Dell EMC Storage for Algorithmic Trading

Problem

Algorithmic quant trading firms are in the midst of transitioning to support growing quant teams on increasingly larger data sets as they move their modeling from intraday to multi-day trading. This coupled with the exponential growth in daily transactions means that quants can no longer store the active trade data sets in memory and today's alternative solutions struggle to deliver the required performance at scale.

Dell EMC PowerScale and Isilon

Dell EMC PowerScale delivers the high performance and extreme concurrency at scale from the edge to the entire data set. This shortens model development time with faster analysis on larger, multi-day trading data and extreme performance for smaller real-time data. Combined with its enterprise features, this enables a simple, efficient solution which accelerate cycles of learning by bridging historical and real-time databases while conforming to regulatory standards and guaranteeing enterprise data protection. PowerScale also offers a smaller footprint and lower cost of entry, ensuring seamless upgrades and compatibility with Isilon nodes.

The Ideal Platforms for Tick Data Analytics

PowerEdge R640

The Hot Edge

PowerScale F200

Outperformed NBBO Scores Antuco benchmarks

UP TO **7.7x** FASTER[1]

PowerEdge R940

Massively Parallel IO

Isilon F800

Outperformed 15 of 16 Kanaga benchmarks

UP TO **24x** FASTER[2]

✓ **High Performance Edge for Data Subsets**

✓ **Massive Parallel IO for the Entire Data Set**

STAC Benchmark: What does it mean?

The results showed the sweet spot for PowerScale based solutions to deliver real-time performance on smaller data sets (<10TB) and near real-time performance on large data sets (>10 TB) at high concurrency (100s into the millions).

Accelerate Model Development

Unrivaled All-Flash throughput and IOPS delivering real-time performance at massive scale to accelerate iterative analytic models to take advantage of time sensitive opportunities.

Unbounded Scale

Endless data scale and high concurrency that allows the same massive data sets to be interrogated by many users and thousands of simultaneous quantitative processes.

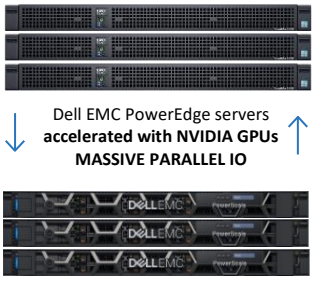
Simple Management and Security

Enterprise grade features delivers out of the box compliance and resiliency, industry leading storage efficiency and transparent data tiering between flash and disk.

[1]The Dell EMC solution is up to 7.7x the speed of a solution involving a parallel file system with 14 database servers and 18 storage servers in the single-user NBBO operation (STAC-M3.β1.1T.NBBO.TIME). Based on a STAC-M3™ Benchmark commissioned by Dell EMC, 'STAC Report: kdb+ 3.6 on 9 x Dell EMC PowerEdge R640 servers accessing a single 3-node Dell EMC PowerScale F200 All-Flash cluster: SUT ID: KDB200914', September 2020, compared to SUT ID KDB200401. Results obtained in STAC-M3 Antuco benchmarks. Full report here: https://stacresearch.com/KDB200914, https://www.delltechnologies.com/en-us/collaterals/unauth/white-papers/products/storage/stac-m3-benchmarks-report.pdf. "STAC" and all STAC names are trademarks or registered trademarks of the Securities Technology Analysis Center, LLC.
[2]Based on a STAC-M3™ Benchmark commissioned by Dell EMC, 'STAC Report: kdb+ 3.6 on 4 x Dell EMC PowerEdge R940 servers with Dell EMC Isilon F800 All-Flash NAS: SUT ID: KDB190430', June 2019, compared to a Lustre-based solution, SUT ID KDB150528. Full report here: https://www.stacresearch.com/news/2019/06/16/KDB190430, https://www.delleмc.com/resources/enus/asset/analyst-reports/products/storage/stac-m3-report-kdb.pdf. "STAC" and all STAC names are trademarks or registered trademarks of the Securities Technology Analysis Center, LLC.

DELL EMC POWERSCALE FOR DEEP LEARNING

Dell EMC Isilon provides flexibility and informed choice with NVIDIA, the leader in AI. By pairing high performance, high bandwidth GPU accelerated compute with high performance, high bandwidth, scale-out flash storage we make AI simple for organizations looking to deploy large scale machine learning and deep learning. Alternatively, for organizations that need to start with a smaller footprint, PowerScale F200 and F600 All-Flash nodes with OneFS 9.0 provides high-performance storage at a lower cost of entry.



All-Dell EMC for AI

BEST OF BREED

Customize your own high-performance scale-out AI environment with the AI optimized Isilon F800, PowerScale and PowerEdge C4140. This solution can also be added as a performance tier to an existing Isilon cluster.

Dell EMC Ready Solution for AI

DEEP LEARNING WITH NVIDIA

This solution is built to simplify AI and deliver faster, deeper insights. Dell EMC's proven expertise lowers risk and shortens deployment times with pre- validated technologies including hardware, AI software, and services.

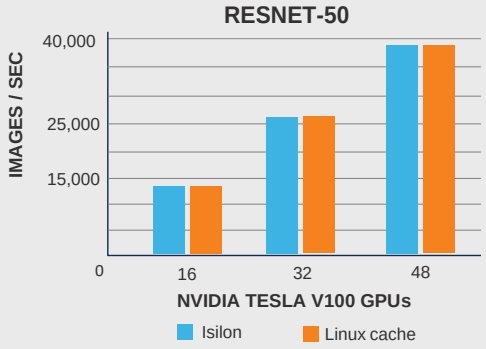
Reference Architecture

ISILON WITH NVIDIA DGX-1 or DGX-2

This architecture pairs NVIDIA's 8-way GPU complex with Dell EMC's scale-out all-flash storage to give more flexibility in deploying large scale AI solutions.

DEEP LEARNING SOLUTION PORTFOLIO VALIDATION

Highlighted below are the results from the ResNet-50 benchmark performing an image classification convolutional neural network (CNN) on labeled images. Using the 148 GB ImageNet dataset as a foundation we made 150 exact copies of each image to create a 'real-life' 22.5 TB dataset which was larger than the server RAM and coherent shared Isilon cache to ensure I/O operations during benchmarking.



SUMMARY BENCHMARK RESULTS

Dell EMC Isilon and NVIDIA DGX-2 GPUs in combination result in:

- Maximize compute ROI with 97% or higher GPU utilization
- The maximum CPU core utilization on the DGX-2 system was 70%. This occurred with ResNet-50.
- There is no significant difference in image throughput between Linux Cache and Isilon
- Linear performance scaling from 8 to 48 GPUs

Read the full performance report here

DELL EMC POWERSCALE ACCELERATES AI BY ELIMINATING THE I/O BOTTLENECK AT SCALE

ACCELERATE INNOVATION

All Flash performance to accelerate model training cycles

INCREASE MODEL ACCURACY

Deeper, higher resolution data sets with access to 10s TBs up to 10s PBs per cluster

IMPROVE DATA SCIENCE PRODUCTIVITY

Flexible in-place analytics and pre-validated solutions for faster, lower risk deployments

SIMPLE SCALE-OUT DEPLOYMENTS

Start small and independently scale-out compute and storage for large scale deployments