



Joint Solution Brief
Artificial Intelligence

AI Data Scientists Test Drive

Equinix Artificial Intelligence (AI) Platform as a Service test drive powered by NVIDIA, NetApp and Core Scientific.

Overview

Equinix, in partnership with NVIDIA, NetApp and Core Scientific, has announced an Artificial Intelligence Platform as a Service (AI PaaS) test drive at Equinix data centers. This solution brings together industry-leading AI infrastructure from NVIDIA and NetApp, in combination with the global interconnection-rich data center platform from Equinix and best-in-class data science workflow software from Core Scientific.

With the test drive program, enterprises can try AI PaaS at Equinix and deploy when they are ready.

Challenge

There is no doubt that groundbreaking machine learning (ML) algorithms and the rise of deep learning (DL) have enabled businesses to attack some of their toughest challenges with the power of AI. Demand for AI compute power will only accelerate in the future. To empower data scientists to prototype, test and iterate on their AI models more rapidly, companies require direct access to advanced computing power and fast local storage. Proximity of datasets to compute resources becomes paramount, and in many use cases, data scientists need access to data coming from external sources such as sensors, cameras and equipment at the edge of the enterprise. To access that data efficiently, they need to host their AI infrastructure at a well-connected location with high-speed, secure access to multiple clouds, private data centers and data brokers. It has become apparent that traditional CPU-based infrastructure constrains data science workflow and does not scale to meet the demands of training increasingly large AI models and datasets.

Data scientists need unfettered access to powerful tools and a platform that enables rapid iteration to deliver the best models with the highest predictive accuracy, in the shortest time possible, if businesses are to reap returns on their AI investments. Data science teams expect “ready-now,” IT-approved AI platforms that can be provisioned quickly and easily from a web UI. Successful AI development requires teamwork with orchestration of hardware and software, including GPU computing in tandem with high-performance storage, data management and high availability. Furthermore, concerns about data center readiness, data locality, escalating cost,

[Equinix.com](https://equinix.com)

About NVIDIA

NVIDIA pioneered accelerated computing to solve problems that normal computers cannot solve. The company innovates at the intersection of graphics, HPC and AI.

[NVIDIA.com](https://nvidia.com)

About NetApp

NetApp is the leader in cloud data services, empowering global organizations to change their world with data. Together with our partners, we are the only ones who can help you build your unique data fabric.

[Netapp.com](https://netapp.com)

About Core Scientific

Core Scientific’s mission is to be the premier AI and blockchain provider, delivering best-in-class infrastructure and software solutions for a rapidly evolving market.

[Corescientific.com](https://corescientific.com)



EQUINIX

scalability, and the rapidly evolving nature of AI innovation and infrastructure can easily delay time to insights. These challenges necessitate a full-service solution that can empower IT and their data science users.

Due to data residency, compliance, performance requirements (of moving large datasets to far-off core clouds) and cost reasons (for backhauling large datasets to core clouds), it becomes critical for businesses to place their AI compute infrastructure co-resident with their data, following the mantra of “train where your data lands.”

Furthermore, AI platforms cannot exist in isolation. They need to be integrated via secure and high-speed networks to an enterprise’s corporate IT systems, which can exist in private data centers and third-party IaaS and SaaS Platforms. Thus, there is a need to connect AI systems with the rest of an enterprise’s IT infrastructure.

Finally, AI platforms need to be situated where they can ingest data from multiple sources to fuel model prototyping and improve accuracy. In many instances these datasets reside in multiple public clouds, data brokers and data centers, and they are also getting generated at the edge. Thus, it is best to host the AI infrastructure at an interconnection hub that has high-speed and secure access to these different data sources.

Solution

Equinix, NVIDIA, NetApp and Core Scientific together offer a fully integrated AI PaaS, available in the test drive program, to help businesses meet the rapidly growing demands of AI development in every industry. The key highlights of this solution are:

- **AI PaaS**
A cloud-native (container-based) AI as a Service offering that makes it easy for data scientists to consume AI services.
- **Industry-Leading AI Technology Stack**
Consisting of high-performing compute, network and storage technologies that are optimized for running all the major AI software frameworks.
- **AI Stack at Global Interconnected Metro Edge**
A global and highly interconnected data center platform that provides high-speed and secure interconnection to IT systems and data sources that are spread across public cloud, private data centers and edge locations.

AI Platform as a Service

This new service at Equinix brings together the critical infrastructure needed for AI innovation, including the world’s leading AI compute system from NVIDIA, with cloud-connected, all-flash storage to provide an on-demand AI solution. It includes Plexus™ from Core Scientific, which includes AI workflow tools, GPU-optimized software containers and advanced orchestration and scheduling, with a single web-based interface to streamline AI model development.

Features

Trusted Partners

Rely on the solution that combines key strengths of Equinix, NVIDIA, NetApp and Core Scientific.

AI at the Metro Edge

Perform AI at 210+ securely connected Equinix International Business Exchange™ (IBX®) data centers worldwide in 56 global metros.

Simplified AI as a Service

Build the data science you need with high-performing tooling, workflows and resource management.

Industry-Leading AI Technology Stack

Run containerized AI frameworks and ML algorithms on industry-leading GPU and storage technologies.



EQUINIX

 Core Scientific Plexus™

Applications			
Web/Shell Access		API Access	
Fastest Data Analytics    		Accelerated AI Platforms    	
Management			
Resource Sharing/Scheduling	Billing	Workload Tuning	High-Availability Operations
Orchestration			
 Kubernetes	 Slurm	Elastic Public Cloud	
AI Optimized Hardware			
Nvidia DGX A100 5 Peta FLOP		NetApp All-Flash Storage Cloud-Connected Data Management	
AI Hosting			
 Global Platform 210+ Data Centers, 56 Markets, 26 Countries	Connected Hub 10,000 Companies, Cross Segment	Connected Metro Edge Low-Latency Public Cloud On-Ramp, Global ECX Fabric®	

Ease of Use with Core Scientific Plexus

- Walk-up tooling, workflows and resource management for data scientists.
- High Availability Kubernetes and Slurm cluster management.
- GPU application portal including NGC containers and leading GPU analytics software applications OmniSci, BrytLyt, FastData.IO and SQream.
- Accelerated AI platforms: TensorFlow, Rapids, PyTorch and MXNet.
- OPEX model with on-demand scaling, both colocated and bursting to the public cloud. (In the test drive setup, there is no bursting of processing to public clouds.)
- Unlike in the public cloud, there are no ingress or egress fees for moving data in or out of the AI setup at Equinix.

Benefits

Keep Control Over Your Data

Avoid cloud lock-in and keep control over sensitive persistent data.

Control Costs, Cut Latency and Satisfy Compliance Regulations

Process data at the edge to cut data transfer costs, perform real-time inferencing and satisfy data residency requirements.

Interconnect with Global Ecosystems

Secure and high-speed access to data sources at financial services, media networks, clouds and enterprise ecosystems.

Scale on Demand

Build capacity to handle surges in data and step it down to save when business slows.



NVIDIA DGX A100 15_{TB} 15TB Gen4 NVME SSD <ul style="list-style-type: none">25GB/sec Peak Bandwidth2X Faster than Gen3 NVME SSDs	6_x 6x NVIDIA NVSwitches <ul style="list-style-type: none">4.8TB/sec Bi-directional Bandwidth2X More than Previous Generation NVSwitch	8_x 8x NVIDIA A100 GPUs with 320GB Total GPU Memory <ul style="list-style-type: none">12 NVLinks/GPU600GB/sec GPU-to-GPU Bi-directional Bandwidth
1_{TB} 1TB RAM <ul style="list-style-type: none">3.2X More Cores to Power the Most Intensive AI Jobs	DUAL 64_{CORE} Dual 64-core AMD Rome CPUs <ul style="list-style-type: none">3.2X More Cores to Power the Most Intensive AI Jobs	9_x 9x Mellanox ConnectX-6 200Gb/s Network Interface <ul style="list-style-type: none">450GB/sec Peak Bi-directional Bandwidth

Industry-Leading AI Technology Stack

The technology stack consists of industry-leading compute, network and storage hardware, along with an AI deep-learning software stack consisting of many open source AI frameworks that have been optimized to run on the [NVIDIA DGX™ A100 systems](#). This technology stack provides industry-leading performance on the leading AI benchmarks for both training and inferencing.¹

■ NVIDIA DGX A100

Is the universal system for all AI workloads, offering unprecedented compute density, performance and flexibility in the world's first 5 petaFLOPS AI system. NVIDIA DGX A100 system² features eight of the world's most advanced accelerator, the NVIDIA® A100 Tensor Core GPU, enabling enterprises to consolidate training, inference and analytics into a unified, easy-to-deploy AI infrastructure.

■ NVIDIA NGC™

A repository of containerized AI frameworks and ML algorithms that have been optimized for the NVIDIA DGX systems and are available as part of this software stack.

■ NetApp AFF A800 Storage System

Cloud-connected, all-flash storage that spans from edge to cloud, can scale up to 20 petabytes and can support up to 400 billion files. Test drive will be a scaled-down version of this storage capacity.

■ NVIDIA Mellanox® Networking

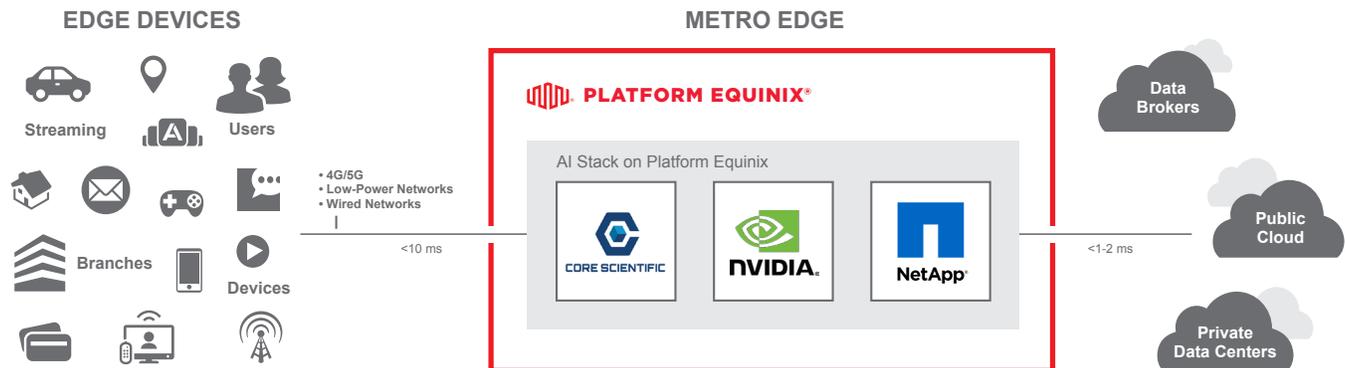
Providing a high-performance AI network fabric connecting each DGX A100 system, with ultra-low latency supporting 100 Gigabit Ethernet connectivity.

¹ NVIDIA Shatters Big Data Analytics Benchmark, Scott McClellan, June 22, 2020 <https://blogs.nvidia.com/blog/2020/06/22/big-data-analytics-tpcx-bb/>

² NVIDIA white paper <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/nvidia-dgx-a100-datasheet.pdf>



EQUINIX



AI Stack at Global Interconnected Metro Edge

In addition to having an industry-leading fully managed AI stack, it is important to host this stack at an interconnection hub that is close to the edge and public clouds for performance, cost and compliance reasons. Hosting an AI stack at an Equinix colocation data center provides the following benefits:

- Global presence and consistent service across 56 markets in 26 countries. This helps to meet enterprise requirements to deploy in multiple locations to satisfy government regulations for data residency.
- Global high-speed and secure interconnection fabric connecting these data centers so that you can securely do AI processing across multiple regions.
- Proximity to the edge in most metros (less than 10 ms from the end devices) allows for AI inferencing and training at the metro edge. This, in turn, provides cost and latency benefits.
- High-speed and low-latency connectivity to the public cloud (between 1 and 2 ms to most major clouds in most markets). This architecture allows customers to have hybrid multicloud AI architectures.
- Ecosystem of 9,700+ companies consisting of clouds, network providers, financial companies, media companies and other enterprises at Equinix. Thus, this AI service allows companies to easily perform AI on their own data at Equinix and get high-speed and secure access to data from other partners. In addition, new companies that join this ecosystem at Equinix can leverage the benefits of AI as a Service.
- Retail colocation data centers that can support from small (10 kW) to larger AI infrastructure deployments (~300 KW).



EQUINIX

Use Cases for This Solution

AI at Equinix is specifically designed for use cases where:

- Enterprises want to do AI training or inference in their private data centers instead of in public clouds for control, cost, performance and privacy reasons. But many may not have the facilities optimized for the unique resource demands of AI infrastructure, and for the complexity of managing AI hardware and software platforms. Thus, enterprises want to access AI Infrastructure as a Service at a colocation data center.
- Datasets need to get processed at the edge instead of hauling the data back to a remote core data center, for cost, latency, privacy and compliance reasons.
- AI applications need to integrate with enterprise IT systems or need to access external data from multiple sources such as clouds, private data centers, data brokers and edge locations. In these hybrid multicloud use cases, it is best to host the AI infrastructure at an interconnection hub where businesses can integrate with distributed IT systems via high-speed and secure networks.
- More than 9,700 enterprises and providers already have their infrastructure at Equinix and are interconnected to each other. We want to make it easy for both existing customers and new customers to do AI processing while leveraging data from this ecosystem.

More Information

Equinix, along with its partners NVIDIA, NetApp and Core Scientific, is offering the AI PaaS test drive at Equinix. By using test drive for both training and inferencing, you can assess the capabilities and simplicity of this joint AI as a Service solution and its benefits. After using test drive, you can deploy this AI platform at Equinix. If you are interested in trying out test drive for free at Equinix, you can find more information at www.equinix.com/AI/TestDrive.

About Equinix

Equinix is the world's digital infrastructure company. Digital leaders harness our trusted platform to bring together and interconnect the foundational infrastructure that powers their success. We enable our customers to access all the right places, partners and possibilities they need to accelerate advantage. With Equinix, they can scale with agility, speed the launch of digital services, deliver world-class experiences and multiply their value.

Equinix.com



The global interconnection platform for a cloud-first world

Globally deploy your infrastructure and services wherever opportunity leads. Directly and privately interconnect to your most important clouds, services and networks. Activate edge services on demand to scale for success. On Platform Equinix®, you'll reach everywhere, interconnect everyone and integrate everything you need to create your best future. Get digital ready with Equinix.

Equinix.com

© 2020 Equinix, Inc.
NVIDIA Equinix AI Solution Brief_US-EN | 12554 | v101620 | 1020 | Q120

Questions? Equinix.com/Contact-Us