High Performance Computing

How NetApp plays in this world

Mohinder Toor

NEMEA E-Series Sales Specialist

NetApp

The “performance” triad

- What customers care about:
  - IOPS
  - Throughput
  - Latency
- The “other stuff” that matters:
  - Read/write mix
  - Large or small I/Os
  - Random or sequential
  - Consider full mix of I/O including host OS feature-related I/O

File System and Application type matter
Customer optimization criteria matters

Customer Challenges and Requirements

Customer Challenges

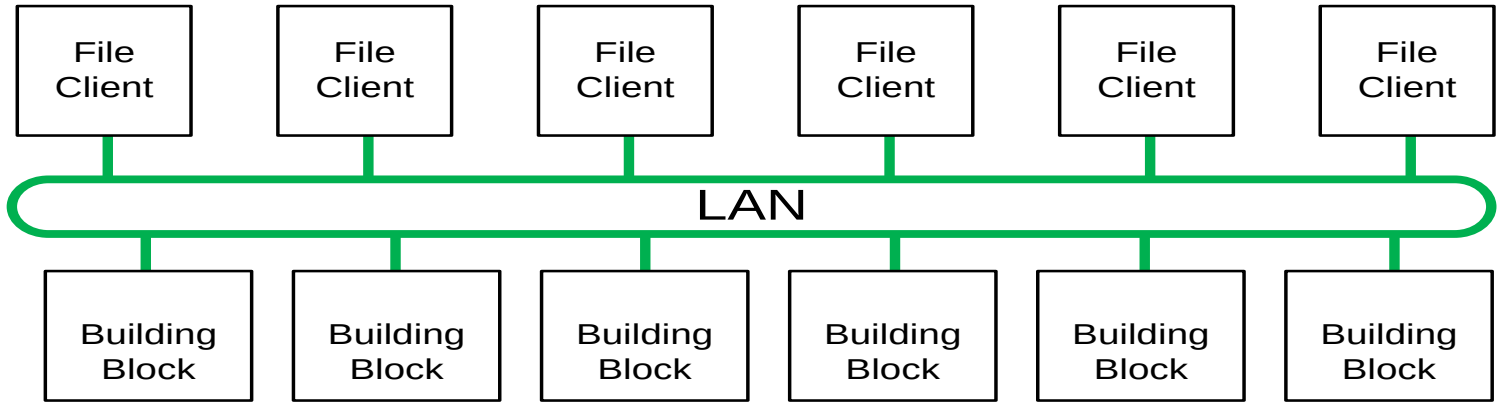
- Larger datasets
- Higher bandwidth
- Large block sequential read after write
- Ingest and analyze data in real time

Customer Requirements

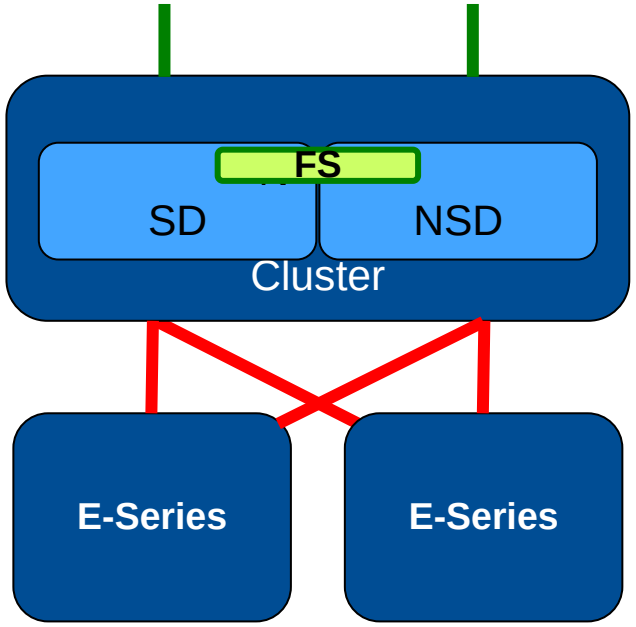
- Higher density rack space and lower power
- Better ROI
- High single stream or aggregate performance
- Reliable and Scalable

Introducing Storage Building Blocks

The concept

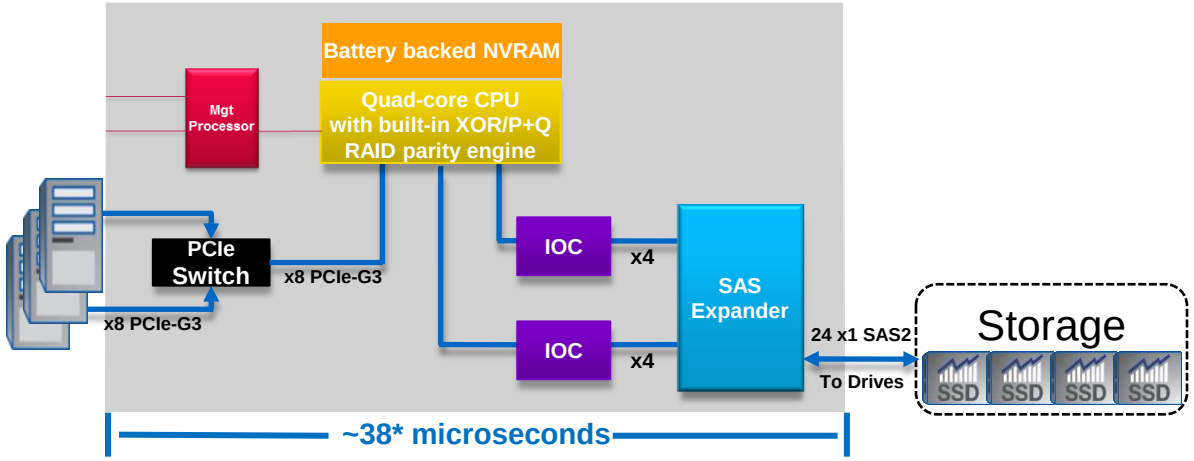


- Comprehensive benchmarked
- Pre-tested clustered server pair tightly coupled to right sized storage
- Storage building blocks simplify system specifications to meet performance and capacity requirements
- File system grows in performance and capacity concurrently and in a linear manner
- Simplifies File System sizing, tuning, and deployment
- SAN Layer [between NSD and storage], which is required in monolithic storage designs is completely eliminated



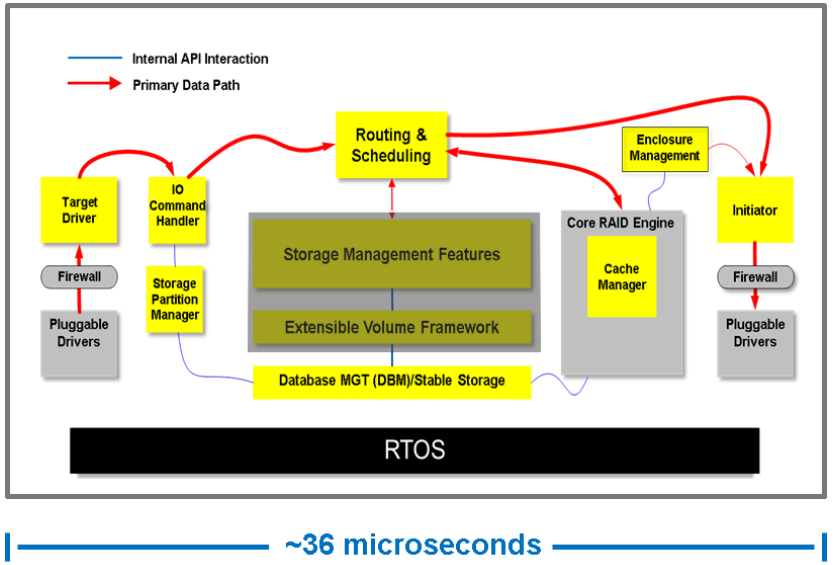
Optimized, Designed and Engineered for Performance

EF Controller & SANtricity OS Architecture



- No Virtual Layer adding latency
- Performance consistent over time
- No overhead for features not in use

- Extreme Parallel Operations
- Eliminate Contention
- Minimal Overhead



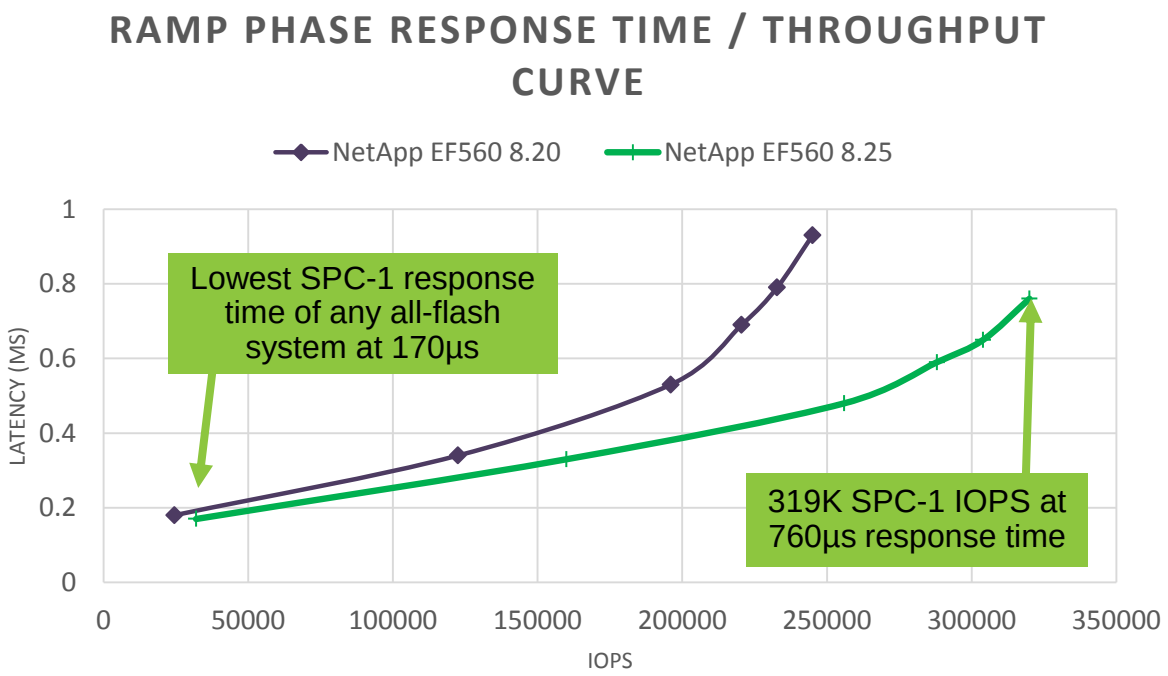
The SPC-1 benchmark

Updated submission to the EF560 SPC-1 benchmark results

Results

- EF560 with previous 8.20 code was
  - 245K IOPS with 930µs latency
  - With 400G PM2 SSDs
  - $0.58 per SPC-1 IOP, was #10 on the top ten list
- EF560 with new 8.25 code is
  - ~320K IOPS with 760µs latency
  - With 800G PM3 SSDs
  - $0.30 per SPC-1 IOP, #6 on the top ten list

Industry-leading all-flash array
Latency and IOPS



- #6 over all ranking
  - $0.30 per SPC-1 IOP
  - #4 ranking for HA systems
- #1 - Lowest Response Time (LRT) leader
- #1 - LRT to 100K IOPS at 250µs
- Consistent microsecond response for third platform workloads

NetApp

The SPC-2 Benchmark

Represents the average data rate, in megabytes per second, across workloads

Large File Processing

- Look for:
  - Scientific computing
  - Large-scale financial processing
- EF results
  - 104 MB/S/Stream
  - 501 MB/S/SSD
  - $9.20 MB/S

Large Data Queries

- Look for:
  - Data mining
  - Business intelligence
- EF results
  - 135 MB/S/Stream
  - 651 MB/S/SSD
  - $7.08 MB/S

Video On Demand

- Look for:
  - Video entertainment
  - Editing / creation
  - Transfer
- EF results
  - 14,000 Streams
  - 550 MB/S/SSD
  - $8.37 MB/S

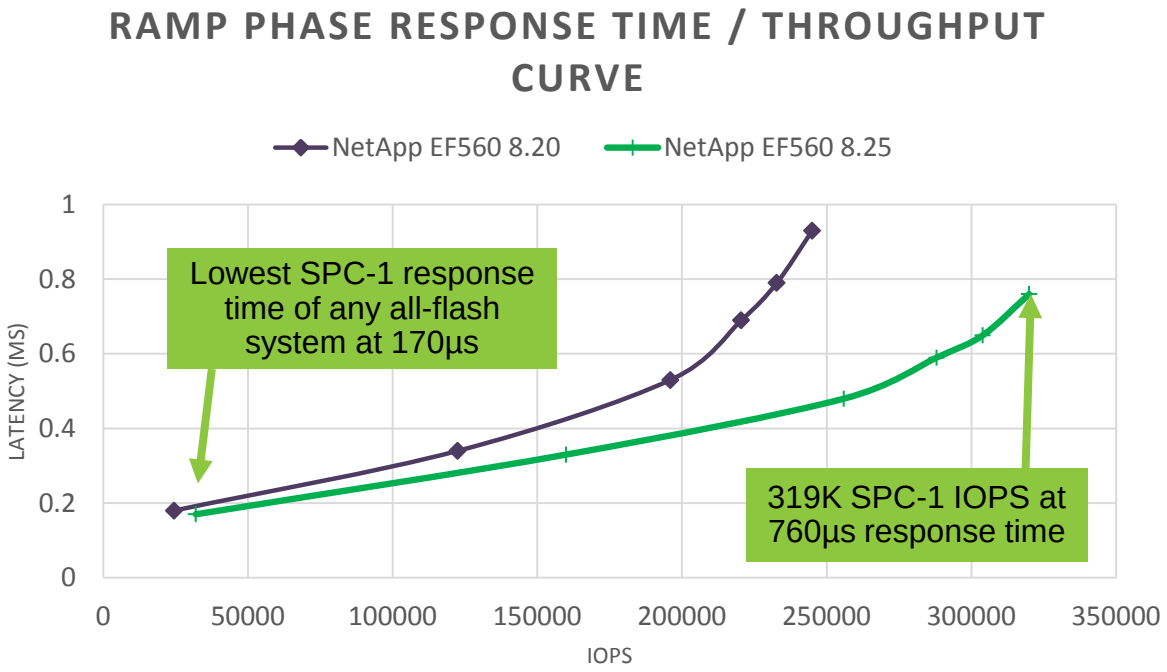
EF560 is a leader in overall solution price

EF560 is the leader $/MB/S solution price

EF560 Performance leadership

The only system to deliver a top-10 ranking in both latency and throughput

Industry-leading all-flash array Latency and IOPS



Industry-leading throughput

SPC-2 Results : By Price-Performance

#	Vendor, Product	$/MB/S	Price in $
1	NetApp® EF560 All Flash	8.12	$92,687
2			916
3			09
4			44
5			93
6			75
7			49
8/9			23

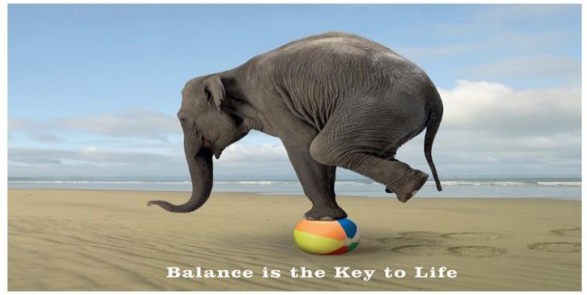
EF560 Results

Overall Composite Score: 11,352 MB/S
Overall $/MB/S Score: #1 $8.12/MB/S
Component/workload scores:
#1 LFP: $9.20
#1 LDQ: $7.08
#1 VOD: $8.37

Faster / Quicker Solutions

Solution Design Strategies - Balanced

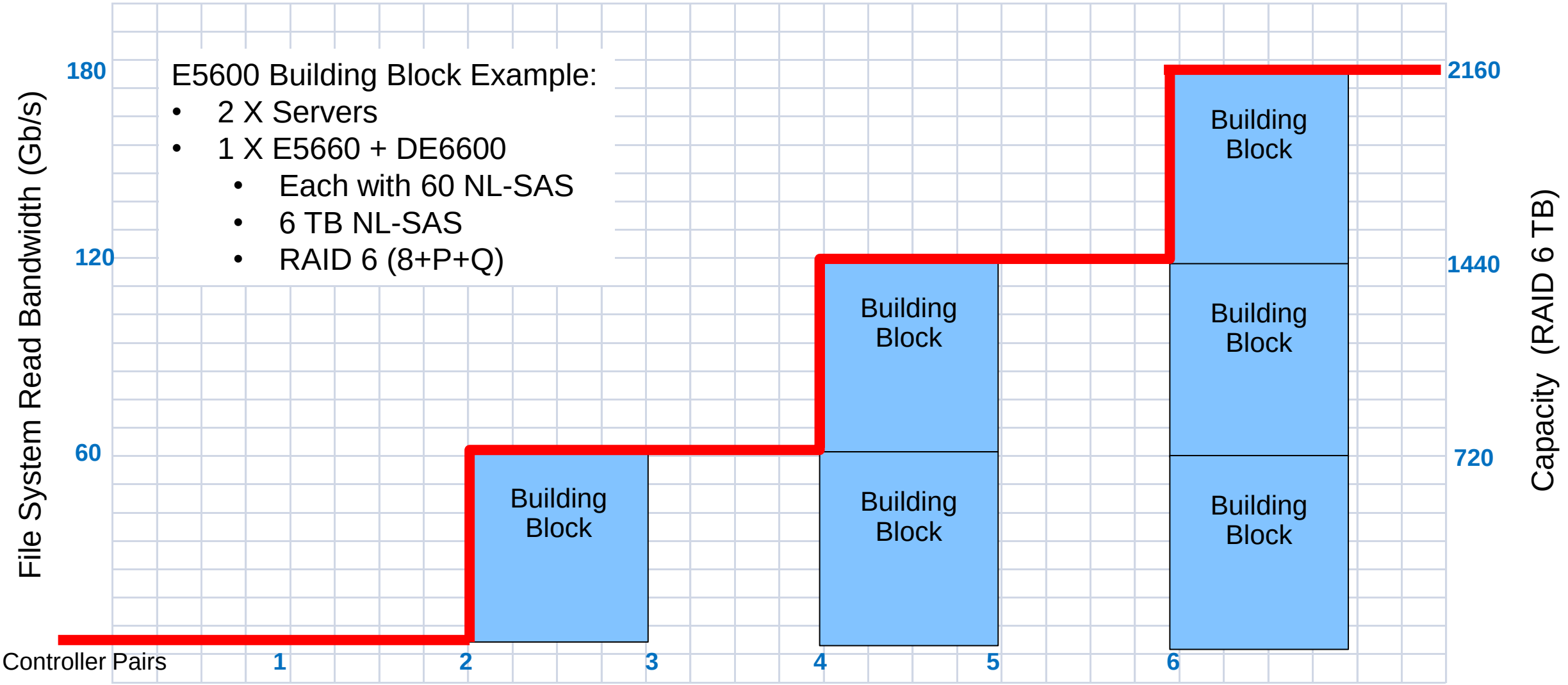
Balanced IO sub-system strategies



- Solutions maximizing capacities
  - Balance # of servers and network adapters with # of controllers. Use a large number of high capacity disks. Bandwidth of disks exceeds bandwidth of controllers
  - Low performance, capacity ratio
- Solutions maximizing performance
  - Balance # of servers and network adapters with # of controllers. Use smaller number of faster disks. Bandwidth of disks matches bandwidth of controllers
  - High performance, capacity ratio
- Solutions providing balanced performance/capacity balance
  - Balance # of servers and network adapters with # of controllers. Use a smaller number of high capacity disks. Bandwidth of disks matches bandwidth of controllers, but capacity is higher
  - Moderate performance, capacity ratio

Flexible scalability – Modular File System

Scale FS performance as you need



Success Stories

National Energy Research Scientific Computing Center - NeRSC

Success Story

National Energy Research Scientific Computing Center: Supporting an Endless Quest for Knowledge

KEY HIGHLIGHTS

Industry
Government

The Challenge
Analyze more data faster; support user demand for research computing services; and accommodate ever-increasing supercomputing power.

The Solution
Improve performance, reliability, and scalability for high-performance computing (HPC) by deploying NetApp® EF-Series all-flash arrays and NetApp E-Series storage systems.

Benefits
- Speeds telescope data analysis sixfold for astrophysics research
- Accelerates project archiving and retrieval for researchers
- Reduces file system scan time by 80%, for fivefold faster backups
- Helps prepare for the coming transition to exascale computing

Solving the Biggest Mysteries of Our Time

The key for pushing science forward has always been taking advantage of the latest and greatest technology. Since 1974, the National Energy Research Scientific Computing Center (NERSC) has used powerful supercomputers to perform research across a range of disciplines.

Located at Lawrence Berkeley National Laboratory, NERSC is a premier HPC facility for scientific research sponsored by the U.S. Department of Energy's Office of Science. More than 6,000 scientists from universities, laboratories, and industry worldwide use NERSC to tackle problems that span everything from climate research, to studying the universe and its evolution, to understanding new materials.

The Challenge

Accelerating scientific discovery

Scientists across a wide range of disciplines are creating increasingly larger datasets at an ever-increasing rate. Keeping up with user demands for storage capacity and performance is an ongoing challenge.

"We support a diverse research community, so user needs can be very different, and they are always increasing," says Jason Hick, storage group leader at NERSC. "Many of our researchers, such as those doing genomic sequencing, need low latency from storage. Others have high bandwidth requirements. If throughput or bandwidth is insufficient, compute jobs can be seriously impacted, which could stall research."

The NERSC Global File System (NGF) provides common filesystems for data that needs to be shared across projects or used on multiple systems. To gather usage statistics and identify changed data that needs to be backed up, NERSC scans its file systems every night.

"We needed better performance for file system metadata," says Rei Lee, a member of the storage systems group. "Nightly scans were taking longer, threatening to push into the next business day. We couldn't let that happen, because it would impede our ability to back up vital research data daily."

"We evaluated a wide range of vendors, looking for flexible, high-density storage that would meet our researchers' requirements," says Jason Hick. "We chose NetApp because it offered the combination of features and functionality we needed."

NetApp EF-Series all-flash arrays have accelerated analysis time six fold. "Being able to do queries on an all-flash array is huge for us," says Peter Nugent. "It used to be that we had to wait an hour from the time the shutter opened on Palomar Observatory's 48-inch telescope to visualize the data. Now, every night, 10 minutes after the shutter opens, we know everything that's on that image.

"With the NetApp systems in place, we've doubled storage capacity for genomic sequencing in support of Department of Energy missions related to clean energy generation and environmental characterization and clean-up," says Ravi Cheema of the storage systems group.

"That's a big productivity improvement, because the previous limit of our disk cache was only five days," says Nick Balthaser, storage systems analyst. "Now, if researchers need to go back to a recent project, they can quickly restore it themselves instead of waiting for us to recover from tape. We're not in their way."

"After putting NetApp EF-Series all flash arrays in, we saw a five fold performance gain in backups alone - Jason Hick"