



Building Optimized Scale-Out NAS Solutions with Avere and Arista Networks

Record-Breaking Performance in the Industry's Smallest Footprint

Avere Systems, Inc.
5000 McKnight Road, Suite 404
Pittsburgh, PA 15237 USA
1-412-635-7170
www.averesystems.com
info@averesystems.com

Part number 0254-002-0171, Rev A, electronic version
Part number 0254-002-0191, Rev A, printed version

Executive Summary

Since 1994, the storage industry has used the SPEC® standard to benchmark NFS performance of different storage systems by different NAS and storage vendors.

In 2011, Avere Systems, Inc. posted a world-record SPECsfs®2008 NFSv3 result of over 1.56 million operations per second with an overall response time (ORT) of 0.99 milliseconds. In addition to the magnitude of performance demonstrated by this result, the posting was notable for several other factors:

- The equipment used to generate the result fit in two and a half standard equipment racks—less than one-quarter of the footprint required by the previous record-holder’s setup.
- Because less equipment was used, both the capital expense (CapEx) of the system and the operating expense (OpEx) of the power and cooling requirements for the record-setting run were also a fraction of those needed by the previous incumbent’s setup.
- The testing configuration used four NAS servers (in Avere terms, *mass storage servers*) whose NFS exports were combined into a global namespace by using Avere’s GNS capabilities. The NAS servers used in the testing configuration ran free and open-source software (OpenSolaris and ZFS) on commodity hardware; however, the Avere solution can be used with any NAS server or servers that support the industry-standard NFSv3 file protocol. The testing configuration could just as easily have included commercial NAS servers from multiple vendors, including NAS servers that themselves export global namespaces, instead of or in addition to the open-source NAS servers.

Avere was able to achieve this recording-breaking performance result with such a small footprint by using its FXT Series appliances in combination with Arista Networks’ 7000-model network switches and commodity backend NAS servers. The combination of Avere and Arista enabled record-setting performance with maximum density, lower power, a small footprint, unprecedented ease of management, and a drastically lower cost. This paper describes the innovations and synergies that resulted in a world record.

The Avere Advantage

The Avere solution, consisting of Avere OS software running on FXT Series appliances (or *nodes*), provides the following main advantages for both new and existing NAS deployments:

- Performance scaling—Avere can accelerate client access to any NAS server, from low-end models and legacy equipment to the latest high-end models.
- Cost reduction—By providing high performance in a small footprint, Avere reduces both the cost of initial equipment purchase and ongoing costs for power, cooling, and data-center space.
- Simplified management—Avere’s browser-based administrative interface provides straightforward management of the cluster, and its Analytics capabilities provide deep insight into all operations across all components of your storage network. Additionally, Avere’s global-namespace (GNS) capabilities enable you to create a single namespace out of multiple NFS exports on multiple NAS servers, simplifying client access.

Performance Scaling

Avere’s FXT Series appliances enable the scaling of performance through the use of *tiered NAS*—that is, intelligently placing data from NFS clients requests into different types of storage (RAM, SAS, SSD) based on the type of request (read-only or read-write; sequential or random), the frequency of requests for a particular data set, and the type of data being requested (large files or small files; whole files or blocks of a particular file or files). Data requested by clients and application servers constitutes an FXT cluster’s working set, which changes dynamically depending on access patterns.

In the Avere solution, FXT Series appliances, or *nodes*, are joined together in a *cluster*; the performance of the cluster increases near-linearly with each additional node as the node’s resources (RAM, SAS, SSD, and CPU) are merged with the other nodes’ resources. Adding one or more nodes to an existing cluster is a one-click task; no additional configuration is required. The FXT cluster sits between your existing NAS servers (in Avere terms, *mass storage servers*) and your clients and application servers. The FXT cluster uses the industry-standard NFSv3 protocol to communicate with both the mass storage servers and clients and application servers. (Note that the FXT solution can also be used with CIFS clients via NFS-to-CIFS translation on the FXT cluster.)

An FXT cluster can be placed in front of an arbitrary number of mass storage servers.

The FXT cluster handles write requests as well as read requests from clients and writes changed data back to the mass storage server or servers at an administratively specified interval. No other product on the market today accelerates write performance as well as read performance.

An FXT cluster can also be used to reduce latency over a WAN, making mass storage servers at remote data centers as responsive to clients as local mass storage servers. With geographically distributed work sites and data centers becoming the standard, faster response times to client requests rapidly add up in the form of cost savings and increased productivity.

Cost Reduction

With the FXT cluster accelerating your NAS performance, you can scale NAS capacity by adding high-density, low-cost SATA drives to your mass storage servers. Typically, achieving satisfactory performance with a capacity upgrade requires the addition of high-cost, low-density and overprovisioned SAS or FC drives, if not replacement of the mass storage server itself with the vendor's latest high-end model. If you need to add more mass storage servers, you can choose less expensive lower-end models or, as Avere did for the SPEC setup, an open-source NFS server running on commodity hardware.

Simplified Management

The FXT cluster is managed from the Avere Control Panel, a browser-based GUI that provides full access to all administrative functions. A typical FXT cluster requires minimal administration after initial configuration. The Avere Control Panel features a Dashboard that provides a high-level overview of the cluster's health and the ability to drill down into detailed information about cluster operations. It also features an extensive Analytics panel that enables you to visualize performance and operations across your entire storage network, from clients to the FXT cluster to mass storage servers. You can choose from a number of predefined graphs or create customized graphs. You can also view histograms to simplify the visualization of "hot spots" across the storage network; among other uses, histograms can help you optimize the distribution of resources across your storage infrastructure.

Although the press coverage of Avere's SPEC result focused on its record-breaking performance, a critical aspect of the result is equally remarkable but has been largely overlooked: By taking advantage of Avere's GNS capabilities, the SPEC result used a single filesystem namespace composed of multiple, disparate volumes (filesystems) from multiple NAS servers. Avere's GNS means

that you no longer need to pay premium prices for a global namespace; it also means that you can create a global namespace out of NAS servers from multiple vendors, including NAS servers that provide their own global namespace!

The Arista Advantage

Arista Networks is a leading network solution provider for building scalable high-performance and ultra-low-latency data center and cloud computing networks. Arista networking solutions are ideal for storage interconnect, providing wire-speed, non-blocking 10Gb, and 40Gb Ethernet switching, with extremely low latency. The lower the latency between a storage read and write, the more efficiently the host can process data applications and deliver more IOPs. Arista Networks switch latency ranges from an industry-leading 500ns with the 24-port 10Gb Ethernet Arista 7124SX, used in this test, to 800ns for a 7050Q-16, 16-port 40Gb Ethernet switch and 4.5usec for the 384 port 10GbE modular Arista 7500 series.

Speed differences in storage and host IO connectivity require network buffering. Arista provides switches with extremely large buffers, up to 100 times the size of competitive solutions. This additional buffering capacity absorbs storage IO bursts of reads /writes and minimizes dropped frames, reducing storage IO retransmissions and optimizing storage performance.

The Arista 7000 Family of switches, like Avere's FXT Series NAS appliances, provide higher performance, simplified management, greater density, and lower power/cooling requirements than competitive products, which reduces ongoing costs for power, cooling, and physical space, three key contributors to operational expense.

The Arista Extensible Operating System (EOS) provides support for many open management toolsets, flexible scripting and scheduling tools such as Chef, Puppet, and Python, along with support for local logging of network performance data and real-time buffer utilization to ensure optimum network performance in demanding storage and compute environments.

Avere uses Arista equipment in its internal infrastructure for the same reasons it used Arista equipment in the SPEC configuration.



Avere: 44 FXT 3500 nodes, 4 mass storage servers, 2.5 cabinets, 6 feet wide



Competitor A: 24 FAS 6240 nodes, 72 disk shelves, 12 cabinets, 24 feet wide



Competitor B: 140 S200 nodes, 1 Infiniband switch, 7 cabinets, 14 feet wide

Stacking Up Against the Competition

In the long history of the SPECsfs NFS benchmark, only three results of over one million operations per second have been posted. The first, from Competitor B, was 1.11 million operations per second. The second, a few months later from Competitor A, was 1.51 million operations per second. The third, only a few weeks after Competitor A's result was posted, is Avere with 1.56 million operations per second. The preceding illustration shows the amount of gear required to achieve each company's result.

As you can see, Avere packs the highest performance and lowest latency into a package that is 79% smaller than Competitor A's result and 65% smaller than Competitor B's result.

The following table provides a detailed comparison of the three top results. In comparison to the competitors' solutions, Avere's solution:

- Costs 51% to 77% less
- Requires 56% to 78% fewer disk drives
- Occupies 65% to 76% fewer rack units

SPECsfs2008 does not measure power or cooling requirements (that is, ongoing operating expenses after initial acquisition costs). In a storage system, the disk drives are the largest consumers of power and dissipaters of heat. Therefore, a good estimate for the power and cooling savings is the disk savings, where Avere is 56% to 78% less.

	Avere	Comp. A	Comp. B	Avere Advantage
Throughput (IOPS)	1,564,404	1,512,784	1,112,705	3% higher than Competitor A 41% higher than Competitor B
ORT, Latency (msec)	0.99	1.53	2.54	35% lower than Competitor A 61% lower than Competitor B
List Price (\$/IOPS)	\$2.5	\$5.1	\$10.7	51% less than Competitor A 77% less than Competitor B
Data Disks (qty)	752	1728	3360	56% fewer than Competitor A 78% fewer than Competitor B
Rack Units (RU)	104	436	288	76% fewer than Competitor A 64% fewer than Competitor B
42U Cabinets (qty)	2.5	12	7	79% fewer than Competitor A 65% fewer than Competitor B
Product Config.	FXT 3500 44-node cluster	Highest-end server; 24-node cluster	Scale-out server; 140-node cluster	

Optimizing Scale-Out Storage

The unique benefits of the Avere/Arista solution are obvious in contrast with the market incumbents' solutions:

- Industry-leading performance that can be scaled to your organization's current **and** future requirements by adding new nodes to an FXT cluster
- The ability to upgrade NAS capacity by adding low-cost SATA drives instead of high-cost SAS or FC drives to your mass storage server
- A small data-center footprint
- Low cost of acquisition and operation
- Simplified management, requiring fewer administrative resources to maintain
- The ability to scale out your existing storage infrastructure instead of replacing it with the latest top-of-the-range gear from your current vendors

In summary, the Avere/Arista solution is the most **efficient** storage solution across all dimensions on the market today. To learn more, contact your Avere or Arista representative.

Avere's record-breaking result was posted to the SPEC website on November 15, 2011. The full result is available at www.spec.org/sfs2008/results/res2011q4/sfs2008-20111031-00204.html.

SPEC® and SPECsfs®2008 are registered trademarks of the Standard Performance Evaluation Corporation (SPEC). See www.spec.org for further information about SPEC and SPECsfs2008.