Vexata File Solution for Machine, Deep Learning and AI

Solution Brief

PREDICTIVE TO PRESCRIPTIVE ANALYTICS

As analytical pipelines move from descriptive to predictive to prescriptive, they leverage emerging data science techniques such as Artificial Intelligence, Machine and Deep Learning, build upon advanced neural nets for training and advancement. Leading analytics platforms such as SAS, Kdb+ time series database, SAP and TensorFlow, all leverage these advanced analytic techniques.

NEW INFRASTRUCTURE CHALLENGES

These neural nets use iterative techniques and require massively parallel compute infrastructure in the form of GPU's. Equally important is need to feed this analytics compute layer with fast data access, driving the need for low-latency storage leveraging next generation solid state media such as NVMe Flash and 3D XPoint™ .

Existing storage stacks deploy flash media in DAS architectures or within first generation all-flash arrays, which are not architected to utilize the new media performance. DAS architectures require data staging and migration prior to feeding the GPU's. This results in idle GPU compute cycles, increasing the data ingest cycles, delaying the neural net training and inference required to make the data actionable for the deep learning environments.

VEXATA VX-100FS FILE STORAGE SOLUTION

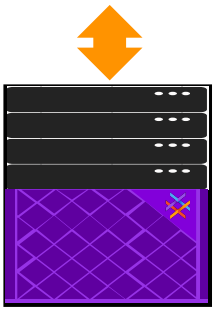
The Vexata VX-100FS File Solution brings unparalleled performance at scale for today's most demanding AI, Machine and Deep Learning workloads. By combining the Vexata Scalable Storage System with a scale-out file system, the VX-100FS enables architects to accelerate next gen data pipelines, gain competitive edge and reduce time to market.

KEY BENEFITS

Accelerate Machine and Deep Learning for Fraud Analytics, Quant Trading and IoT use cases

- 500K Images/sec with Imagenet benchmarks
- Massive parallelism, bandwidth and lowest latency
- 1M+ Random File IOPS
- Sub-millisecond response
- 50GB/s Read, 25GB/s Write, 30 GB/s Mixed bandwidth

NFS, GPFS, HDFS
16 x 40bGE



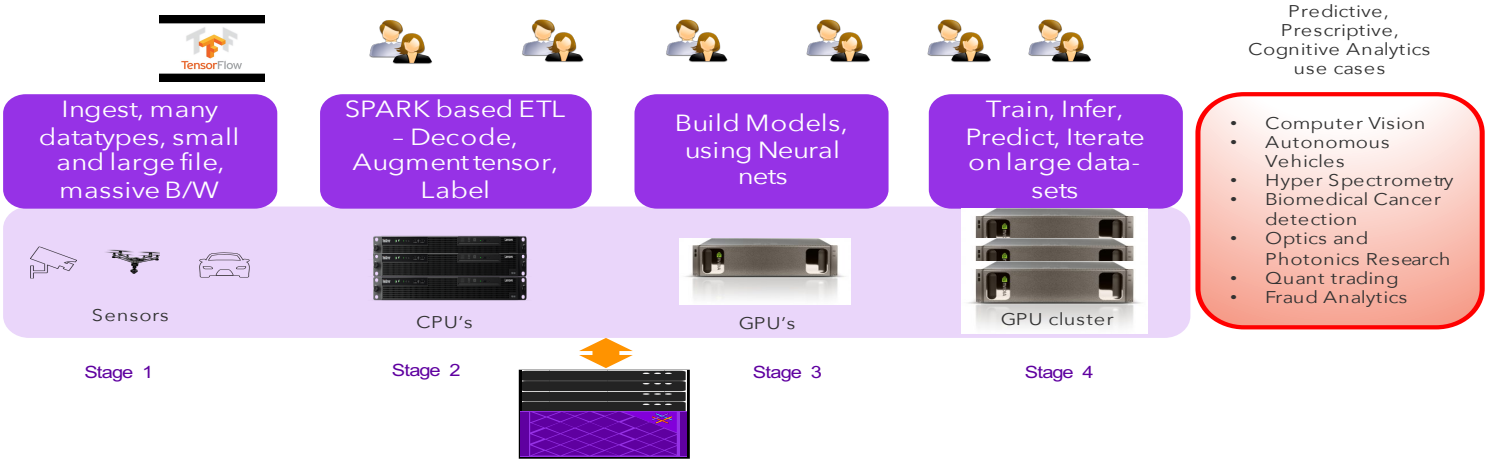
Consolidate Ingest, ETL, Build and Train stages on same shared storage fabric

- Massive ingest, Random R/W, small file handling
- Concurrent work-streams

VX-100FS File Storage System

Optimize Data pipelines

- Optimally utilize valuable GPU cycles, deliver quicker results to Data Scientists
- Aggregate edge data sources, draw fast inference at core
- Enterprise availability, readiness and data management
- Simple to consume operations using Docker pipelines



Vexata File Solution for Machine, Deep Learning and AI

Solution Brief

CUSTOMER CASE STUDY

Customer use case

- Customer utilized advanced Machine Learning/AI pipeline for Computer Vision, Hyper-Spectrometry in this cancer detection use case.
- Pipeline involved ingest from several sensors, ETL with SPARK, neural net modeling on Nvidia DGX-1 and neural net training with DGX-1 Farm.

Customer Challenges

- Siloed pipeline stages were required to meet different storage performance needs, forced to use different storage tiers – high ingest, random R/W, small file handling
- Initially used a DAS storage architecture which needed to be staged to local SSD's, resulted in very long neural net training cycles of 20 days, wasted GPU cycles ($$) and Data Scientist time ($$$)
- Attempted to deploy first generation all-flash based file storage solution, but that system had insufficient bandwidth and small file performance characteristics

Vexata VX-100FS File Storage System

- Customer deployed the Vexata VX-100FS system with a single, shared storage backend that delivered ingest for the entire data pipeline, thus consolidating and eliminating data transfer between stages
- The Vexata solution simultaneously supported 500K images/sec plus 25 GB/s of bandwidth left for Ingest, ETL and modeling stages
- Distributed training time was reduced from over 20 days to 6 hours

Customer Benefits

- Competitive advantage, due to faster time to results and accurate prediction
- Better TCO, due to pipeline stages consolidation and Data scientist timeslots
- Better utilization of Nvidia DGX-1 training farms

Nvidia	Leading All Flash File solution				Vexata Solution			
1 DGX-1	File Size	B/W	**Images/sec**	Ingest, ETL, Model Bandwidth	File Size	B/W	**Images/sec**	Ingest, ETL, Model Bandwidth
	50KB	1 GB/s	20K	9 GB/s	50 KB	2.5 GB/s	50K	47.5 GB/s
	115 KB	1.5 GB/s	13K	13.5 GB/s	115 KB	2.5 GB/s	21.8K	47.5 GB/s
10 DGX-1 Training Cluster	50 KB	10 GB/s	200K	**No Bandwidth**	50 KB	25 GB/s	500K	**25 GB/s**
	115 KB	15 GB/s	113K	**No Bandwidth**	115 KB	25 GB/s	218K	**25 GB/s**

- Imagenet pre-trained models used
- Inception V3, Resnet – 5, Resent – 152, Alexnet, VGG16 container images
- Supervised Learning, labelled images, 1.28M, 1000 categories

ABOUT VEXATA: Vexata is the leader in active data management solutions. Vexata's unique breakthrough enterprise offerings enable transformative performance and scale from database and analytics applications. With unparalleled ability to consume the latest in media like NVMe Flash and now with Intel Optane™ SSDs, Vexata systems deploy simply and seamlessly into existing storage environments. Learn more at www.vexata.com Contact Vexata: info@vexata.com