



# VEXATA VelocityAI FOR AI AND MACHINE LEARNING



## ENABLING PREDICTIVE AND COGNITIVE ANALYTICS VelocityAI USE CASES

Machine Learning (ML) workloads are increasing in volume and complexity as organizations look to reduce training and operational time-lines for artificial intelligence (AI) use cases. This has given rise to massively parallel GPU servers like the Nvidia DGX-1, delivering massive compute power to run these machine learning frameworks.

In order to accelerate training and operational cycles, storage systems that power these AI/ML pipelines must maintain ultra-low latency, massive ingest bandwidth and heavy mixed random and sequential read/write handling. Architectures using direct attached storage (DAS) limits performance and data mobility, while existing all-flash arrays lack the sustained performance to deliver timely insights at scale. Only VelocityAI can deliver the performance and scale of NVMe to accelerate ML workloads.

- Fraud Analytics
- Quant Trading
- Industrial IoT
- Computer Vision
- Speech Recognition
- Hyper Spectrometry
- Biomedical Cancer Detection



VelocityAI

## ACCELERATE ML/AI WORKLOADS WITH VelocityAI

### VelocityAI is purpose built to overcome these machine learning challenges

Reduce training and inferencing time from days to hours, improving data scientist productivity

- Accelerated data path with deterministic low latency performance for better DGX-1 GPU utilization
- Native NVMe solid state performance delivered with standard NVMe-oF (RoCE) and file system interfaces

Access large training and inferencing data-sets

- Accelerated non-blocking access to NVMe media for large data ingest with low latency IO performance.

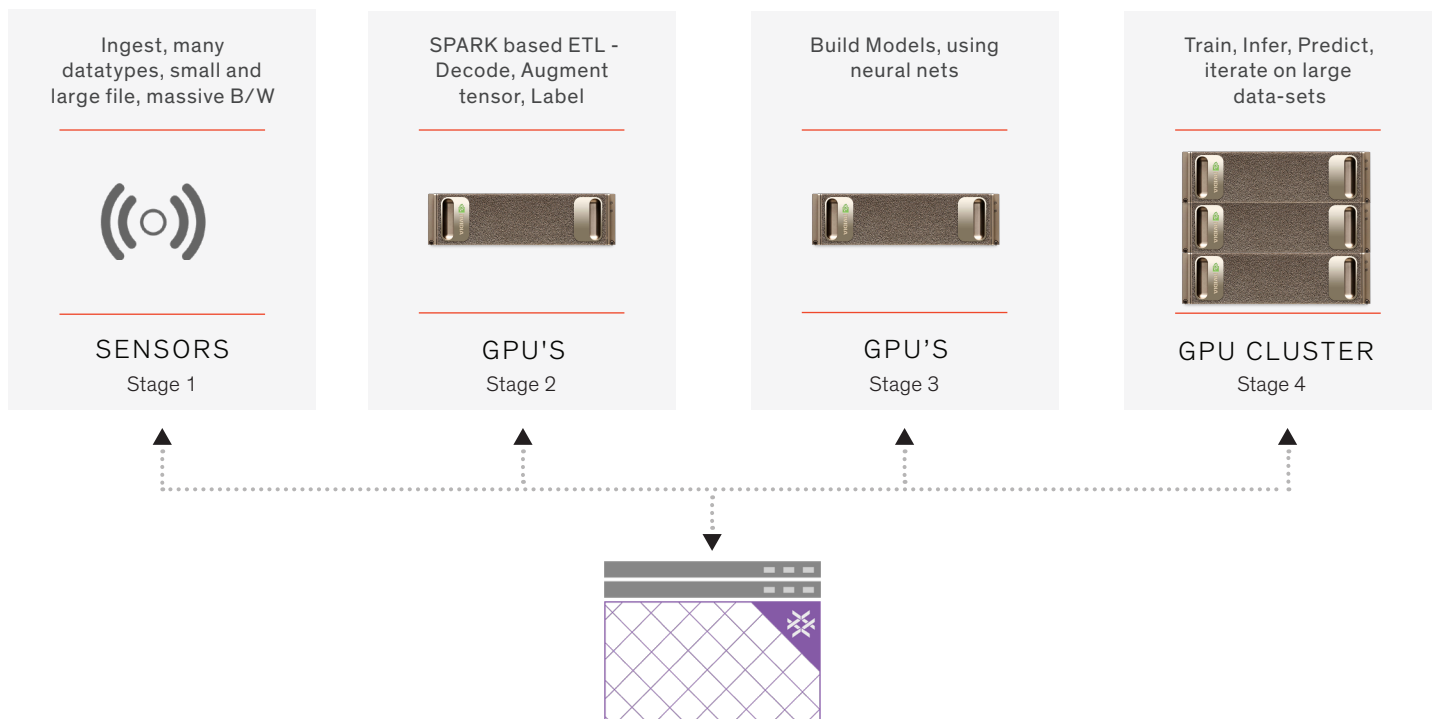
Consolidate and eliminate movement between data pipeline stages

- Shared storage to handle all data pipeline stages without performance degradation
- Simultaneously supports small block random IO , large block sequential IO, mixed Read/Write IO
- In-place data analytics with flexibility of ingest protocols (FC, NVMe-oF, NFS, SMB, S3)

Enterprise storage services, security and data protection

- Instantaneous snapshots and clones, replication, RAID 5/6 protection to eliminate 3 copy replication, compression and 256bit encryption

### Accelerated Data Pipeline





# VEXATA VelocityAI FOR AI AND MACHINE LEARNING



## VelocityAI Reference Architecture

### Compute

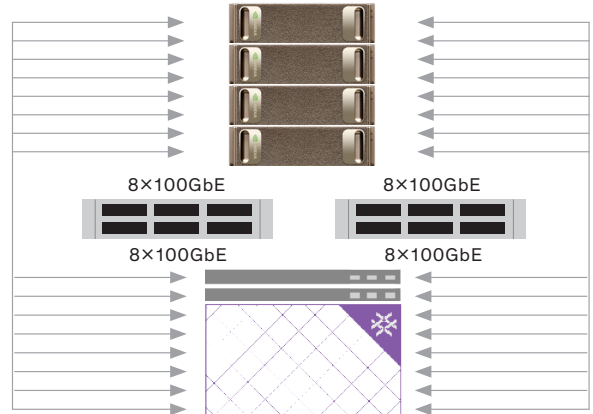
- Four DGX-1 systems (8 Tesla V100 GPU's, 2x Intel E5-2698 v4)
- 4\*100GbE DGX IB ports are configured to run 100 GbE Ethernet
- 4 PFLOPS of Deep Learning performance
- Container based Nvidia GPU Cloud Deep Learning stack with machine learning frameworks

### Networking

- Mellanox SN2700 100 GbE x 32 switch (2 switches)

### Storage

- Vexata VX-100FS NVMe-oF scale-out storage system
- 430 TB of fast file tier
- 50 GB's of bandwidth
- Scale – Add DGX's, add head nodes, add arrays



	VEXATA NVIDIA SOLUTION- 1 DGX SERVERS, 4 BLADES, 2 HEADS			VEXATA NVIDIA SOLUTION- 2 DGX SERVERS, 8 BLADES, 2 HEADS			VEXATA NVIDIA SOLUTION- 4 DGX SERVERS, 16 BLADES, 4 HEADS		
File Size	Available B/W – training/ inference	Images/sec	Remaining B/W	Available B/W – training/ inference	Images/sec	Remaining B/W	Available B/W – training/ inference	Images/sec	Remaining B/W
150 KB	6.25 GB/s	41K	6.25 GB/s	12.5 GB/s	83K	12.5 GB/s	25 GB/s	166K	25 GB/s
1 MB	6.25 GB/s	6.25K	6.25 GB/s	12.5 GB/s	125K	12.5 GB/s	25 GB/s	25K	25 GB/s

### TEST CONFIGURATION:

- Bandwidth equally divided between training/Inferencing and Ingest/ETL/Build
- Imagenet pre-trained model – Alexnet used because it is storage IO heavy
- Inception V3, Resnet – 50, Resnet – 152, Alexnet, VGG16 container images
- Supervised Learning, labelled images, 1.28M, 1000 categories
- Standard docker file - nvcr.io/nvidia/tensorflow:18.04-py2
- Batch\_size = 64
- Horovod = 0.11.3

**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](http://www.vexata.com)

Contact Vexata: [info@vexata.com](mailto:info@vexata.com)