

MATRIX

DEEP LEARNING UNLEASHED

JUMP START YOUR AI

The MATRIX is an end to end Deep Learning solution for fast tracking AI development and deployment. As a comprehensive line of plug & play "DL-in-a-Box" appliances, the MATRIX is packaged with all tools necessary to support complete cycles of Deep Learning development with faster time to productivity. Fully-integrated with Bitfusion Flex, the MATRIX provides a complete dev environment featuring intuitive user interface that natively supports model iteration and optimization of workloads. Advanced features such as batch job scheduling, dynamic user management, and GPU virtualization to aggregate local as well as networked GPUs make the MATRIX on-premise solutions as flexible and feature rich as a public cloud, but with superior performance at significantly better TCO.



[SMART]RACK AI
HIGH-PERFORMANCE
MACHINE LEARNING CLUSTER



MATRIX 280
HIGH DENSITY DEEP LEARNING
SERVER



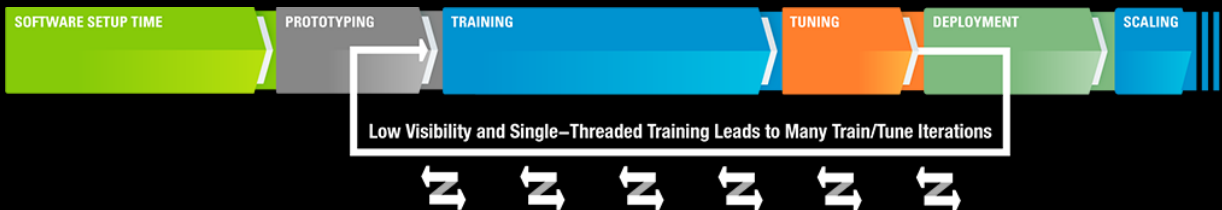
MATRIX 400
THE WORLD'S MOST POWERFUL
DEEP LEARNING DEV BOX

10x Faster Deep Learning Development with MATRIX vs. Do-It-Yourself Approach



“AMAX was instrumental in helping us move off a costly AWS monthly spend to a performance-optimized, on-premise GPU infrastructure featuring their MATRIX products for our Deep Learning workloads.”

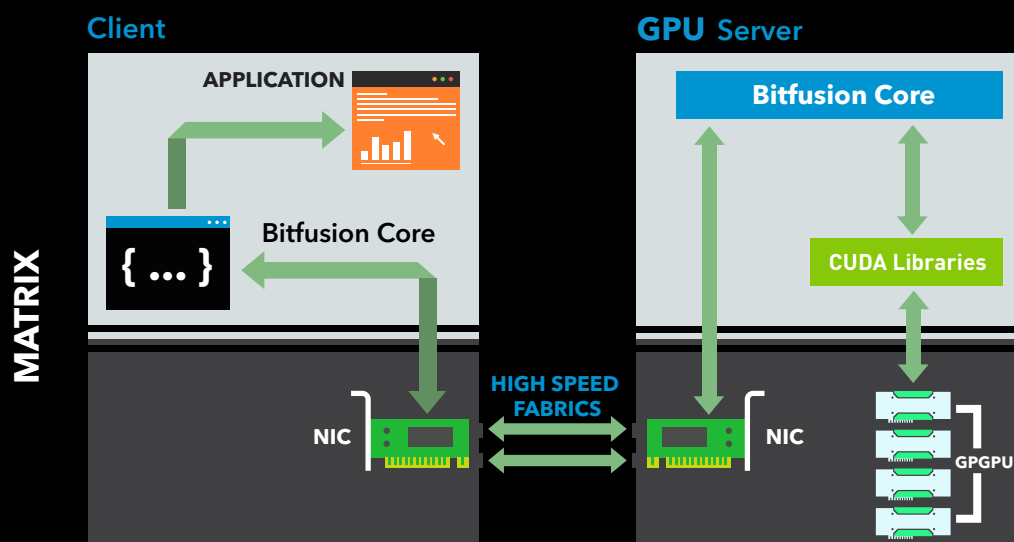
- GEIRI North America



Features

- Fully-inclusive, container-ized Deep Learning solution designed to work out-of-the-box
- Load pre-built Docker environments featuring the latest Deep Learning frameworks and data science libraries within seconds, or build your own
- Pull the latest containers (environments) into cluster with single command without disruption to running workloads
- Full integration with Jupyter Notebook to create Interactive Workspaces that connect seamlessly with your code and data
- GPU over Fabrics Technology enables sharing & scaling of large numbers of GPUs across systems for multi-tenancy and highly-customizable self-service features
- Dynamically allocate GPUs across multiple jobs and users for optimal resource utilization & efficiency
- Attach unlimited network-attached data locations or local NFS file system for seamless access to data
- Easily scalable from single servers to multiple racks to support future expansions.

How the MATRIX Works



User-Friendly Software Interface

The screenshot shows the MATRIX bitfusion FLEX User-Friendly Software Interface. The interface includes a navigation bar with tabs for WORKSPACES, ENVIRONMENTS, and NODES. The main content area displays a table of GPU resources.

GPUs	IP Address	Memory	Allocation	Attached Workspaces	Health
>	10.1.68.217	480.280B	100% 100% 100% 100% 100% 100%	1	OK
>	10.1.62.184	480.280B	100% 100% 100% 100% 100% 100%	1	OK

Below the GPU table, there is a section for CPUs:

CPUs	IP Address	Memory	Attached Workspaces	Health
	10.1.64.13	14.690B	1	OK

The interface also includes an "Expand All" button and a chat icon in the bottom right corner.

- The MATRIX software, powered by Bitfusion Flex, fully integrates Jupyter Notebook application, for a sharable GUI development environment that supports 40 different languages, including Python, Julia, R, Scala, etc.
- Track development workloads anywhere via both GUI and CLI
- Unmatched flexibility in GPU resource management via auto-allocation, and interactive workspace sharing across multiple users and workloads.

The MATRIX was designed as an end to end “DL-in-a-Box” solution with everything needed to fast track AI development and deployment. As a comprehensive line of fully-validated plug & play appliances optimized for Deep Learning, the MATRIX eliminates hardware compatibility and setup overhead so users can quickly achieve time to productivity. Fully integrated with a revolutionary software built upon the Bitfusion Flex suite, the MATRIX provides an iteration environment with intuitive user interface, high visibility and control over resources and workflow, and a compute virtualization capability that enables easy scaling to massive data sets across massive amounts of GPU resources.

Deploy these solutions as standalone engines, or as building blocks to scale highly elastic, self-service AI infrastructures supporting virtually any bare metal, virtualized or container environment.

	MODEL	FF	CPU	GPU	STORAGE	NETWORK	REF #
	MATRIX 200	μWS	1x E5-1600/2600v4	2x GTX 1080Ti	3x 2.5/3.5" drives LSI3008 12Gb/s SAS HBA	Intel X540 Dual 10GbE	Q176759
	MATRIX 400	WS	1x E5-1600/2600v4	4x GTX 1080Ti	2x 5.25", 2x3.5" 4x 2.5" bays 10x SATA 6Gb/s	Intel X550 Dual 10GbE	Q176758
	MATRIX 140	1U	2x E5-2600v4	8x Tesla P40/P100	2x 2.5" hot swap 2x 1.8"SSD	Dual 25GbE (RoCE) IB FDR opt.	Q176762
	MATRIX 280	2U	2x E5-2600v4	8x Tesla P40/P100	6x 2.5" hot swap 2x 2.5"NVMe	Dual 25GbE (RoCE) IB FDR opt.	Q176763
	MATRIX 480	4U	2x E5-2600v4	8x Tesla P40/P100 Single Root Complex	24x 2.5" hot swap	2x 1G (opt. 1) 2x 10G (opt. 2)	Q176764
	[SMART] Rack VM	42U	Up to 24x E5-2600v4	Up to 96x Tesla P40 for 1152 TFLOPs/ 96x P100 for 864 TFLOPs	Up to 92Tb All-Flash HA (StorMax™ SDS)	Dual 25G to each node 1G management network	Q176404 (42U rack)

