Quantum

Reference Architecture for High-Performance Video Workloads Using StorNext 7, Quantum F-Series, SPEC SFS 2014 SP2 VDA

Everett Ward, Technical Marketing Engineer

Executive Summary

The generation of video and different forms of digital images are growing in all industries, for entertainment and communication purposes, for training and marketing, and for surveillance. In all of these use cases, extremely high-performance storage systems are required to handle the ingest and processing of this video content.

In order to demonstrate the performance of Quantum StorNext File System running on Quantum F-series NVMe storage servers, Quantum used the SPEC SFS 2014 SP2 Video Data Application ("VDA") workload which is a publicly-available benchmark test designed to simulate a high-performance video workload at scale.

StorNext® is a high-performance file system ideal for high throughput, low latency workloads, and is used extensively in media and entertainment, different forms of scientific research and computing, and for video surveillance capture and recording. Quantum built this reference architecture and performed the testing using commercially available off-the-shelf products, and the results set new benchmarks for:

- The highest number of concurrent streams (**7,450**)
- The lowest latencies (**overall response time of 0.90 msec**)
- The highest aggregate throughput (**34,391 MB/sec**)

Further, StorNext demonstrated these results at a **much lower $ per concurrent stream** with:

- 57% fewer storage nodes
- Significantly less expensive hardware (less CPU & memory)
- Fewer NVMe drives

More details on the results and the reference architecture are provided in this paper.

Perspective on What These Results Mean

This reinforces that StorNext provides the **best performance**, and **best value**, for any workloads that require massive streaming performance with a large number of concurrent users or processes.

StorNext provides parallel file system access for high throughput, low latency workloads, and also reduces storage costs by using policy-based data movement to manage data across its lifecycle. StorNext software can move files and folders between NVMe and nearline disk, between primary storage and secondary storage options such as public and private object stores, as well as tape for offline long term archiving.

Use cases in media and entertainment, video surveillance capture and retention, and in earth and life sciences with large digital image files can all benefit from the streaming performance and data lifecycle management of StorNext.

SPEC SFS 2014 Test Results

The output of the testing is shown in the figure below.

Business Metric (Streams)	Average Latency (msec)	Streams Ops/Sec	Streams MB/Sec
745	0.8	7455	3440
1490	0.8	14910	6870
2235	0.8	22365	10320
2980	0.8	29819	13758
3725	0.8	37275	17192
4470	0.8	44730	20643
5215	0.8	52185	24077
5960	0.9	59640	27539
6705	1.1	67095	30947
7450	1.7	74550	34391

This reference architecture demonstrated the following results:

- Streams = **7450**
- Overall Response Time = **0.90 msec**
- Aggregate streaming throughput = **34,391 MB/sec**

The tables below show the success of StorNext in this benchmark. StorNext delivers a 2.5x price-to-performance advantage over the second place vendor, faster performance with 57% fewer storage nodes,

A record-breaking ORT that is 42% lower than the closest competitive vendor with 57% less storage nodes, 28% fewer NVMe drives, and more than 70% fewer processors and memory.

Metric	Quantum	Second Place Vendor	Summary
Streams	7450	6800	10% more Streams
ORT	0.9	1.56	42% lower ORT
MB/sec	34391	31375	10% more MB/sec

Hardware	Quantum	Second Place Vendor	Summary
Storage Nodes	10	23	57% less Storage Nodes
NVME Drives	100	138	28% less NVMe Drives
Processor	10	46	78% less Processors
Memory	80	276	71% less Memory
Network Interface	20	23	13% less Interface cards
100GbE switches	2	10	80% less switches
clients	14	18	22% less clients

Applicability to Media and Entertainment Workflows

The SPEC SFS VDA workload most closely simulates a video surveillance capture use case, where camera data streams with 100% sequential writes on VDA1 and a companion stream, VDA2, that is 84% random reads. Each stream is 36Mbit/sec bit rate, and the results indicated consistent performance across the range of testing. Because of the large number of streams and low bit rate, the benchmark records the number of streams, overall bit rate (ORT), and total MB/sec as metrics to gauge how well the solution performs.

Other streaming workloads, outside of this benchmark, such as 4K/8K high resolution, high frame rate workloads are commonly found in the media and entertainment industry.

Separate testing conducted on this reference architecture resulted in:

- **Aggregate throughput speeds of 230GB/sec**, with 12 clients performing 4K uncompressed and unconstrained concurrent streams.
- **Read performance from a single client exceeding 20GB/sec** with two ports of 100GbE, directly connected via iSER/DRMA to the storage cluster.

This makes this reference architecture well suited for media and entertainment workflows that require large single stream performance, as well as concurrent access.

Reference Architecture Details

SPEC SFS 2014 is an industry standard performance benchmark. Using the generally available products listed below, Quantum designed this test bed, configured it as noted in this section, and used the SPEC SFS test workload to validate the performance results. The design for Quantum is to provide a solution that adheres to the guidelines put forth by the SPEC SFS committee, which include:

- The solution under test must be available within 6 weeks of the first publication of the results.
- Any modifications to hardware or software must be published or disclosed.
- As for the performance run, the test is started and must pass 10 data points without stopping or any failure notices.
- Once completed the results are collected and submitted to the SPEC committee for review, after a review process, the results will be published, usually within 2 to 3 weeks.
- A full list of the run rules are located here.

A diagram and description of the reference architecture are below.

The benchmark runs on a group of workstations and measures the performance of the storage solution that is providing files to the application layer. Because the benchmark runs at the application system call level, all components of the storage solution impact the performance of the solution – this includes the load generators themselves as well as any physical or virtual hardware between the load generators and where the data ultimately rests on stable storage.

There are several connectivity options for the load generating client, such as NFS, SMB, iSCSI, FC. In a typical SPEC SFS test configuration, a series of load generating clients are directed through a network at file system/s shared or exported from a file server. This solution consists of the load generators, the network, the file server, and finally the stable backend storage. For the Quantum solution the load generating clients are 1U servers provided by Quantum known as Xcellis® Workflow Extenders.

The network is a 100GbE dual network/subnet Arista switch configuration, with clients and storage connected via iSER (iSCSI Extensions for RDMA), each component has a dual link to the switch fabric.

Our solution had a single file system created and shared out to the clients. The file server is Quantum's Xcellis Workflow Director running StorNext 7.0.1.

For our validation the workload referred to as VDA (Video Data Acquisition) was tested. This workload generally simulates applications that store data acquired from a temporally volatile source (e.g. surveillance

cameras). A stream refers to an instance of the application storing data from a single source (e.g. one video feed), maintaining a minimum fixed bit rate per stream and secondarily about maintaining the fidelity of the stream. The goal is to provide as many simultaneous streams as possible while meeting the bit rate and fidelity constraints.

The business metric for the benchmark is VDA. This benchmark consists of two workload objects: VDA1 (data stream) and VDA2 (companion applications). Each stream corresponds to a roughly 36 Mb/s bit rate, which is in the upper range of high definition video. The validation criteria for VDA is as follows.

- Threshold Value Per proc oprate >= 75% of requested
- Global average oprate >= 95% of requested
- Workload oprate ratio variance <= +/- 5%

Products Used in the Reference Architecture

Item No	Qty	Type	Vendor	Model/Name	Description
1	1	Parallel File System	Quantum	StorNext V7.01	High-Performance, parallel file system, scales across storage nodes, capacity and performance, multiple OS support
2	10	F1000 Storage Node	Quantum	F-Series NVMe Storage	Single Node, F1000, each node has, 10 Micron 9300 MTFDHAL15T3TDP 15.36TB NVMe SSD, Single AMD EPYC Processor (7261 8 -Core Proc @2.5GHz), 64GB Memory, 2 x Dual port Mellanox ConnectX-5 100GbE HBA (MCX518A-CCAT). 2 x 100GbE connections to the switch fabric, 1 per ethernet adaptor
3	14	Clients	Quantum	Xcellis Workflow Extender (XWE) Gen2	Quantum XWE, each single 1U server has, 192GB memory, Dual CPU Intel(R) Xeon(R) Silver 4110 CPU @ 2.10GHz 8 Core, 2 x Dual Port 100GbE Mellanox MT28800 [ConnectX-5 Ex]. 2 x 100GbE connections to the switch fabric, 1 per ethernet adaptor
4	1	Metadata Controller	Quantum	Xcellis Workflow Director (XWD) Gen2	Dual 1U server with HA (high availability), each server has, 192GB memory, Dual CPU Intel(R) Xeon(R) Silver 4110 CPU @ 2.10GHz 8 Core, 1 x Dual Port 100GbE Mellanox MT28800 [ConnectX-5] each ConnectX-5 card connects with a single DAC connection to switch infrastructure, only for administrative purposes. Note: Secondary node is also being used as "SPEC Prime".
5	2	100GbE Switch	Arista	Arista DCS-7060CX2-32S-F	32 Port, 100GbE Ethernet switch
6	1	1GbE switch	Netgear	ProSAFE GS752TP	48 Port, 1GbE Ethernet switch
7	1	1GbE switch	Dell	PowerConnect 6248	48 Port, 1GbE Ethernet switch

Network Requirements

The SPEC Benchmark does not have a minimum network connection. It's based purely on testing the performance of the attached solution. According to the user guide, the solution may have several connectivity options ranging from NFS, SMB, direct attached fiber channel and 100GbE Ethernet. For our solution all clients and storage nodes are connected via Dual 100GbE Ethernet (iSER/RDMA). Additionally, there is a 1GbE network for metadata and administration of the solution.

StorNext File System Configuration

The storage solution configured for the benchmark was designed as a single file system, all storage nodes combined into a single stripe group. The connectivity option chosen is 100GbE iSER/RDMA. Each storage node has two 100GbE connections to the switch fabric. The switch fabric is configured into two subnets with and MTU set to 9000. No other special alterations were done to the storage, hardware, or software.

StorNext Tuning

StorNext tuning is done within the SN UI. Certain parameters are important depending on the type of workflow, type of media and performance that is needed for concurrent real-time playback or acceptable interactivity on the system. For an in-depth look at file system tuning and configuration, see the StorNext Documentation Center.

For the benchmark testing discussed in this document the following configuration parameters were implemented:

1. File system block size = 4k
2. Allocation Session Reservation (ASR) = 0 (off)
3. inodeStripeWidth = 4GB
4. Stripe Groups = one, combined metadata, journal, user data
5. LUNs: 20
6. Stripe breadth = 2.5MB
7. bufferCacheSize = 8GB
8. inodeCacheSize = 512k

F1000 Storage Node Configuration

1. Default configuration from manufacturing
2. Enable jumbo frames, set MTU = 9000

SPEC SFS 2014 SP2 Prime Configuration

The "Prime" is defined as the manager for the benchmark. The configuration files reside on the machine along with the license. It maintains all the SPEC tools used in gathering and submitting reports based on the results. The testing sequence is always launched from the prime.

1. Install SPEC SFS software as per user guide documentation, page 5. Be sure to review the prerequisites on page 5 as well, very important
2. Be sure to follow steps for either Linux installation or Windows.
3. Edit the configuration file in root directory of the SPEC installation
4. Install the license on Prime only

SPEC SFS 2014 SP2 Client Configuration

1. Install the SPEC SFS 2014 benchmark on all clients or load generators
2. Configure clients per the documentation on SPEC SFS 2014 SP2 site
3. Special tuning for clients in this solution
   i. Jumbo Frames = 9000 - set Jumbo frames to 9000
   ii. nr_requests =512 - maximum number of read and write queued at one time
   iii. scheduler =noop - I/O scheduler in Linux
4. 3/ii and 3/iii above are set in /usr/cvfs/config/deviceparams on each client
5. Entry in /etc/fstab
   cachebufsize=128k,buffercachecap=16384,dircachesize=32m,buffercache_iods=32,bufferlowdirty=6144,bufferhighdirty=12288

References

The documents below were referenced to configure the software and systems for validation of this reference architecture.

Document Title	Download URL
Xcellis Workflow Extender	Link to Document
StorNext	Link to Document