequencing, for example, and Qumulo lists a chip vendor that today stores roughly 69 billion files.

The second limitation with existing scale-out NAS is that at significant levels of scale where hundreds of billions or even trillions of files are involved, it becomes crucially important for the system to have a real-time understanding of the data it's storing – and how it's being used – in order to optimize for both performance and efficiency. So, for example, it would be useful for admins to know which users or projects might be slowing down overall cluster performance, or get a better understanding of which data can be archived or saved for data analytics. Existing systems lack any such data awareness beyond fairly simple metadata, Qumulo says. Indeed, this is where many organizations come to understand the limitations of storage. Qumulo's mission is to provide data visibility through a storage platform that is essentially 'invisible.'

Notably, Qumulo's Core product has been shipping to customers since August, and the company says it has about 15 paying customers, with the majority in production. Qumulo's early customers include Ant Farm, BLIND, Densho, Sinclair Oil, Sportvision, TELUS Studios, University of Utah Scientific Computing and Imaging Institute, and Zoic Studios.

Products

Qumulo says it has engineered its Qumulo Core product and its Qumulo Scalable File System (QSFS) software to address both of these shortcomings, and more. It is designed as a platform to allow administrators to 'curate' vast amounts of digital content. Qumulo Core is a software-only product that runs as a user application on Linux. The software can run as an appliance on any commodity hardware or as a VM. This means the system could run in a blended on-premises and public cloud environment. Qumulo has built a 1U hybrid storage appliance that offers 24TB of raw HDD capacity (4x6TB Helium HDDs) and 1.6TB of raw SSD capacity. Systems start with a four-node minimum but can scale to more than 1,000 nodes in a single cluster and namespace. Nodes are interconnected using 10GbE. At this stage, Qumulo is supporting clusters of up to 10 nodes on its own hardware, although it says it has tested hundreds of nodes in its lab with no scaling issues. Pricing for a 100TB raw, four-node cluster starts at $50,000.

QSFS is a fully consistent system, supporting POSIX semantics. It supports SMB, NFS and REST-based protocols, as well as a REST-based API, offering fully programmable management. From an architecture point of view, Qumulo says it has separated the data-protection function from the file system itself; running data protection under the file system provides the basis for greater scale and reliability, it says. It uses mirroring for resiliency today, although it's also developing an erasure coding approach for future use. Qumulo says the software is 'absurdly easy to use.'

Another key advantage of its architecture is the ability to support real-time analytics, Qumulo says. It has built a dashboard offering a variety of views into multiple dimensions of system performance, examples being IOPS by client, IOPS by path and IOPS hot spots. It has built a proof-of-concept system running four billion files in 300,000 directories; it was able to add another node to the four-node cluster in 90 seconds, and says it rebuilt a failed 6TB HDD in 17 hours.

One aspect still missing from Qumulo's software is support for data-protection features such as snapshots. The company says that although systems such as Isilon can take snapshots on a per-directory level, this is not necessarily helpful for directories that contain millions of files. Qumulo says it will develop a more granular snapshot method to get around this, although it has nothing to announce at this stage.

Competition

While the block storage market has always been a crowded space with several dozen well-established players and startups jockeying for position – especially at this time as flash drives new disruption opportunities – the file storage market has always been relatively much less crowded. This is good news for emerging players such as Qumulo, although it also reflects that this is a smaller market compared to the block storage space. That said, there is still a sizable opportunity here; Qumulo believes the NAS market opportunity is a $10bn total available market – roughly half of the SAN space overall, but with far fewer competitors.

Fewer rivals should not be confused with a lack of competition, however, especially when one of those rivals is EMC, which has carved out a string position in the scale-out NAS market following its purchase of Isilon. Meanwhile, NetApp remains strong in the core NAS market, even if its focus over time has extended to scale-out NAS and more VM-centric workloads. In terms of differentiation, Qumulo says Qumulo Core is designed to offer the small-file performance of NetApp combined with the scalability of Isilon.

However, outside of EMC and NetApp, few other major players have a real strength in enterprise NAS. Vendors such as IBM (with GPFS), HDS (with BlueArc) HP (with IBRIX) and Dell (With Exanet) have tried to gain a substantial foothold in this space, although in truth, all have struggled to break the effective EMC-NetApp duopoly. However, ZFS continues to enjoy some support, being promoted by companies such as Nexenta and Tegile. Meanwhile, startups such as Primary Data and DataGravity are also begging to talk up the opportunity around 'data aware' storage. Meanwhile, DataDirect Networks and Seagate are among those pushing NAS into the HPC market.

Elsewhere, object storage approaches from the likes of Cloudian, Scality, Cleversafe, Caringo, DataDirect Networks and Exablox are often cited as alternatives to NAS for very large data volumes. And NAS caching companies such as Avere Systems are teaming with object storage vendors to provide an alternative to scale-out NAS. Qumulo believes that despite growing interest in object storage, most of its target customers are put off because their applications were built on a file system. In response to this, some object storage vendors are now adding file-system and even block-level support to their object stores. Conversely, Qumulo can cater to applications requiring object-level access with its REST support.

SWOT Analysis

Strengths

Few are as qualified to build a scale-out NAS company as Qumulo's founders, having already successfully created and sold one such company: Isilon. The fact that the company already has a dozen or so customers is another plus.

Weaknesses

It's still early days for Qumulo, and its product initially lacks features that some enterprises value, such as snapshots.

Opportunities

Qumulo believes its total available market is $10bn. Regardless of how close to the truth that it, there's no doubt that more organizations are increasingly challenged when it comes to managing, storing and optimizing their unstructured data.

Threats

The NAS space is still dominated by just two companies – NetApp and EMC – which remain formidable. Additionally, the NAS space has not generated as much attention as the SAN space recently, especially in the context of the flash opportunity.