S62794 - Build Secured AI Cloud Infrastructure at Scale

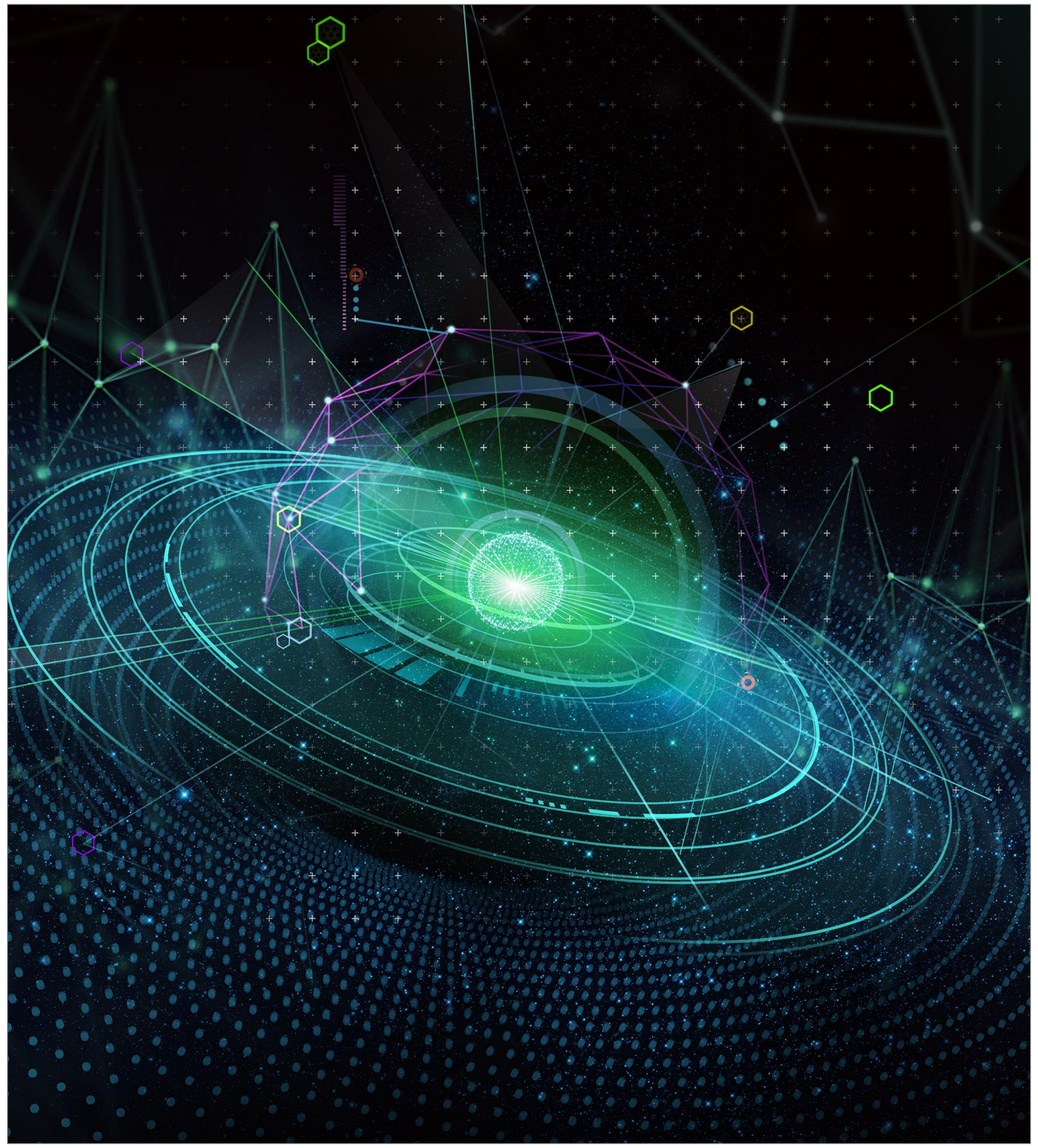
Gera Dorfman, Vice President Network Security, Check Point
Yaël Asseraf Shenhav, Vice President of Product, Networking, NVIDIA

Generative AI is Transforming Cloud Data Centers

Modern Data Centers are Undergoing a Seismic Shift to Meet the Demands of Generative AI

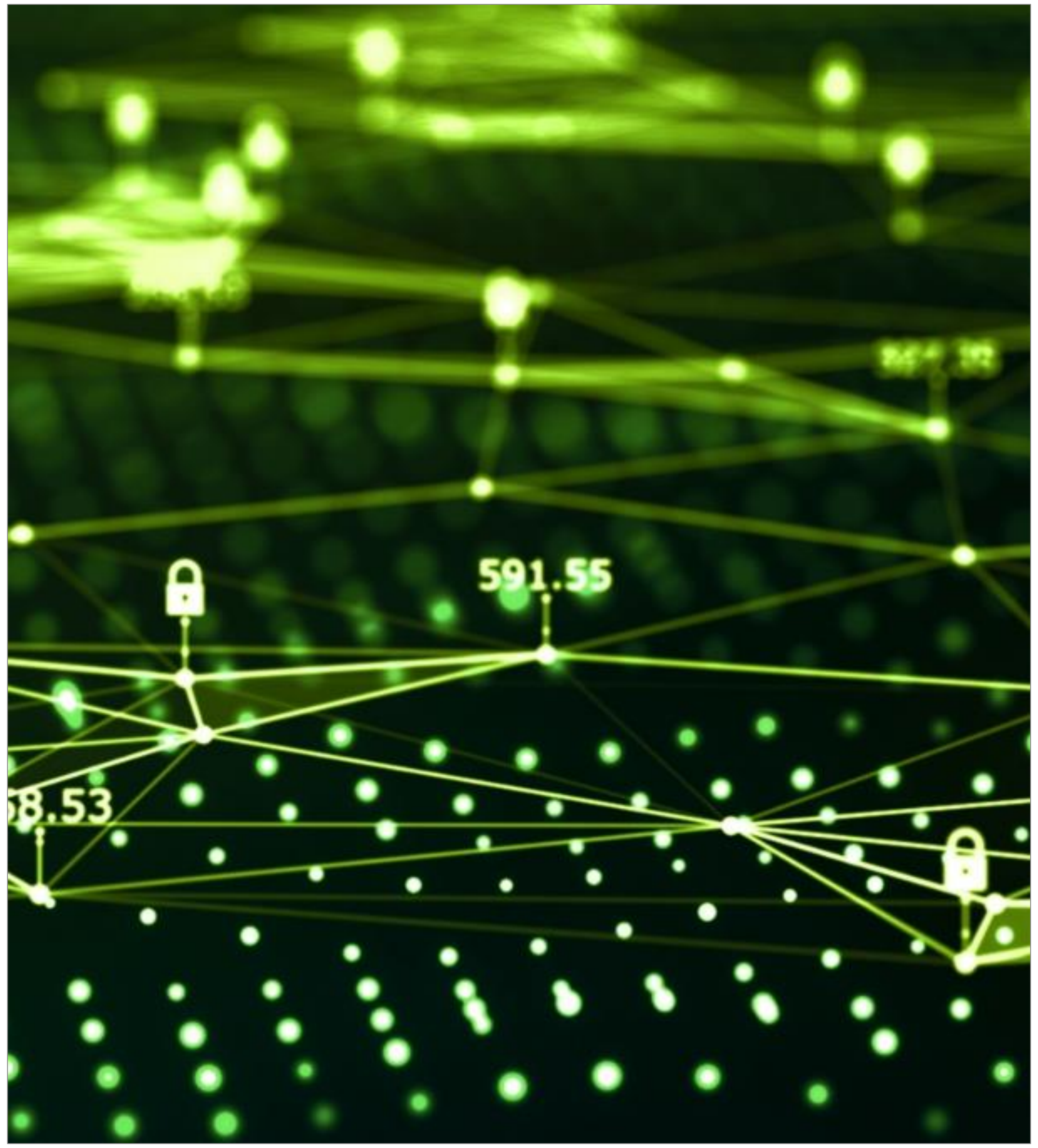
Rise of GPU Computing

Cloud data centers prioritizing GPU-accelerated infrastructure



Massive Data Volumes

Advent of AI data platforms for expanding data capacities

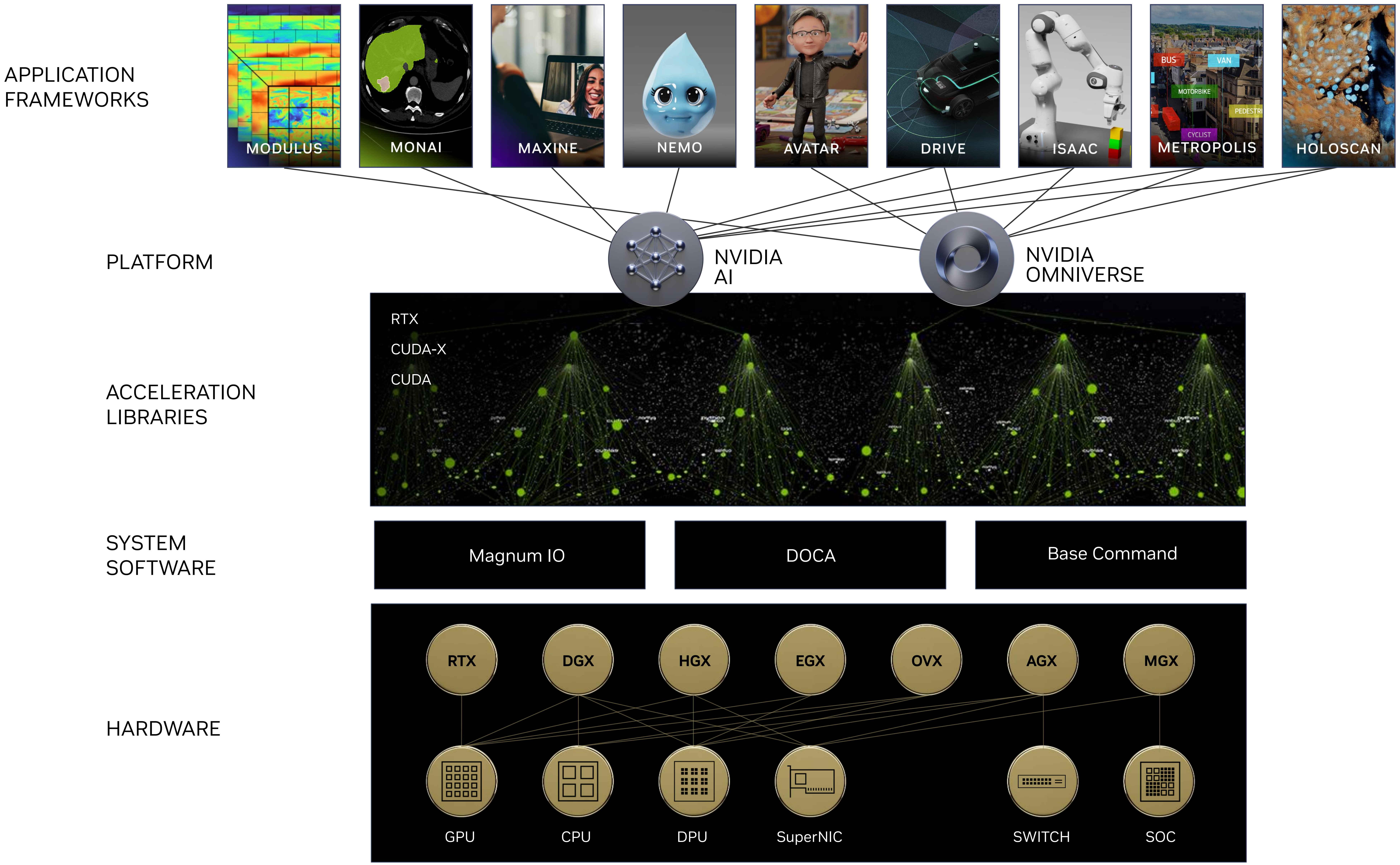


New Frontier in Cybersecurity

New level of sophistication is emerging in cyber attacks

NVIDIA Accelerated Computing for Generative AI

Accelerated Computing Services, Software and Systems Enabling New, Enhanced Business Models



APPLICATION FRAMEWORKS

MODULUS, MONAI, MAXINE, NEMO, AVATAR, DRIVE, ISAAC, METROPOLIS, HOLOSCAN

PLATFORM

NVIDIA AI

NVIDIA OMNIVERSE

ACCELERATION LIBRARIES

RTX

CUDA-X

CUDA

SYSTEM SOFTWARE

Magnum IO

DOCA

Base Command

HARDWARE

RTX, DGX, HGX, EGX, OVX, AGX, MGX

GPU, CPU, DPU, SuperNIC, SWITCH, SOC

New Class of Cloud Data Centers for Generative AI

Workload Performance is Paramount

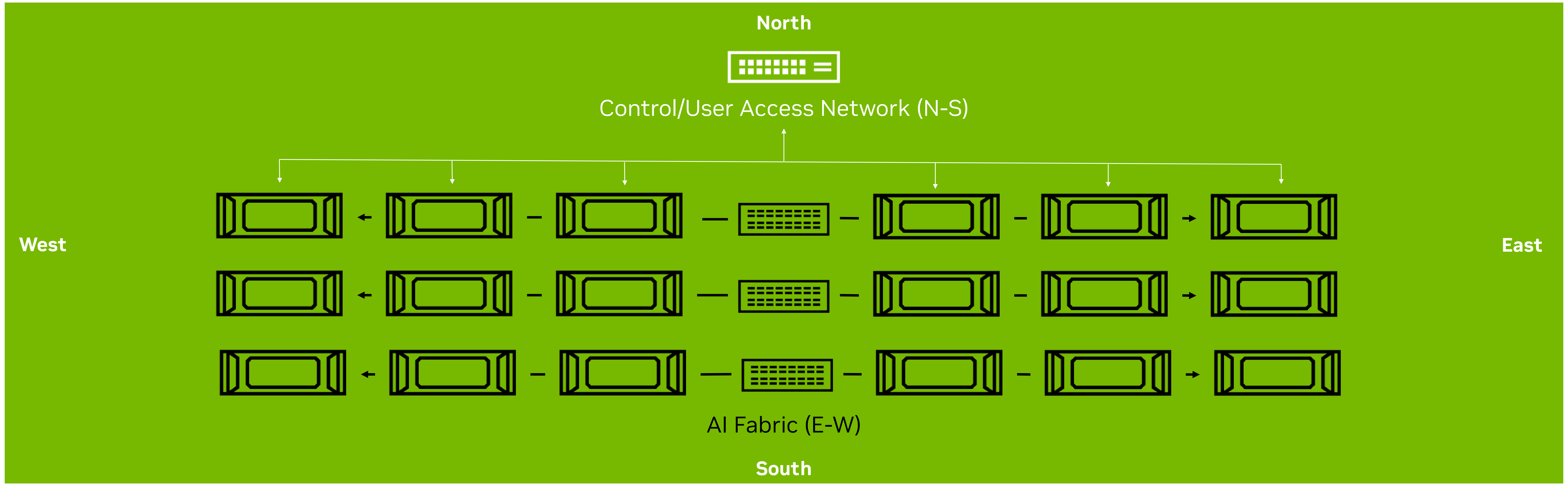
NVIDIA DGX SuperPOD

Purpose-built for the unique demands of AI

- Highest Efficiency and Performance
- Consistent & Low-Jitter Performance
- Maximized Networking and Compute Availability
- Multi-tenancy – Zero Trust and Performance Isolation
- Fully Integrated Solution

Modern AI Cloud Data Center Architecture

Optimized Networking for Peak AI Workload Efficiency



- Multi-Tenant Environments
- Increased Attack Surface
- Explosion of Data, Users, Devices, and Apps
- AI Cyber Threats

AI Cloud Security Requires a New Paradigm

Traditional cybersecurity tools not equipped to protect generative AI cloud data centers

CHECK POINT™

The LEADING Global Cyber Security Company

GLOBAL LEADER

100,000+ CUSTOMERS,
88+ COUNTRIES, 6,200+ PARTNERS

CUTTING-EDGE TECHNOLOGIES

OVER 30 YEARS OF EXPERTISE,
INDUSTRY'S MOST VISIONARY PLAYER

INNOVATION LEADERSHIP

HIGHEST NUMBER OF AI REAL-TIME
PREVENTION TECHNIQUES

6,500+

EMPLOYEES WORLDWIDE,
TOP TALENT

TRADED ON NASDAQ

1996 I CHKP

WORLD'S BEST EMPLOYER

BY FORBES,
#1 CYBER SECURITY VENDOR

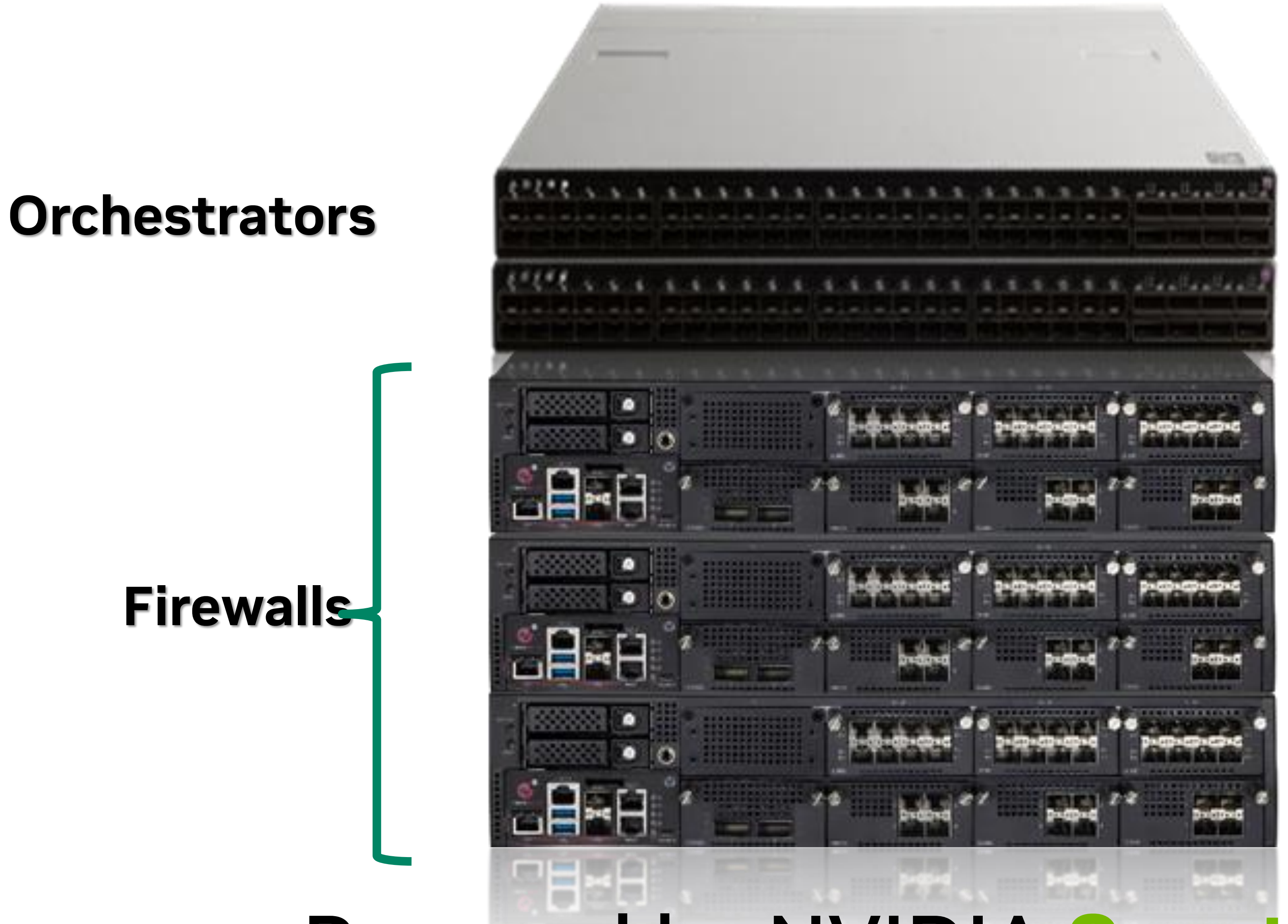
TRUSTED BY

FORTUNE 500

COMPANIES

Check Point and NVIDIA Long History of Innovation

Check Point Maestro Hyperscale System
w/ Intelligent Load Sharing



Powered by NVIDIA **Spectrum** networking

Check Point Quantum Force Security Gateways

Powered by NVIDIA **ConnectX-7**

THREATCLOUD AI

50+ AI engines powered by **CUDA**

AI Training Attacker Tactics, Techniques, and Procedures

- Confidentiality
  - **Data Exfiltration** to expose intellectual property
    - Training data set
    - AI/ML model, artifacts

- Integrity
  - **Backdoor the model** to change its behavior to produce incorrect output
    - Adjust the model weights to control the model output
    - Include malicious code with a downloaded model to try and infect the host machine with malicious code
  - **Data Poisoning** - Poison training data set to use specific triggers to influence the output of the model

- Availability
  - **Denial of Service** - Attack component(s) of AI infra to degrade performance and reduce its capacity

[Note] Prompt Attacks are another attack vector, but are not relevant for Training

The Implication of a Security Incident

A successful attack against AI infrastructure will mostly result in a significant financial loss

Research shows that minor changes to training dataset may render the model useless at long term

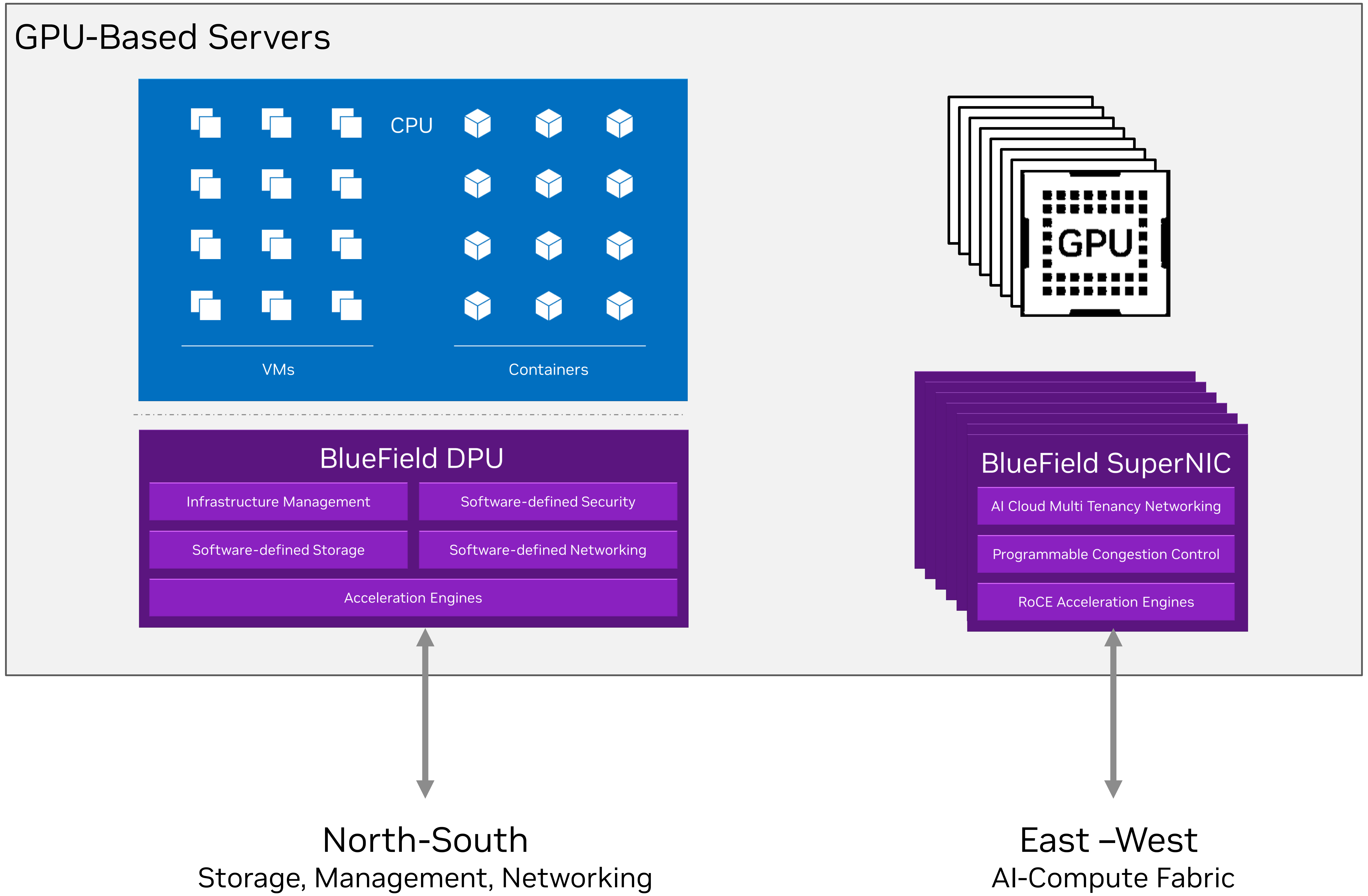
Considering the costs of model training:

- For example: training GPT3 on **40B** tokens would take about a **month** on **160 GPUs** and cost **$380k**

Access control to AI cloud infrastructure has higher significance compared to traditional datacenters

NVIDIA BlueField-3 AI Platforms

DPUs and SuperNICs


GPU-Based Servers
CPU
VMs
Containers
GPU
BlueField DPU
Infrastructure Management
Software-defined Security
Software-defined Storage
Software-defined Networking
Acceleration Engines
BlueField SuperNIC
AI Cloud Multi Tenancy Networking
Programmable Congestion Control
RoCE Acceleration Engines
North-South
Storage, Management, Networking
East –West
AI-Compute Fabric

Runtime Security for Modern Data Centers & AI Clouds Using BlueField

A Security Sentry Validating the Integrity of Compute Nodes

BLUEFIELD

Automated Response

Firewalling, Segmentation & Visibility

Network & Application Behavioral Monitoring

Drift Detection

Indicators of Compromise

Application Network Access Control

Application Control

Realtime Breach Detection

Realtime Visibility

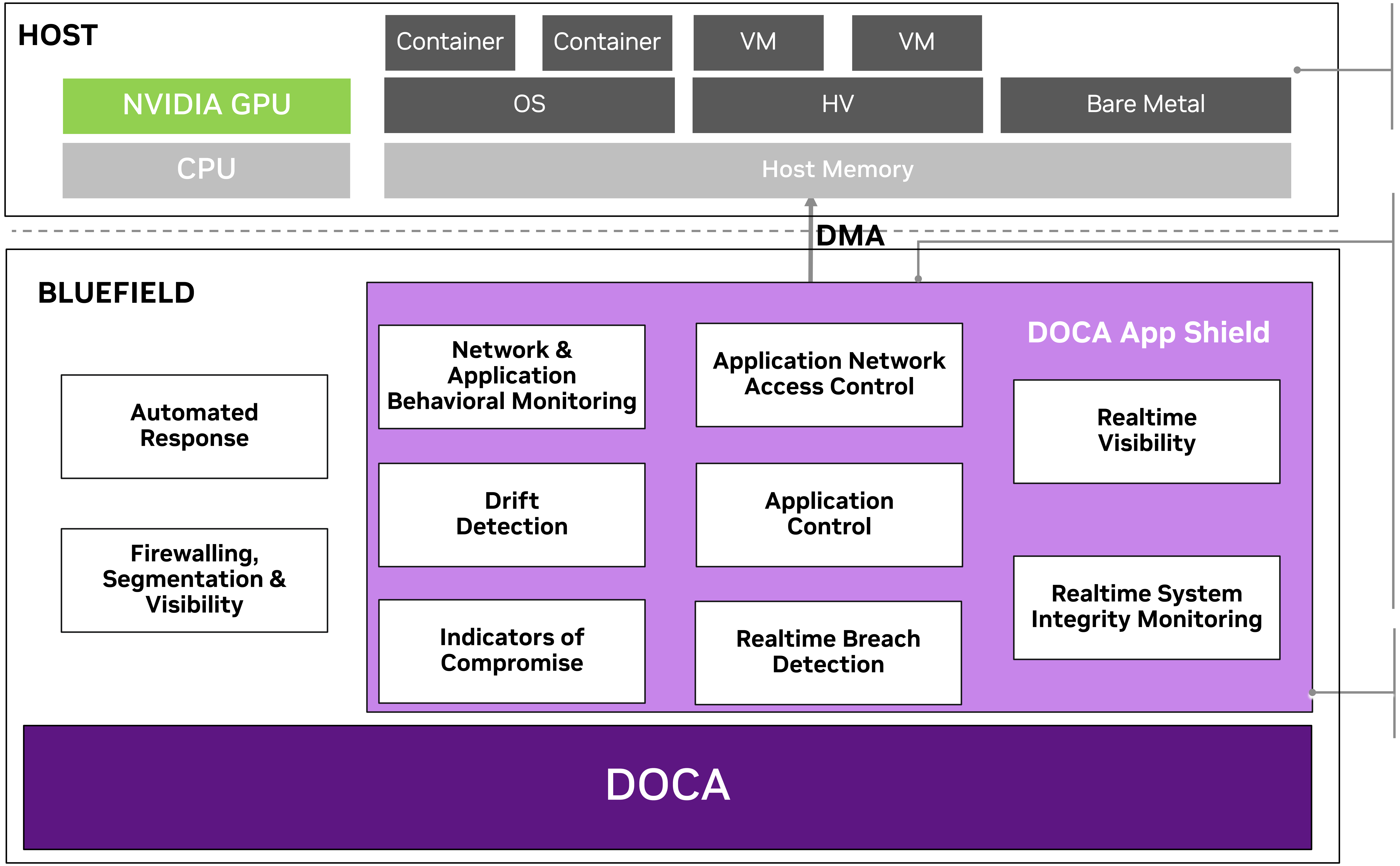
Realtime System Integrity Monitoring

DOCA

SIEM / XDR / SOAR

NVIDIA Morpheus

DOCA App Shield Service



Apps & services running on host, protected by App Shield

Provides comprehensive information about host compute nodes by analyzing memory.
For example, providing information about processes, connections, the integrity of running processes, and more

App Shield Runtime security feature set

CHECK POINT™ + NVIDIA®

Partnering to Deliver
Secure AI Cloud Solutions

Take Your AI Security to the Next Level

Comprehensive Network and Host Level Security

Tailored Security at the network and application level designed to protect AI cloud Infrastructure.

Out-of-the-box Security with Effortless Deployment

Plug and Play deployment using standard orchestration tools at high scale

Zero Impact Over AI Performance

Running in BlueField with no impact on AI performance

Comprehensive Network and Host Level Security

Tailored Security at the network and application level designed to protect AI cloud Infrastructure.

identities based on **IdP, Geo Location, Network**

content type based on **data type signatures**

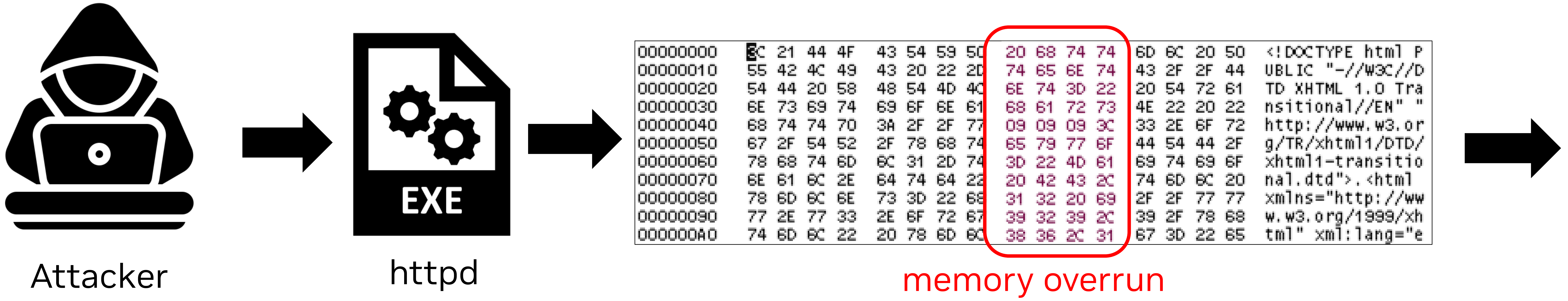
No.	Name	Source	Destination	Services & Applications	Content	Action	Track
▾ AI Infrastrcutre Access (1-2)							
1	Upload data to LLM	Data Science Team	HGX AI servers	https	Upload Traffic Source Code	Accept	Extended Log Accounting
2	Admin Access	AI admins	HGX AI servers	ICMP Protocol ssh_version_2	Any	Accept	Log

Servers and VMs based on **Orchestrator tags**

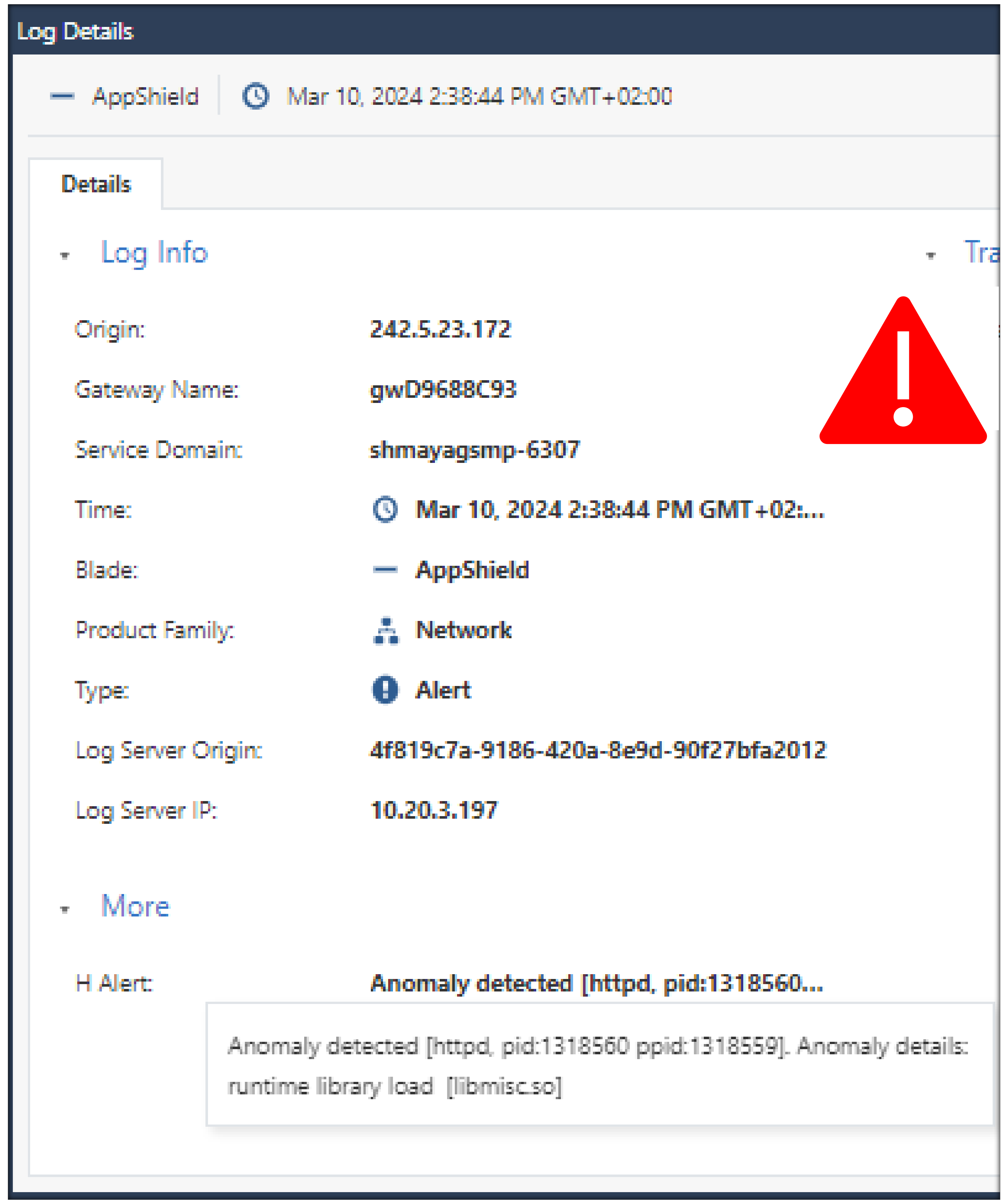
- Industry-leading threat prevention capabilities integrated into AI cloud infrastructure
- Prevent Data Poisoning, Model Inversion and Model Theft attacks targeting AI systems

Comprehensive Network and Host Level Security

Tailored Security at the network and application level designed to protect AI cloud Infrastructure.



```
00000000  3C 21 44 4F  43 54 59 50  20 68 74 74  6D 6C 20 50  <!DOCTYPE html P
00000010  55 42 4C 49  43 20 22 2D  74 65 6E 74  43 2F 2F 44  UBLIC "-//W3C//D
00000020  54 44 20 58  48 54 4D 4C  6E 74 3D 22  20 54 72 61  TD XHTML 1.0 Tra
00000030  6E 73 69 74  69 6F 6E 61  68 61 72 73  4E 22 20 22  nsitional//EN" "
00000040  68 74 74 70  3A 2F 2F 77  09 09 09 3C  33 2E 6F 72  http://www.w3.or
00000050  67 2F 54 52  2F 78 68 74  65 79 77 6F  44 54 44 2F  g/TR/xhtml1/DTD/
00000060  78 68 74 6D  6C 31 2D 74  3D 22 4D 61  69 74 69 6F  xhtml1-transitio
00000070  6E 61 6C 2E  64 74 64 22  20 42 43 2C  74 6D 6C 20  nal.dtd">.<html
00000080  78 6D 6C 6E  73 3D 22 68  31 32 20 69  2F 2F 77 77  xmlns="http://ww
00000090  77 2E 77 33  2E 6F 72 67  39 32 39 2C  39 2F 78 68  w.w3.org/1999/xh
000000A0  74 6D 6C 22  20 78 6D 6C  38 36 2C 31  67 3D 22 65  tml" xml:lang="e
```



Mar 10, 2024 2:38:44 PM GMT+02:00

Log Info

Origin:	242.5.23.172
Gateway Name:	gwD9688C93
Service Domain:	shmayagsmp-6307
Time:	Mar 10, 2024 2:38:44 PM GMT+02:...
Blade:	AppShield
Product Family:	Network
Type:	Alert
Log Server Origin:	4f819c7a-9186-420a-8e9d-90f27bfa2012
Log Server IP:	10.20.3.197

More

H Alert:	Anomaly detected [httpd, pid:1318560...

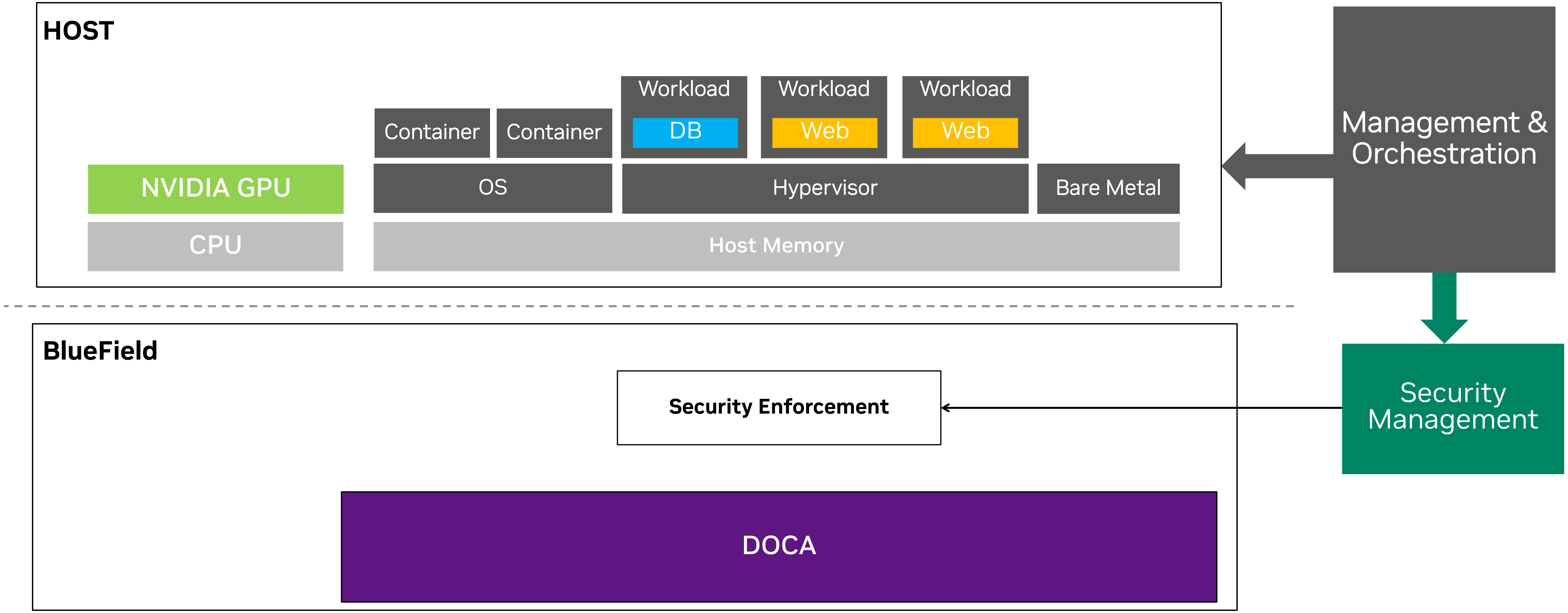
Anomaly detected [httpd, pid:1318560 ppid:1318559]. Anomaly details: runtime library load [libmisc.so]

BlueField-3

Unique host-level process and application protection, preventing local code execution and privilege escalation attacks

Out-of-the-box Security with Effortless Deployment

Plug and Play Deployment using Standard Orchestration Tools at Scale
Automatic Integration across Cloud and on-premises deployments

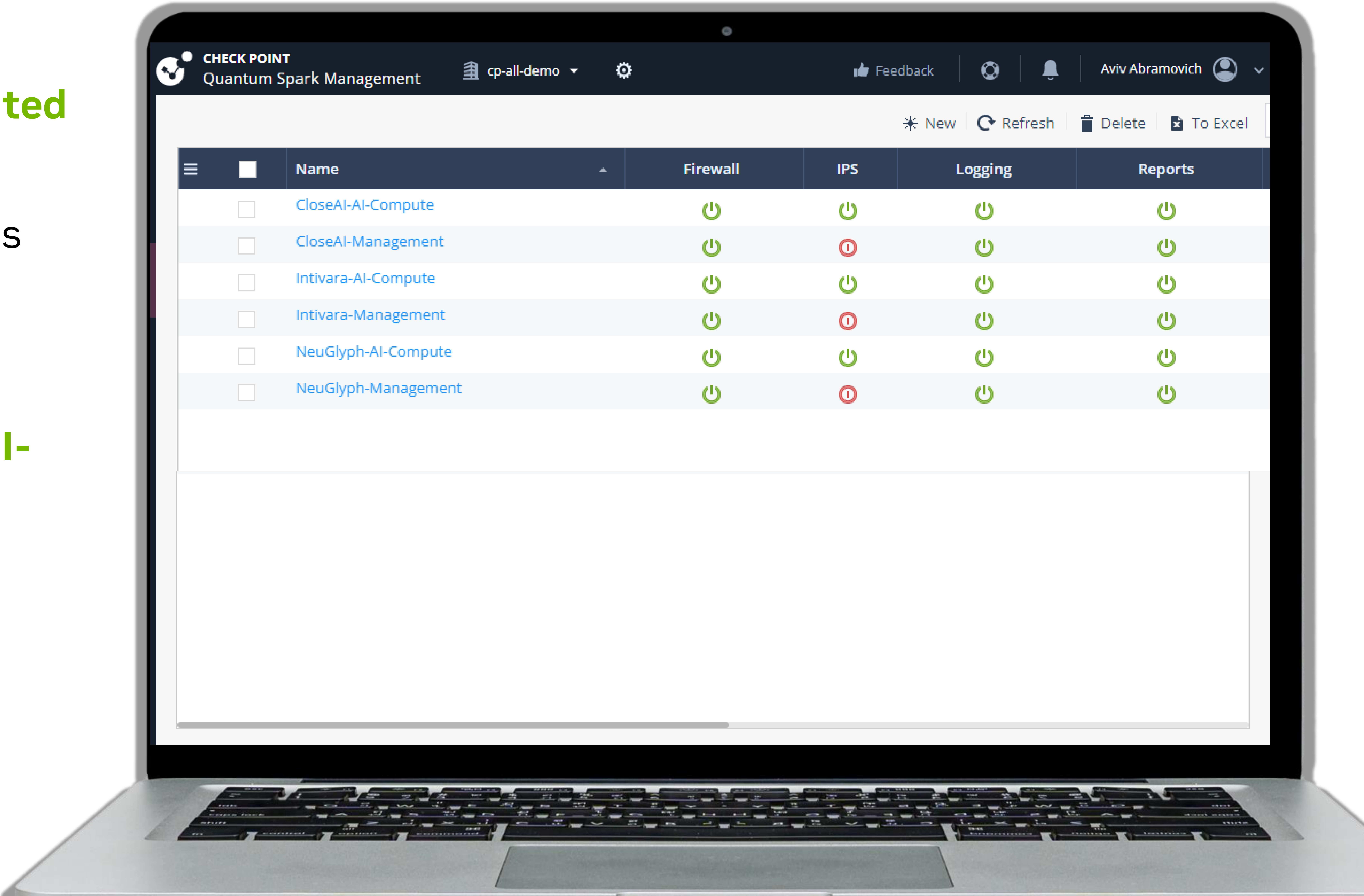


Out-of-the-box Security with Effortless Deployment

Plug and Play deployment using standard orchestration tools at high scale

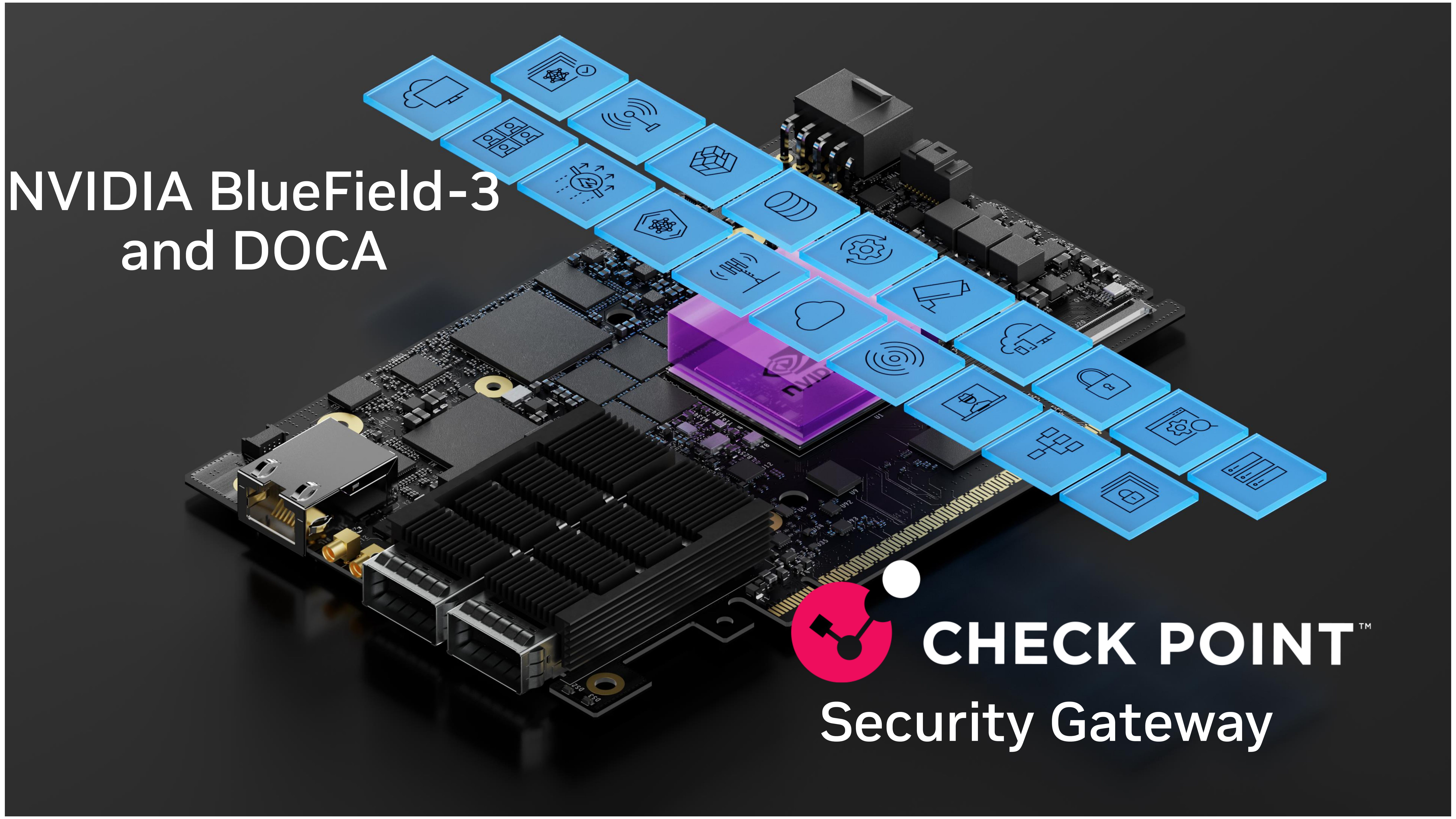
- **Multi-tenanted**, **unlimited scale**, security management with pre-defined security policies designed for AI.

- Supporting **agile** AI environments using **API-driven** security management and monitoring



☐	Name	Firewall	IPS	Logging	Reports
☐	CloseAI-AI-Compute	🟢	🟢	🟢	🟢
☐	CloseAI-Management	🟢	🔴	🟢	🟢
☐	Intivara-AI-Compute	🟢	🟢	🟢	🟢
☐	Intivara-Management	🟢	🔴	🟢	🟢
☐	NeuGlyph-AI-Compute	🟢	🟢	🟢	🟢
☐	NeuGlyph-Management	🟢	🔴	🟢	🟢

Zero Impact on AI Performance



Take Your AI Security to the Next Level

Check Point and NVIDIA Bring Together an Innovative Approach for Securing AI Clouds

Comprehensive Network and Host Level Security

Out-of-the-box Security with Effortless Deployment

Zero Impact for AI Performance

NVIDIA®