

HIGH PERFORMANCE COMPUTING

HPC



CONTENTS

WHAT'S INSIDE?

Boston: Powering innovation for over 30 years	05	Liquid Cooling	40
HPC cluster architecture and workflow	06	Types of Liquid Cooling	40
Boston HPC solution stack selector	07	The GlacierCore Series	44
Compute Technology	08	Castrol	46
Supermicro Compute Solutions for HPC	09	Submer	48
AMD EPYC™ 9005 Series processors	10	Supermicro	49
AmpereOne	11	Zutacore	50
The Intel® Xeon® 6 processor family	12	Networking Solutions	51
Micron MRDIMM Memory	13	NVIDIA Ethernet Spectrum®-4	52
NVIDIA® Grace	14	NVIDIA InfiniBand Quantum™-2	53
Accelerated Solutions	15	NVIDIA Spectrum™-X	54
AMD Instinct™ MI300A series platform	16	Software	55
AMD Instinct™ MI355X series accelerator	17	DKube	56
Intel® Gaudi® 3 AI Accelerator	18	NVIDIA Base Command Manager™	58
NVIDIA® GB300 Grace Blackwell Superchip	19	NVIDIA Mission Control™	60
NVIDIA® RTX Pro 6000 Blackwell	20	OpenHPC	62
Rack-Scale Solutions	21	TrinityX	63
NVIDIA® GB300 NVL72 SuperPOD	22	Boston Services	64
NVIDIA® DGX SuperPOD with DGX B200	23	Boston Technology Consulting	65
An Introduction to AI Factories	24	Boston Training Academy	66
AI Factory Digital Twins with NVIDIA	27	Boston Labs	67
HPC Workloads & NVIDIA Optimised Toolkits	28	Professional Services	68
Architecture Biomedical Science	29	Support and Onsite Warranty	70
Storage Solutions	30		
Boston Igloo AI+	32		
DDN Exascalr	34		
The VAST data platform	36		
WEKA IO	38		



WHAT SETS US APART?



SOLUTION CO-DESIGN & CUSTOMISATION

Complex, multi-vendor and end-to-end solutions built to meet specific customer needs and workloads



EDGE TO EXASCALE

From edge computing to the most demanding exascale environments — solutions for any scale



INDEPENDENT & VENDOR-NEUTRAL GUIDANCE

Our agnostic approach ensures the best-fit solution for every customer, appraising options from industry respected vendors



LABS, TESTING & PROOF-OF-CONCEPT

In-house labs deliver rapid verification, certification, benchmarking and value demonstration — accelerating decision making and reducing risk



ELITE PARTNERSHIPS & INDUSTRY RECOGNITION

Accredited and recognised partnerships with industry leaders like NVIDIA (Elite Partner, SuperPOD) — giving you early access to innovative, ground-breaking products and technologies



AI & HPC CONSULTANCY + TRAINING

Strategic guidance paired with practical training (incl. NVIDIA DLI), helping customers adopt and scale to match their complex computing challenges



GLOBAL REACH, LOCAL SUPPORT

Our international presence ensures responsive support — wherever your operations are



SUSTAINABILITY THROUGH INNOVATION

We integrate energy-efficient, future-facing technology into every solution, supporting greener IT and reducing TCO while simultaneously improving performance and density



FULL-SERVICE SUPPORT: DESIGN, DEPLOYMENT, OPERATION

Professional services that span through design, build, testing, integration, ongoing managed services and support

BOSTON: POWERING INNOVATION FOR OVER 30 YEARS

WHO ARE WE?

Boston Limited has spent over three decades redefining what's possible in high-performance computing. From our early days pioneering bespoke server solutions to becoming a global leader in AI, cloud and advanced computing, our journey has been shaped by a relentless drive to push technology forward.

We design tailor-made HPC clusters, delivering end-to-end solutions for complex HPC challenges, including computational fluid dynamics (CFD), molecular dynamics, molecular biology and particle physics amongst many others.

At Boston, we don't just build systems—we engineer solutions tailored to the demands of the most complex and performance-hungry workloads. Whether it's AI and machine learning, high-frequency trading, cloud infrastructure or large-scale storage, our platforms are designed for efficiency, reliability and uncompromising performance.

But what truly sets us apart is our approach; every solution we create is backed by expert integration, rigorous testing and deep technical insight. We offer fully branded, custom-built platforms, global lifecycle management and hands-on support that extends far beyond deployment. From training and advisory services to full-scale implementation, we help organisations harness the power of HPC, automation and data to drive real results.

With demand for high-performance technology accelerating, Boston continues to expand—bringing our expertise to more industries, more markets and more innovators who refuse to settle for off-the-shelf solutions.

If performance, precision and partnership matter to you, Boston is ready to power your success.

HPC CLUSTER ARCHITECTURE AND WORKFLOW

HPC clusters are enabling breakthroughs in AI, scientific research, financial modelling and beyond. But building a truly optimised cluster isn't just about stacking servers—it's about engineering an ecosystem where compute, networking, storage and software work seamlessly together.

At the heart of every high-performance cluster is a carefully balanced architecture. CPUs handle serial workloads and orchestrate processes, while GPUs accelerate parallel computations, delivering the raw power needed for AI training, simulations and complex modelling. High-speed networking ensures that data moves swiftly between nodes, preventing bottlenecks and enabling real-time processing. Storage solutions—whether high-capacity disk arrays or lightning-fast NVMe drives—are selected to match the needs of data-intensive applications. And on top of it all, the right software stack ensures that every component operates in perfect unison, from job scheduling to monitoring and optimisation.

Designing and deploying such an ecosystem requires expertise. Poorly planned clusters can lead to inefficiencies, underutilised resources and spiralling costs. IT teams often struggle with incompatible software, networking misconfigurations and systems that fail to scale effectively. That's where Boston comes in.

f Boston understands the end-to-end process of designing, implementing and managing clusters—ensuring your project is delivered on time and within budget.

With decades of experience, Boston takes the complexity out of cluster deployment. Our team pre-engineers, tests and optimises complete HPC solutions in our state-of-the-art Boston Labs, ensuring seamless integration of hardware and software. Whether leveraging open-source or commercial cluster management tools, we tailor every system to match our customers' environments, workflows and performance goals.

The result? Clusters that are not just powerful, but also easy to deploy, manage and scale. With Boston, you don't just get a collection of servers—you get a finely tuned, high-performance computing ecosystem, ready to drive innovation from day one.



BOSTON HPC SOLUTION STACK SELECTOR

BOSTON SERVICES	Solution Design & Consultancy	Installation	Warranty & Support	Maintenance & Managed Services
SOFTWARE	OpenFOAM	STAR-CCM+	ANSYS Fluent	LS-DYNA
CLUSTER MANAGEMENT	ClusterVision TrinityX	NVIDIA Mission Control	NVIDIA Base Command	OpenHPC
JOB SCHEDULERS	Slurm	Altair Grid Engine	Altair PBS Professional	IBM Spectrum LSF
CONTAINERS	Apptainer	Kubernetes	OpenShift	Rancher
COMPUTE	AMD 5 th Gen EPYC™	Intel® Xeon® 6	NVIDIA Grace	AmpereOne
ACCELERATED SOLUTIONS	AMD Instinct™ MI300 Series	Intel® Gaudi® 3	NVIDIA Blackwell Architecture	NVIDIA Ada Lovelace Architecture
NETWORKING	NVIDIA Spectrum-X	NVIDIA Quantum-2 InfiniBand	NVIDIA Spectrum-4 Ethernet	SONiC Ethernet
STORAGE	Boston Igloo AI+	VAST	WEKA	DDN
COOLING	Direct Liquid Cooling	Liquid Immersion	Rear Door Heat Exchanger (RDHx)	Air Cooling
DATA CENTRE SERVICES	Colocation	Self-Hosted	Managed Services	Modular

COMPUTE TECHNOLOGY



Supermicro's compute solutions deliver high-density, multi-node server systems designed for scalability, reliability and energy efficiency. From rackmount to blade servers, they provide exceptional performance for HPC, data centres and cloud computing.



AMD's compute solutions excel in HPC, AI and enterprise workloads. EPYC processors power advanced HPC clusters, while AMD's high-density, multi-node solutions offer scalability and efficiency for demanding applications.



Ampere empowers enterprise computing with their high-performance AmpereOne processors, offering exceptional core counts, clock speeds and memory capacity. Built on ARM architecture, Ampere's solutions deliver power-efficient, scalable performance tailored for cloud-native and high-performance workloads.



The Intel® Xeon® 6 processors deliver high performance and energy efficiency, making them ideal for HPC, cloud-native applications and high-density deployments. Designed for demanding workloads, they offer exceptional compute power for AI, scientific research and enterprise applications.



Micron's MRDIMM memory doubles data rates with on-module multiplexer technology, delivering up to 8800 MT/s to overcome traditional memory limits. Designed for AI/ML, in-memory databases and HPC, it boosts efficiency, accelerates performance and scales for future data centre needs.



The NVIDIA Grace CPU is purpose-built for AI supercomputing, HPC and large-scale data processing, delivering high memory bandwidth and energy efficiency for demanding applications. Its advanced memory integration with LPDDR5X makes it ideal for complex simulations, AI model training and real-time analytics.

SUPERMICRO COMPUTE SOLUTIONS FOR HPC

Supermicro and Boston have been delivering HPC solutions for over three decades, powering some of the world's most demanding workloads. Supermicro's high-density, multi-node platforms are engineered for maximum efficiency; a perfect fit for HPC, AI and scientific research. Their liquid cooling solutions further enhance performance by reducing energy consumption and improving thermal management—critical for sustainable, large-scale computing.

Their reference designs for enterprise HPC, AI and deep learning provide a strong foundation for customised solutions, allowing organisations to fine-tune their infrastructure to the minutest of detail. Boston trusts Supermicro as a primary building block for our HPC solutions, ensuring customers benefit from optimised performance, energy efficiency and seamless scalability.

Supermicro systems have been deployed in HPC clusters worldwide, including CERN, where they contribute to groundbreaking scientific discoveries. A significant portion of data centres globally also leverage Supermicro's technology due to its reliability, density and competitive performance-to-price ratio.

From computational fluid dynamics (CFD) and finite element analysis (FEA) to weather prediction, oil and gas research, automotive, aerospace and naval engineering, Supermicro's solutions are trusted across industries to accelerate workloads while minimising operational costs.

HPC requires immense computing power, but efficiency is key. Supermicro champions greener computing with power-efficient configurations and liquid cooling technology, helping businesses reduce total cost of ownership while advancing their sustainability goals.





AMD EPYC™ 9005 SERIES PROCESSORS

AMD has taken you on a journey through Naples, Rome, Milan and Genoa. Now, the next stop is Turin-codename for AMD EPYC™ 9005 Series processors. The EPYC family has consistently delivered exceptional performance, efficiency and security, and the EPYC 9005 Series processors continues this legacy. With architectural advancements and cutting-edge memory and I/O capabilities, it redefines compute power for modern workloads.

AMD EPYC 9005 Series processors features up to 192 “Zen 5” cores, delivering industry-leading core density. With DDR5 memory support and 12 channels, it provides immense bandwidth and capacity for data-intensive applications. PCIe® Gen 5 ensures high-speed connectivity, making it the foundation of next-generation enterprise computing. All this translates to lower total cost of ownership (TCO) and improved energy efficiency, making it a smart investment for data centres.

KEY FEATURES:

BENEFITS:

- **Extreme Core Density:** Up to 192 “Zen 5” cores deliver exceptional socket-level performance for compute-heavy workloads and parallel processing
- **High-Capacity, High-Speed Memory:** Supports up to 6TB of DDR5-6400 across 12 memory channels, enabling rapid data access and large in-memory computing
- **Expansive I/O Capabilities:** Offers 128 PCIe® Gen 5 lanes (up to 160 in dual-socket systems) for high-throughput connectivity to GPUs, storage and networking devices
- **CXL 2.0 Support for Greater Efficiency:** Enables expanded memory capacity and improved efficiency—ideal for large-scale simulations and memory-intensive tasks
- **Energy-Efficient Scalability:** Delivers top-tier performance per watt while reducing total cost of ownership, supporting sustainable data centre operations
- **Accelerated Insights:** Optimised architecture drives faster results and improved time-to-insight, powering innovation across AI, HPC and analytics workloads



AMPEREONE

The AmpereOne™ family delivers high-efficiency compute with a scalable architecture optimised for cloud, HPC and edge workloads. Designed for energy efficiency and performance per watt, AmpereOne™ processors provide a balanced memory and compute structure, ensuring seamless execution of demanding applications.

With high core counts, PCIe® Gen 5 connectivity and efficient power management, AmpereOne™ processors are ideal for Telco Edge, Enterprise and HPC environments, offering a sustainable solution without compromising performance. Engineered for AI compute, AmpereOne delivers maximum flexibility-removing the need for rigid, single-purpose GPU infrastructure.

KEY FEATURES:

BENEFITS:

- **Workload Versatility:** Optimised for enterprise, HPC and telco edge workloads, offering scalable performance across diverse deployment scenarios
- **High Core Count for Parallelism:** Up to 192 cores for powerful multi-threaded processing, accelerating compute-intensive tasks
- **Fast Clock Speeds:** Operates at up to 3.7 GHz, enabling responsive execution and reduced latency for demanding applications
- **Configurable Power Envelope:** Adjustable TDP from 250W to 400W enables a balance between energy efficiency and performance, tailored to specific workload demands
- **Next-Gen Connectivity:** PCIe® Gen 5 support ensures high-bandwidth connections for modern accelerators and storage, maximising system throughput
- **Energy-Efficient Arm Architecture:** Built on Arm v8.6+ architecture to deliver exceptional performance per watt, ideal for power-sensitive environments





THE INTEL® XEON® 6 PROCESSOR FAMILY

The Intel® Xeon® 6 processor family is designed to power HPC, cloud and enterprise workloads with a scalable and efficient architecture. Featuring Performance Cores (P-Cores) for maximum compute power and Efficient Cores (E-Cores) for energy-optimised scaling, the Xeon 6 family delivers tailored solutions for demanding workloads.

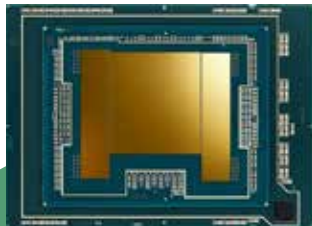
For HPC environments, the Xeon 6 6900P (Advanced Platform - AP) and 6700P (Scalable Platform - SP) processors provide the highest performing cores, I/O capabilities and memory bandwidth to accelerate complex computations. The Xeon 6 6900E and 6700E processors prioritise energy efficiency, making them ideal for high-scale cloud workloads where power consumption is a key factor.

KEY FEATURES:

- **High Core Count:** Offers a large number of Cores per socket, allowing for parallel processing of complex tasks
- **P-Cores (Performance Cores):** Dedicated high-performance Cores for computationally intensive workloads
- **Advanced Matrix and Vector Engines:** Optimised for efficient matrix operations, crucial for AI and scientific calculations

BENEFITS:

- **High Memory Bandwidth:** Supports MRDIMM technology, with speeds up to 8800 MT/s, enabling fast data transfer between the processor and memory.
- **AI Acceleration:** Integrated AI engines to accelerate machine learning and deep learning tasks
- **Scalability:** Designed to be easily scaled to meet the demands of large HPC clusters



MICRON MRDIMM MEMORY

Micron's MRDIMM, or Multiplexer Combined Ranks Dual In-line Memory Module, represents a groundbreaking advancement in server memory technology designed to address the growing performance demands of data centers. At its core, MRDIMM utilizes an innovative on-module multiplexer that enables the two memory ranks to be accessed simultaneously. This clever design allows the module to function like a single, ultra-fast rank, effectively doubling the data rate and providing a significant boost in memory bandwidth. This architecture overcomes the speed limitations inherent in traditional RDIMM and LRDIMM modules, unlocking new levels of performance for modern CPUs.

The impact of this technology is substantial, especially for workloads that are frequently bottlenecked by memory speed. With initial speeds starting at 8800 MT/s and a roadmap extending beyond 12000 MT/s, MRDIMM provides the massive bandwidth required for next-generation AI/ML training, in-memory databases, and high-performance computing (HPC) applications. By ensuring that powerful processor cores are consistently fed with data, Micron's MRDIMM technology is a critical enabler for the future of computing, promising to enhance server efficiency and accelerate insights from complex datasets.

KEY FEATURES:

- **On-module multiplexer technology:** Enables simultaneous access to two memory ranks, effectively doubling the data rate
- **Ultra-high memory bandwidth:** Starts at 8800 MT/s with a roadmap beyond 12000 MT/s, surpassing traditional RDIMM and LRDIMM limitations
- **Optimised for next-gen workloads:** Designed to support AI/ML training, in-memory databases and HPC applications

BENEFITS:

- **Improved server efficiency:** Keeps modern CPUs consistently fed with data, removing memory bottlenecks
- **Accelerated performance:** Supports faster data throughput for demanding tasks, enhancing application responsiveness
- **Future-ready infrastructure:** Delivers the bandwidth scalability needed for evolving data centre requirements



NVIDIA® GRACE

Designed to tackle the growing demands of AI, HPC and large-scale data analytics, NVIDIA Grace delivers exceptional speed, efficiency and scalability. With a purpose-built architecture, Grace unlocks new levels of performance while reducing power consumption-making it ideal for next-generation data centres.

Grace is available as a standalone CPU or as part of a hybrid CPU-GPU solution for AI-driven applications. The Grace Hopper Superchip pairs the Grace CPU with an NVIDIA Hopper GPU, optimised for AI training and inference, while Grace Blackwell integrates a Grace CPU with NVIDIA's Blackwell GPU architecture, delivering high-performance computing for large-scale workloads. Engineered for seamless integration, Grace leverages Arm Neoverse V2 cores and NVIDIA's high-bandwidth interconnect technology, ensuring streamlined data transfer and accelerated processing for even the most complex workloads.

KEY FEATURES:

- **144 Arm Neoverse V2 cores:** Ensures high single-threaded and multi-threaded performance for demanding workloads
- **Up to 1024 GB/s Memory Bandwidth:** Utilises LPDDR5X memory with ECC to provide fast, reliable data access
- **NVLink-C2C Interconnect:** Delivers high-speed, low-latency connectivity between Grace CPUs and NVIDIA GPUs for unified performance

BENEFITS:

- **Optimised AI and HPC Workloads:** Powers complex AI models, data analysis and simulations with industry-leading speed
- **Energy-Efficient Design:** Offers up to 2x the energy efficiency of traditional x86 systems, reducing overall operational costs
- **Scalability for Data Centres:** Supports flexible configurations to accommodate evolving data centre needs

ACCELERATED SOLUTIONS



AMD's MI300A is a versatile APU that integrates Zen 4 CPU Cores and CDNA3 GPU architecture on a single package, designed for HPC and AI workloads with a focus on energy efficiency and reduced latency.



The AMD Instinct™ MI355X is built on the CDNA™ 4 architecture and features 288GB of HBM3E memory, making it ideal for large-scale AI training and inference. It offers high memory bandwidth and FP8 precision support, optimised for LLM and transformer workloads.



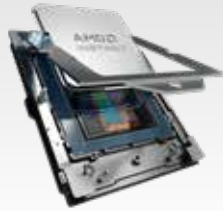
The Intel® Gaudi® 3 AI Accelerator is engineered to deliver high-performance, cost-effective scalability for complex AI deployments, supporting a wide range of industry AI models and frameworks.



The NVIDIA® GB300 Grace Blackwell Superchip combines a Grace CPU with a Blackwell GPU over high-speed NVLink-C2C, delivering fast memory access and powerful performance for AI, HPC and trillion-parameter model workloads.



The NVIDIA RTX PRO 6000 Blackwell delivers extreme performance for professional AI, rendering and simulation workloads, making it ideal for generative AI, digital twin modelling and large-scale visualisation.



AMD INSTINCT™ MI300A SERIES PLATFORM

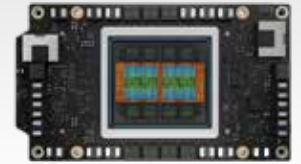
The AMD Instinct™ MI300A is a powerful accelerated processing unit (APU) designed to handle demanding server and supercomputer workloads, including cutting-edge AI, machine learning and large-scale simulations.

Built with advanced CDNA3 architecture and Zen 4 CPU Cores, and leveraging 5nm process technology, the MI300A provides exceptional performance for a variety of intensive computing tasks. Equipped with 128 GB of HBM3 memory and Infinity Fabric technology, the MI300A ensures reduced latency and faster data access for memory-intensive applications. This makes it a versatile solution for AI inferencing, HPC and hybrid workloads where both general-purpose and accelerated computing are essential.

KEY FEATURES:

BENEFITS:

- **High Memory Capacity and Bandwidth:** Features 128 GB of HBM3 memory and Infinity Fabric™ Link for ultra-low latency and rapid data access, accelerating insight and decision-making
- **Optimised for Mixed Workloads:** Integrated Zen 4 CPUs and CDNA™ 3 GPUs enable efficient execution of AI inferencing, scientific simulations and complex data workflows
- **Enhanced Built-in Security:** Advanced security capabilities protect AI models and mission-critical data, offering peace of mind for regulated and high-risk environments
- **Hybrid Architecture for Versatility:** Combines Zen 4 CPU cores and CDNA™ 3 GPU architecture to support a wide range of workloads, from AI and ML to traditional HPC, offering deployment flexibility
- **Robust Security for Sensitive Workloads:** Includes Secure Memory Encryption and AMD Infinity Guard to protect IP and critical data—ideal for security-conscious sectors
- **Energy-Efficient Performance:** Optimised for performance per watt, reducing power consumption and operational costs while supporting greener data centre initiatives



AMD INSTINCT™ MI355X ACCELERATOR

Powering the next era of AI and HPC performance

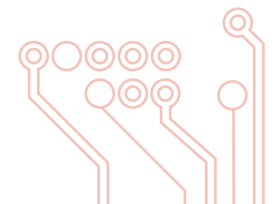
The AMD Instinct™ MI355X GPU represents the latest leap forward in AMD's accelerator lineup—built to support the compute demands of modern AI training, inference and high-performance computing workloads. Based on the advanced CDNA™ 4 architecture and manufactured using an industry-leading 3 nm process, the AMD Instinct™ MI355X GPU combines 288 GB of HBM3E memory, 8 TB/s of memory bandwidth and up to 20 PFLOPS of FP4/FP6 performance with sparsity.

With a liquid-cooled OAM form factor, support for PCIe® 5.0, and Infinity Fabric™ interconnects, the AMD Instinct™ MI355X GPU is engineered for high-density, rack-scale deployments. It is ideal for powering large language models (LLMs), generative AI and advanced simulations across enterprise and research environments. As an AMD Elite Partner, Boston offers full integration, configuration and benchmarking support—enabling you to deploy the AMD Instinct™ MI355X GPU effectively within your infrastructure.

KEY FEATURES:

BENEFITS:

- **288 GB HBM3E high-bandwidth memory** per GPU, delivering up to 8 TB/s bandwidth
- **Up to 20 PFLOPS** of performance (FP4/FP6 with sparsity), ideal for inference-heavy AI workloads
- **OCP Accelerator Module (OAM) form factor**, supporting 1.4 kW sustained TDP with optional direct liquid cooling
- **Up to seven Infinity Fabric™ links per GPU**, enabling high-speed, low-latency multi-GPU scaling
- **Support for PCIe® 5.0, SR-IOV GPU virtualisation**, ECC and secure boot
- **Tackle massive models without memory bottlenecks** – with 288 GB of HBM3E and ultra-fast access
- **Scale efficiently and sustainably** – liquid cooling and 1.4 kW thermal headroom support high-density deployments
- **Secure and robust deployments** – full-chip ECC, secure boot, DICE attestation and ROCm™ software ecosystem





INTEL® GAUDI® 3 AI ACCELERATOR

Building upon the proven performance and efficiency of its predecessors, the Intel® Gaudi® 3 AI Accelerator is engineered to meet the escalating demands of large-scale AI applications. It delivers a significant leap in performance and productivity for AI training and inference on popular LLMs and multimodal models.

The Gaudi® 3 architecture is optimised for deep learning workloads, offering enhanced performance and scalability-ideal for enterprises advancing their AI initiatives. Central to this scalability is an integrated high-speed network fabric that supports efficient communication across accelerators. Built-in Ethernet-based NICs provide remote direct memory access (RDMA) with minimal latency and high bandwidth, enabling rapid data movement without CPU overhead.

KEY FEATURES:

BENEFITS:

- **High-Efficiency Deep Learning:** Provides up to 1,678 TFLOPS of FP8 and BF16 compute performance, ensuring faster AI model training and inference
- **Massive Memory Bandwidth:** Equipped with 128 GB of High-Bandwidth Memory 2e (HBM2e) and a memory bandwidth of 3.7 TB/s to handle large, complex datasets
- **Integrated Networking:** Features 24 integrated 200-Gigabit Ethernet (GbE) ports, delivering an aggregate bandwidth of 4.8 TB/s in each direction for seamless, high-speed data transfer
- **Enhanced AI Training Speed:** The high compute power and memory bandwidth reduce training times for complex AI workloads, improving overall productivity
- **Scalable Architecture:** Supports large-scale deployments with efficient, industry-standard Ethernet networking, eliminating the need for proprietary solutions
- **Flexible Deployment Options:** Available in both custom Open Accelerator Module (OAM) and standard PCIe form factors, with power configurations of 900W and 600W, respectively



NVIDIA® GB300 GRACE BLACKWELL SUPERCHIP

The GB300 is NVIDIA's next-generation Grace Blackwell Superchip, featuring the new B300 Blackwell Ultra GPU paired with an Arm-based Grace CPU. Designed for ultra-dense AI reasoning and HPC workloads, each chip delivers up to 20 petaflops of FP4 AI performance in desktop, edge or rack configurations.

The GB300 Blackwell Ultra platform is ideal for applications including Agentic AI, which uses sophisticated reasoning and iterative planning to autonomously solve complex, multistep problems. AI agent systems go beyond instruction-following. They can reason, plan and take actions to achieve specific goals. Physical AI, enabling companies to generate synthetic, photorealistic videos in real time for the training of applications such as robots and autonomous vehicles at scale.

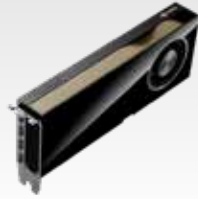
KEY FEATURES:

BENEFITS:

- Enhanced FP4 throughput—**1.5x** faster than GB200 in inference workloads
- Massive **288 GB HBM3e** memory (12-stack, up from 192 GB on GB200)
- Dual-reticle Blackwell Ultra die on TSMC 4NP process with **208 billion transistors**
- **Maximised model capacity:** 1.50x the HBM capacity vs GB200 enables larger models
- **Higher energy efficiency:** Stronger FP4 performance per watt helps cut inference cost
- **Simplified deployment:** Single-chip or rack-scale config, ready for AI factories and next-gen AI labs

USE CASES:

- **Trillion-parameter LLM training and fine-tuning**, including GPT and open-source generative models
- **Real-time AI inference** for search, recommendation systems, and large-scale digital assistants
- **AI-driven scientific computing and simulation**, including molecular dynamics and climate modelling
- **Digital twin and industrial metaverse applications**, requiring high-throughput, low-latency compute



NVIDIA® RTX PRO 6000 BLACKWELL

The NVIDIA RTX PRO 6000 Blackwell is NVIDIA's most powerful workstation and server GPU to date, delivering exceptional compute performance and memory capacity for professional AI, rendering and simulation workloads. Built on the Blackwell architecture, it's designed to accelerate generative AI, high-fidelity visualisation and real-time data analysis in both workstation and data centre environments.

Available in active, blower and passive-cooled variants, the RTX PRO 6000 Blackwell supports up to 96 GB of ECC GDDR7 memory and 24,064 CUDA cores. With advanced Tensor and RT cores, plus support for MIG and PCIe Gen5, this GPU is ideal for multi-user environments, high-performance computing clusters and enterprise creative workflows.

KEY FEATURES:

- 96 GB GDDR7 ECC memory with 1.8+ TB/s bandwidth
- 24,064 CUDA cores and 752 next-gen Tensor Cores
- Available in workstation and server configurations (passive/active cooling)

BENEFITS:

- Supports large-scale AI model training and rendering without memory bottlenecks
- Scalable for data centre deployment with up to 8 GPUs per server
- Built-in support for secure multi-instance GPU (MIG) compute and virtualisation



RACK-SCALE SOLUTIONS



The NVIDIA GB300 NVL72 SuperPOD is a liquid-cooled, rack-scale AI platform built on 72 Blackwell Ultra GPUs and Grace Blackwell CPUs. It delivers exceptional density and throughput, providing top-tier performance for large-scale training and inference in a compact data centre-ready system.



The NVIDIA B200 GPU is a high-efficiency solution for edge and embedded AI. With Blackwell architecture, large memory, fast interconnects and low power use, it delivers strong performance where space and energy are limited.



NVIDIA AI Factories unite accelerated computing, networking and software to create purpose-built AI infrastructure. They enable organisations to train models and deploy intelligent services at scale, turning data into actionable outcomes.



NVIDIA Omniverse™ plays a foundational role in the AI Factory Blueprint, enabling real-time simulation, industrial digital twins and synthetic data generation. Integrated with Omniverse Nucleus, it supports collaborative pipelines and predictive modelling across robotics, autonomous systems and complex industrial environments.



NVIDIA accelerates HPC with integrated platforms and optimised toolkits. Designed for complex scientific and engineering tasks, they shorten time to results and scale efficiently across data-intensive environments.



Boston and NVIDIA provide a validated reference architecture for biomedical AI. Combining high-performance systems with NVIDIA AI Enterprise software, it supports breakthroughs in genomics, drug discovery and clinical research.



NVIDIA® GB300 NVL72 SUPERPOD

The NVIDIA GB300 NVL72 SuperPOD is NVIDIA's most powerful rack-scale AI system to date — purpose-built for trillion-parameter LLMs, accelerated HPC and AI factory-scale deployments. Powered by 36 GB300 Grace Blackwell Superchips (page 19) — each combining one Blackwell GPU and one Grace CPU — the system forms a unified 72-GPU NVLink domain with groundbreaking speed and scalability.

Built to deliver real-time inference at exascale, the GB300 NVL72 offers up to **1.4 exaFLOPS of FP4 AI performance and 130 TB/s NVLink bandwidth** within a single rack. The system is fully liquid-cooled, energy-efficient and ready to deploy with Boston's reference architectures and storage integrations.

Boston provides fully integrated GB300 NVL72 solutions in collaboration with industry leaders like VAST Data, WEKA and DDN — ensuring optimised performance, scalability and ease of deployment.

KEY FEATURES:

- **Unified Blackwell Architecture:** Combines Grace CPUs and Blackwell GPUs in each Superchip for full-stack acceleration across AI and HPC workloads
- **Unrivalled NVLink Bandwidth:** 5th Gen NVLink Switch System delivers 130 TB/s interconnect bandwidth across 72 GPUs
- **AI-Ready Compute:** Achieves up to 1.4 exaFLOPS of FP4 performance in a single rack for cutting-edge generative AI and LLM applications

BENEFITS:

- **Scalable AI Infrastructure:** Designed to scale horizontally across racks for hyperscale AI factories and HPC deployments
- **Turnkey Reference Design:** Boston delivers GB300 NVL72 as a pre-integrated solution, with liquid cooling and certified storage partners
- **Built for Next-Gen Workloads:** Supports NVIDIA AI Enterprise, NVIDIA NIMs and the full NVIDIA software stack out of the box



NVIDIA® DGX SUPERPOD WITH DGX B200 SYSTEMS

The NVIDIA DGX SuperPOD, now featuring DGX B200 systems, offers a scalable AI infrastructure designed to meet the demands of large-scale AI and high-performance computing workloads. Each DGX B200 system is equipped with eight NVIDIA Blackwell GPUs, delivering exceptional performance and efficiency.

Built for maximum performance, the DGX B200 delivers up to 72 petaFLOPS for training and 144 petaFLOPS for inference, supported by 1,440GB of GPU memory. With 5th Gen NVLink providing 14.4 TB/s bandwidth per system and PCIe Gen6 at 256 GB/s, the DGX SuperPOD ensures rapid data transfer, optimising AI model training and inference.

KEY FEATURES:

- **High-Performance Computing:** DGX B200 systems deliver unparalleled computational power, enabling rapid training and inference for complex AI models
- **Scalable Architecture:** The SuperPOD design allows for seamless expansion, accommodating growing AI workloads and datasets
- **Integrated Networking:** High-speed interconnects ensure low-latency communication between nodes, optimising performance for distributed AI applications

BENEFITS:

- **Accelerated AI Development:** With the combined power of DGX B200 systems and the SuperPOD infrastructure, organisations can significantly reduce time-to-insight for AI projects
- **Simplified Deployment:** The turnkey nature of the DGX SuperPOD streamlines the implementation of AI infrastructure, allowing enterprises to focus on innovation rather than integration
- **Future-Proof Scalability:** Designed with growth in mind, the SuperPOD enables organisations to effortlessly scale their AI capabilities in response to evolving business needs





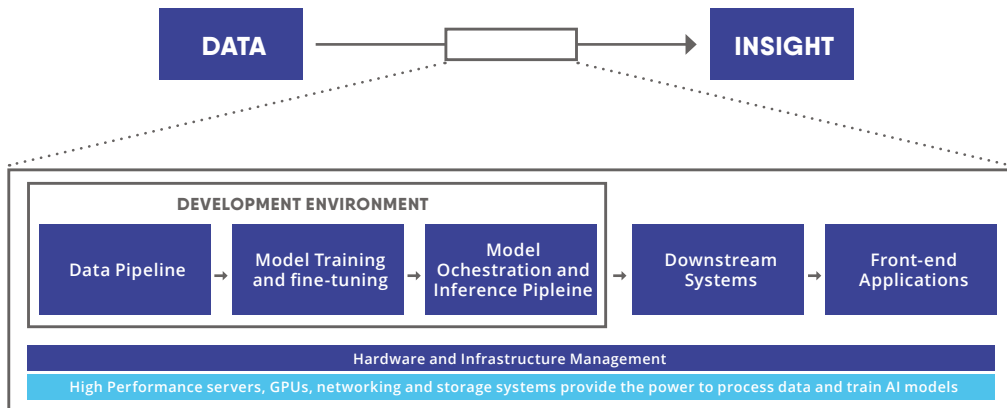
AN INTRODUCTION TO AI FACTORIES

An AI Factory is an integrated computing environment engineered to operationalise the development, training and deployment of artificial intelligence at scale. It consolidates high-performance computing resources, low-latency networking, accelerated storage systems and optimised software frameworks into a unified architecture.

Within this environment, data is systematically ingested, processed and prepared for model training. Models are then iteratively developed, validated and deployed, creating a closed-loop workflow that is repeatable, measurable and optimised for throughput. By standardising and industrialising these processes, AI Factories transform AI from ad-hoc experimentation into a sustained, high-efficiency capability.

This approach enables organisations to execute multiple AI initiatives in parallel, reduce development cycles and maintain consistency and quality across AI-driven operations.

To understand the outputs of this process, consider the analogy of how a brick factory transforms raw clay into uniform bricks, an AI model transforms raw data into its fundamental unit of production: the token. Each token is a discrete digital building block. The model's internal machinery learns patterns and relationships within massive datasets, and assembles tokens one by one into coherent outputs, whether that be an essay, a line of code or a spoken sentence. In this way, the AI factory manufactures intelligence, piece by piece, from data.



Boston Limited designs and delivers end-to-end, application-defined AI Factories. Our accredited architects and specialists manage the entire project lifecycle, from requirements analysis, design and system configuration to installation, software integration, optimisation and long-term maintenance. Every solution is engineered around specific workload demands, ensuring optimal performance and scalability.

At the core of many AI Factories is the NVIDIA SuperPOD, a fully integrated high-performance computing platform capable of supporting the most demanding AI and HPC workloads.

As an established NVIDIA SuperPOD specialist, Boston applies deep expertise to design and implement SuperPOD based solutions tailored to business and research objectives. We ensure that each deployment is aligned to the application requirements, network architecture and data workflows of the client's environment.



SuperPOD Specialised Partners in UK



Access to the latest NVIDIA advanced technology



Expert guidance and tailored solutions



AI FACTORIES CONTINUED

THE BOSTON VALUE PROPOSITION:

- **Workload-specific architecture:** Solutions are defined by the applications and datasets they will run, ensuring measurable gains in performance and efficiency.
- **Comprehensive delivery:** From strategic consultancy and system design through to installation, software stack integration and ongoing optimisation, Boston manages every stage of the deployment.
- **Specialist expertise:** A team with deep knowledge in HPC, AI, networking and storage, capable of integrating complex technologies into a single, operationally robust system.
- **Vendor-neutral approach:** Leveraging a broad partner ecosystem to ensure the best-fit technologies are selected for each project, without compromise.
- **Lifecycle partnership:** Long-term support, maintenance and upgrade paths to keep systems aligned with evolving workloads and emerging technologies.
- **Proven track record:** Successful delivery of high-value HPC and AI projects across industries including research, finance, manufacturing and healthcare.

Boston enables organisations to move from concept to scale with confidence, efficiency and technical precision. Contact us to learn how NVIDIA SuperPOD solutions can transform your business.

E: sales@boston.co.uk
T: +44 (0) 1727 876 100
W: www.boston.co.uk



AI FACTORY DIGITAL TWINS WITH NVIDIA OMNIVERSE™

Building an “AI Factory” demands a sophisticated, integrated approach to design and operations, a challenge NVIDIA addresses with its Omniverse Blueprint. This innovative framework leverages OpenUSD libraries to aggregate vast amounts of 3D data from diverse sources, including the building’s structure, NVIDIA’s powerful accelerated computing systems (like DGX SuperPODs and GB300 NVL72) and critical infrastructure components from partners such as Schneider Electric and Vertiv. By unifying the design and real-time simulation of billions of components, the blueprint empowers engineers to overcome complex hurdles such as component integration, space optimisation and the meticulous fine-tuning of cooling, power distribution and networking topologies.

This simulation-first methodology breaks down traditional engineering silos, enabling seamless, collaborative iteration across disciplines, leading to optimised energy usage, the elimination of potential failure points and the accurate modelling of real-world operational conditions, ultimately accelerating decision-making, and ensuring designs are validated before physical deployment.

BY USING OMNIVERSE IN AN AI FACTORY, ENTERPRISES CAN:

- Create industrial digital twins for predictive simulation and optimisation
- Integrate Omniverse Nucleus into real-time collaboration pipelines across multiple tools
- Support AI development with synthetic data generation and scenario testing for robotics, autonomous systems and more

When paired with Boston’s NVIDIA SuperPOD systems, Omniverse becomes more than a visualisation tool — it becomes a dynamic layer for real-time insight, digital feedback loops and end-to-end AI lifecycle integration.





HPC WORKLOADS & NVIDIA OPTIMISED TOOLKITS

REAL-WORLD HPC APPLICATIONS & NIM TOOLKITS

Boston supports a wide array of HPC and AI workloads across industries — backed by NVIDIA's expanding range of NeMo Inference Microservices (NIMs) and domain-specific SDKs.

Whether you're running simulations, training LLMs or performing edge inference, Boston's infrastructure and software stack accelerates results with optimal performance per watt and node.

POPULAR WORKLOAD AREAS:

- Computational Fluid Dynamics (CFD)
- Molecular Dynamics & Biology
- Climate Modelling
- Digital Pathology & Genomics
- High-Frequency Trading
- Generative AI for Healthcare & Finance

SUPPORTED TOOLKITS & OPTIMISATIONS:

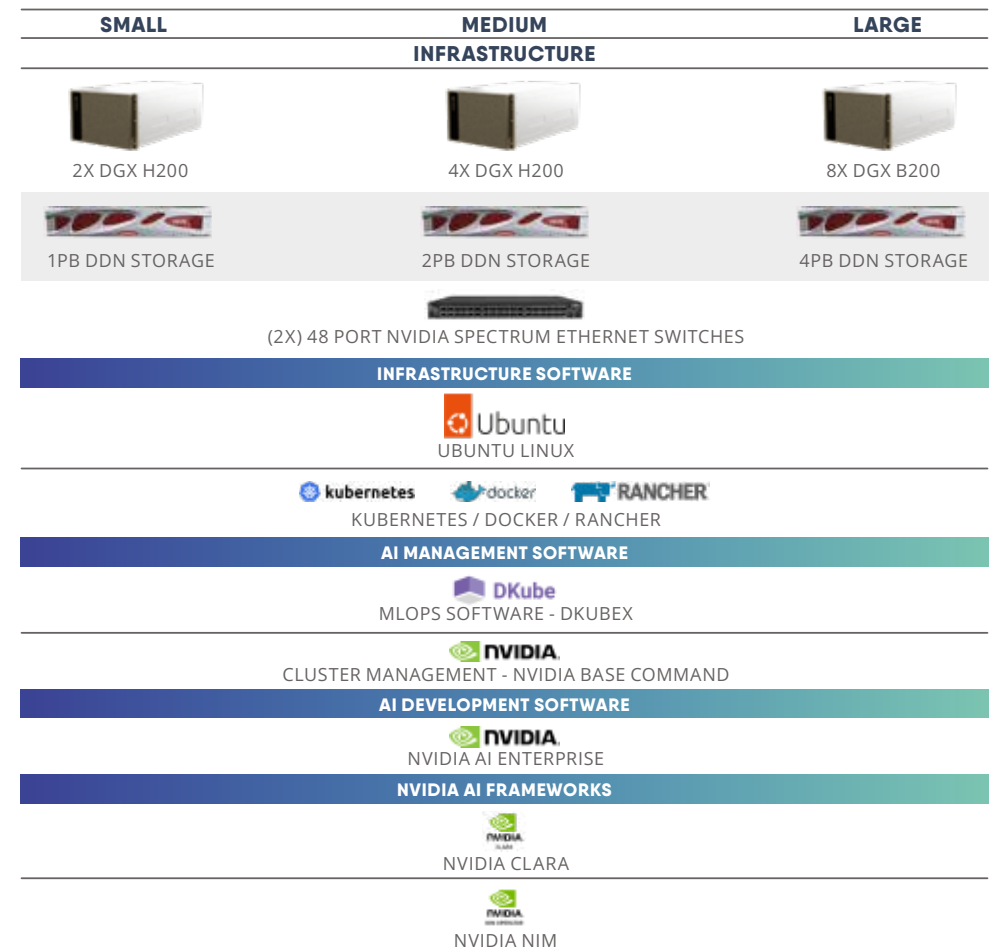
- NVIDIA Modulus (Physics-informed ML)
- NVIDIA Clara™ (Healthcare)
- NVIDIA cuQuantum™ (Quantum simulations)
- NVIDIA NeMo & NIMs (Pre-trained AI model deployment)
- NGC Containers (GPU-optimised apps)

Boston's team works with customers to select and optimise the right toolkits for each application, ensuring maximum throughput, stability and ROI.

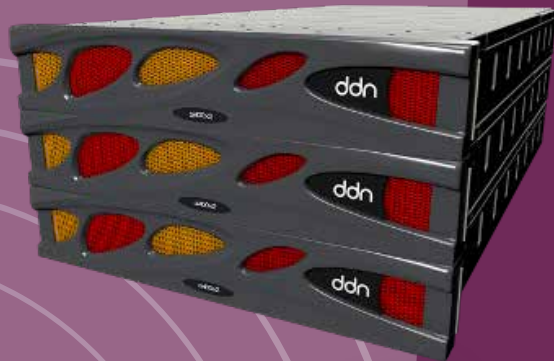
FOCUS ON OPTIMISED ARCHITECTURE FOR NVIDIA CLARA FOR HEALTHCARE & BIOMEDICAL SCIENCE

A MODULAR DATACENTRE SOLUTION FOR SCALABLE AI IN HEALTHCARE & LIFE SCIENCES

Boston Limited, in partnership with NVIDIA, delivers a flexible, scalable AI infrastructure tailored for biomedical applications such as diagnostics, drug design, and precision medicine. This architecture enables both AI training and inference across small to large-scale deployments with full support and resilience built in.



STORAGE SOLUTIONS



Purpose-built for AI workloads, the Boston Igloo AI+ delivers high-performance storage at an accessible price point, making AI projects viable without compromise. With ultra-low latency, scalable architecture and seamless GPU integration, it ensures fast, efficient data access tailored for AI-driven enterprises.



Designed for large-scale AI and HPC environments, DDN Exascaler delivers high-bandwidth, low-latency storage with parallel file system technology. Trusted by leading research institutions and enterprises, it enables rapid AI model training, real-time analytics and seamless data movement across hybrid cloud infrastructures.



A revolutionary unified storage and compute platform, VAST Data eliminates storage silos by integrating high-performance flash storage with database and containerised computing. Ideal for multi-modal AI training and generative AI workloads, it provides limitless scalability with enterprise-grade efficiency.



Engineered for data-heavy AI and ML applications, WEKA®FS™ delivers unprecedented storage throughput and IOPS, ensuring massive-scale AI training, deep learning and GPU-accelerated workloads operate without bottlenecks. With a cloud-native, software-defined architecture, WEKA maximises performance from edge to core to cloud.

BOSTON IGLOO AI+



Investment in accelerator technology is critical to the success of an AI project. The Boston Igloo AI+, powered by PEAK:AIO, delivers exceptional performance to support AI workloads without compromising investment in accelerators. The NFS protocol enables multiple GPU-accelerated hosts or entire HPC clusters to access the storage simultaneously, ensuring seamless scalability as AI compute demands grow.

The Boston Igloo AI+ strikes the perfect balance between cost and high performance, making it an ideal solution for storage requirements in the hundreds of terabytes-ensuring accelerators are continuously fed with data for fast model development.

For smaller environments with a single HPC host, the Boston Igloo AI+ can be connected directly, eliminating the need for a switch. This allows for greater investment in additional compute resources, ensuring long-term scalability and efficiency.

SPECIFICATIONS

MODEL	IGLOO AI+ 1U FORM FACTOR	IGLOO AI+ 2U FORM FACTOR
Drives	6 or 12x NVMe	6, 12 or 24x NVMe
CPU	(2x) Intel Gold 6442Y	(2x) Intel Gold 6442Y
Memory	256GB	256GB
Network Adapter	NVIDIA Connect-X 7 Dual Port	NVIDIA Connect-X 7 Dual Port
Installation	Local or remote	Local or remote
Capacity	Up to 368.6TB RAW	Up to 737.3TB RAW
Throughput	40GB/s+	80GB/s+

KEY FEATURES:

- **Flexible Storage Access:** Supports both block and file access, providing compatibility with a wide range of applications and workflow
- **High-Speed Throughput:** Delivers over 80 GB/s of sustained throughput, enabling fast data ingestion and model training
- **GPU Direct Compatibility:** Supports GPU direct for reduced latency and increased data movement efficiency between storage and compute
- **Massive Storage Capacity:** Scales up to 737.3 TB of RAW storage, providing room for large datasets and long-term growth

BENEFITS:

- **Exceptional Value:** Offers high performance and capacity at a competitive price point, delivering strong ROI
- **Turnkey Deployment:** Supplied as a fully built and configured appliance with remote installation by Boston experts
- **HPC-Ready Scalability:** NFS protocol support allows seamless integration of additional HPC client systems over time
- **Low Maintenance:** Designed for simplicity-run it with minimal administration or ongoing oversight

EXAMPLE USE CASE:

CHALLENGE

Many organisations face a choice when buying storage to support NVIDIA DGX technology. They wish to invest as much as possible into DGX only to find that little of the budget remains for fast parallel storage. Enterprise All Flash does not deliver enough performance, while a traditional parallel storage system is too expensive.

SOLUTION

The Boston Igloo AI+ meets the need by providing blistering storage performance at a price point that can support the limitations of a small budget. The NFS file system means that DGX systems can be added to storage over time while not compromising on performance needed to support the AI workloads.

ADVANTAGE

The Boston Igloo AI+ has been designed especially for this common use case. The low cost of this fast storage solution means that on many occasions budget can be directed into more GPUs.



DDN EXASCALER

For decades, DDN has designed, developed, deployed and optimised systems, software and storage solutions that enable enterprises, service providers, universities and government agencies to generate more value and to accelerate time to insight from their data and information on-premise and in the cloud.

Developed and optimised using the latest advances in file system software technology, the DDN EXAScaler storage appliance delivers extreme performance, scalability, capability, reliability and simplicity.

EXAScaler introduces several new data management and integrity file system features developed by DDN and is only available in its appliances and cloud offerings. Stratagem is a powerful data orchestration engine that gives users comprehensive data residency controls using policy-based placement.

KEY FEATURES:

- **Enterprise Features in an HPC Array:** Combines robust enterprise storage capabilities with HPC-level performance, enabling broad use case support from day one
- **API-Driven Data Management:** Simplifies orchestration with an advanced integration tool for seamless data movement, automation and control
- **Comprehensive Data Services:** Includes snapshots, encryption, data integrity, quotas, workload analytics (SIA), multitenancy and multicloud readiness for enterprise-grade functionality

BENEFITS:

- **Multi-Protocol Flexibility:** Supports file (NFS & SMB), object (S3) and containerised environments, offering unmatched versatility across diverse infrastructures
- **Optimised Media Utilisation:** Leverages the full performance range of various storage media, ensuring the right balance of speed, cost and capacity for each data profile
- **Built for Hybrid Workloads:** Secure, flexible and performance-driven—ideal for organisations managing data across on-prem, edge and cloud environments



MULTICLOUD DEPLOYMENT OPTIONS



USE CASE:

CHALLENGE

A leading life sciences university required shared computational resources to accelerate genomic medicine research while staying within strict budget constraints.

SOLUTION

Enhanced existing infrastructure with a cost-effective, high-performance computing solution, integrating 200G HDR InfiniBand, 40/10 Gbps Ethernet and a DDN Storage system to improve simulations, data analysis and collaboration.

IMPACT

Delivered up to 90 GB/sec performance per appliance, significantly boosting throughput, scalability and research efficiency, enabling faster data processing and improved collaboration.



THE VAST DATA PLATFORM

The VAST Data Platform is a breakthrough approach to data-intensive computing that serves as the comprehensive software infrastructure required to capture, catalogue, refine, enrich, and preserve data through real-time deep data analysis and deep learning. This system is designed to provide seamless and universal data access and computing from edge-to-cloud, all from a platform that is designed for enterprises and cloud service providers to deploy on the infrastructure of their choosing.

The system takes a new approach to marrying unstructured data with structured data with declarative functions and can store, process and distribute data as a global data-defined computing platform. The VAST Data Platform also takes a first-principles approach to simplifying the data experience, by introducing several new architecture conventions that break long-standing trade-offs in data-intensive computing.



KEY FEATURES:

BENEFITS:

- **Massive Flash Scalability:** Scale to exabytes of affordable flash storage with shared NVMe accessible by every container—designed for high-growth, data-driven environments
- **High-Performance Fabric Support:** Compatible with data centre-scale Ethernet and InfiniBand, delivering low-latency, high-throughput access across your infrastructure
- **Flexible Client Connectivity:** Supports NFS, SMB, S3, GPUDirect and Kubernetes CSI, enabling seamless integration across diverse applications and architectures
- **Unified Storage Architecture:** Multi-protocol design breaks down silos by consolidating file, object and container storage into a single, high-performance platform
- **Optimised for Enterprise Workflows:** Built to handle mission-critical and data-intensive workloads, providing fast, unified access to all your data across the enterprise
- **Intelligent Infrastructure Efficiency:** VAST innovations enable always-on availability, ultra-resiliency and exceptional resource efficiency—powering the modern flash cloud

USE CASE:

CHALLENGE

Having developed as a lead provider of biological information to biopharma companies, a premier research organisation sought to build, host and manage an AI platform for its customers and partners to drive biological scientific discovery. This was at a time when the need for performance never stood still due to the need to keep pace with the evolving complex calculations that the platform needed to deliver.

SOLUTION

Having evolved various options, the institution chose to deploy the VAST Data Platform. The decision was made due to the value of an all-flash feature rich VAST storage to complement the Boston designed managed service solution and to deliver the performance needed to support scientific discovery.

ADVANTAGE

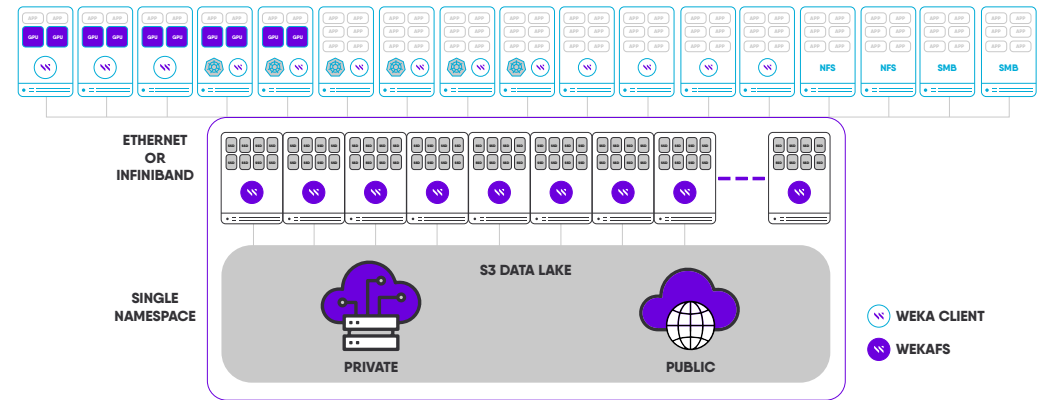
The solution supports the need to address many use cases. The all-flash architecture will contribute to energy savings over the lifetime of the solution and the researchers and partners will be supported by a storage platform that will enable researchers to deliver results rapidly in a world where the storage will also need to support the complexity associated with evolution of more complex calculations.



WEKA IO

Built for NVMe flash and cloud-native, WEKA is a parallel file system designed to enable organisations to maximise the full value of their high-powered IT investments - compute, networking and storage. By leveraging existing technologies in new ways and augmenting them with engineering innovations, WEKA's software delivers a more powerful and simpler solution that would have traditionally required several disparate storage systems. The resulting software solution delivers high performance for all workloads (big and small files, reads and writes, random, sequential and metadata heavy).

WEKA's Data platform was built from the ground up to shatter the limits imposed by legacy storage architectures by taking a unique software-only approach to address the needs of modern workloads.



KEY FEATURES:

- **Unified Data Management:** Manage all file and object data across NVMe, SSD flash and disk within a single global namespace for simplified operations and improved visibility
- **Intuitive User Interface:** A clean, easy-to-use GUI enables rapid provisioning, policy management, real-time system health monitoring and seamless data protection adjustments
- **Optimised for Mixed Workloads:** With no manual tuning required post-deployment, WEKAFS™ is built to handle varied workloads effortlessly, from AI to enterprise storage tasks

BENEFITS:

- **Operational Simplicity:** Designed to remove complexity and guesswork from data management, allowing teams to focus on innovation rather than administration
- **User-Friendly Experience:** Whether scaling out storage or fine-tuning policies, the system ensures a smooth management experience for both technical and non-technical users



USE CASE:



CHALLENGE

Genomics England had previously implemented a scale-out NAS solution from a leading vendor to support the 100,000 genome project; however, it had already hit its limit on storage node scaling and performance suffered when the system was near capacity.

SOLUTION

A two-tier architecture that takes commodity flash and disk-based technologies, presenting it as a single hybrid storage solution. The primary tier consists of 1.3 Petabytes of high performing NVMe flash storage which supports the working data sets.

ADVANTAGE

Genomics England was able to realise no limit on capacity scaling, over 10x improvement in performance, 75% reduction in storage cost per genome, as well as now embedding a full disaster recovery strategy and offering integration with public cloud for compute elasticity.

LIQUID COOLING



Liquid cooling is rapidly becoming an inevitable approach to HPC cooling. Why is that? The short answer is that air cooling will not be able to keep pace with the wattage of components. The long answer is that on top of meeting temperature requirements, it also works towards sustainability goals as liquid cooling requires less energy to maintain, and generally there is less maintenance with a liquid cooled setup.

Using liquids to cool a system can come in many shapes and sizes. There are two categories which then also break into two types; single-phase and dual-phase liquid cooling. Each approach can be applied directly on the chip or by immersing components, or even the entire system.

DIFFERENT TYPES OF LIQUID COOLING

DIRECT LIQUID COOLING (DLC)

In the realm of liquid cooling, single-phase DLC maintains a consistent liquid state, efficiently absorbing heat from components and subsequently dissipating it through cooling mechanisms before being recirculated, finding applications in both direct-on-chip and immersion cooling setups.

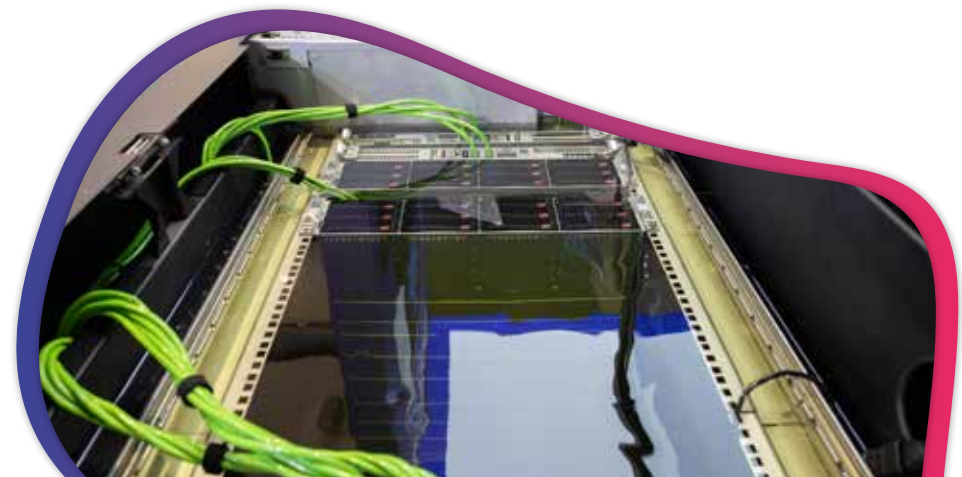
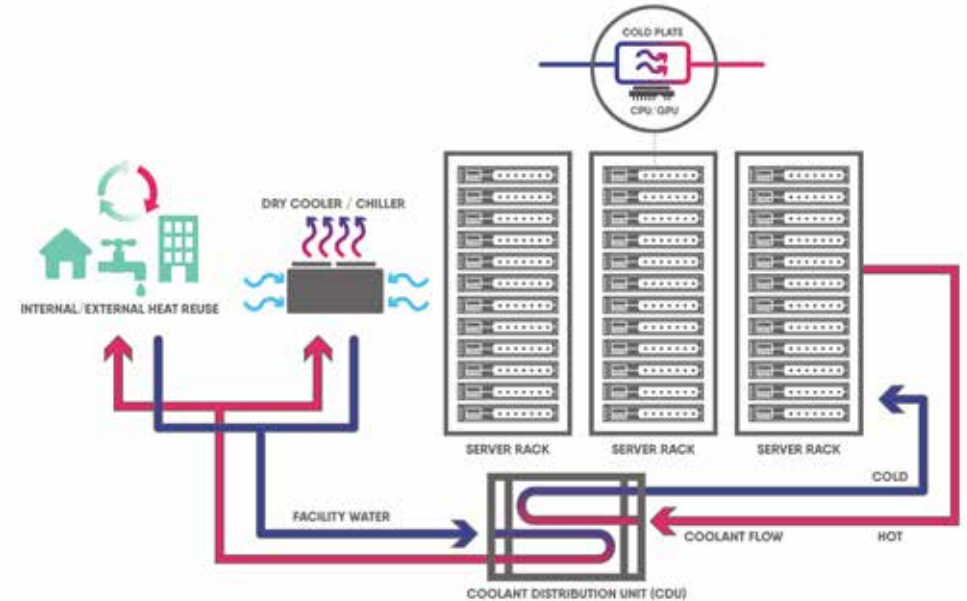
DUAL-PHASE DIRECT LIQUID COOLING (2-PHASE DLC)

Dual-phase DLC involves an intricate process where a liquid undergoes a phase change into vapour upon contact with heat and, after vapour cooling, it condenses back to its liquid form, a method often employed in immersion cooling; it relies on a precise choice of a fluorocarbon-based liquid that transitions states around 50°C, demanding careful handling.

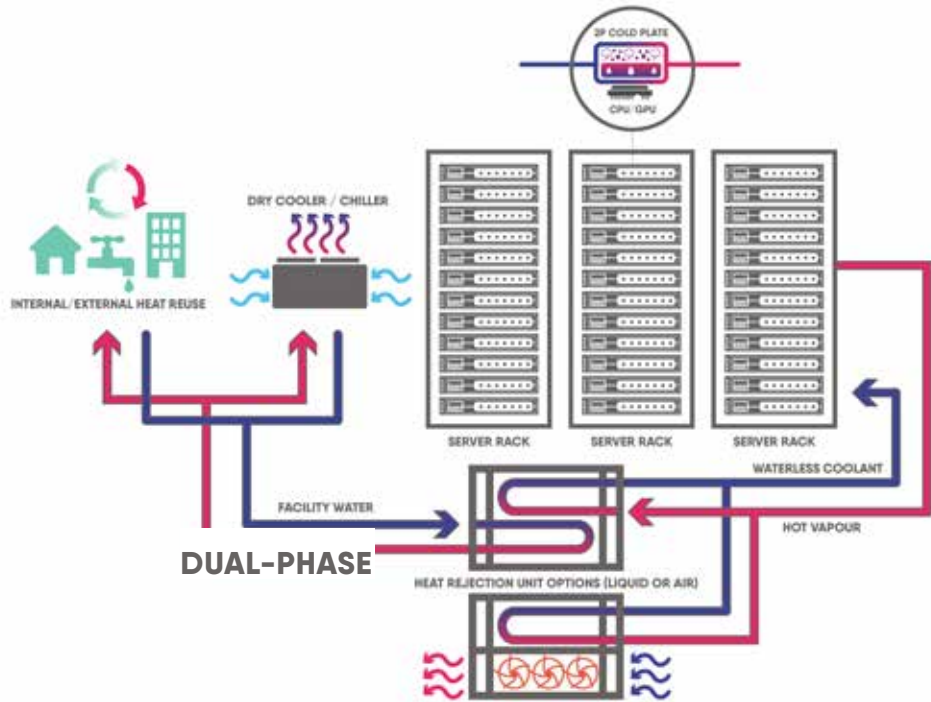
IMMERSION COOLING

Immersion cooling submerges components within a dielectric fluid, offering the choice of both single and dual-phase cooling loops, its dielectric properties ensuring safety as it efficiently absorbs and dissipates heat from electronic parts; with a secondary heat exchanger in play, the captured heat can be redirected for dissipation or repurposing, making it particularly well-suited for high-power applications such as data centres and industrial processes.

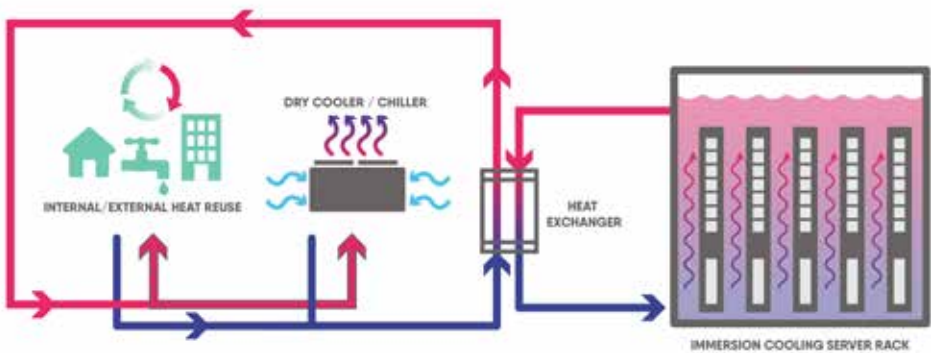
DIRECT LIQUID COOLING



TWO-PHASE DIRECT LIQUID COOLING



SINGLE-PHASE IMMERSION



LIQUID COOLING

BOSTON
Servers | Storage | Solutions

The Boston GlacierCore series is a range of liquid cooling-ready servers, designed for businesses that require robust, reliable performance with fully validated and warranted designs.

Castrol
ON

Boston Limited has partnered with Castrol, a global leader in lubricants and thermal management, to bring advanced cooling solutions to the data centre industry.

submer

Submer specialises in single-phase liquid immersion cooling, offering solutions that scale from edge data centres to hyperscale deployments, addressing the evolving challenges of high-density computing.

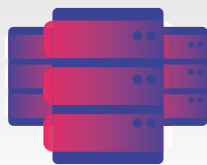
SUPERMICRO

Supermicro provides a comprehensive end-to-end liquid cooling ecosystem, from direct-to-chip cooling solutions to rack-level infrastructure and cooling tower integration, reducing data centre electricity costs by up to 51%.

ZUTACORE

ZutaCore is a two-phase direct-to-chip liquid cooling specialist, delivering high-performance, energy-efficient solutions that maximise compute density, system longevity and operational efficiency.

THE GLACIERCORE SERIES



PURPOSE-BUILT FOR LIQUID COOLING

Our GlacierCore range of servers is designed from the ground up for immersion cooling, ensuring maximum efficiency and performance. Built to handle the demands of modern data centres, these servers leverage Castrol's cooling fluids to optimise performance, reduce thermal stress and support sustainable operations.

GLACIERCORE MODELS OVERVIEW:

GlacierCore 1100

A compact and cost-effective solution ideal for entry-level deployments, offering efficient immersion cooling for essential tasks.



GlacierCore 1110

A powerful choice for high-density applications, designed for businesses requiring robust computing performance with reliable cooling.



GlacierCore 1200

Built for advanced workloads, this model supports intensive applications such as AI training, benefiting from high-performance cooling.



GlacierCore 1210

Designed to support data-intensive and mission-critical environments, this model ensures stability even under heavy processing demands.



GlacierCore 1220

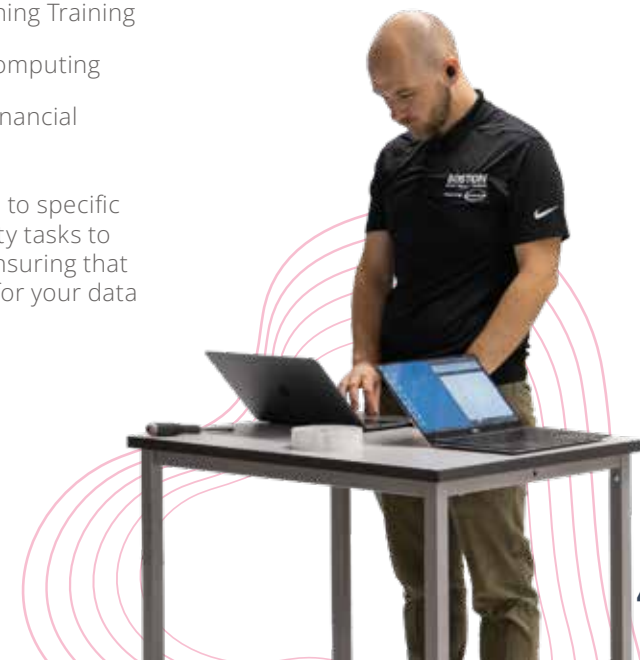
The top-tier choice for extreme processing needs, this model is optimised for AI, ML and HPC tasks, delivering sustained performance with unparalleled cooling efficiency.



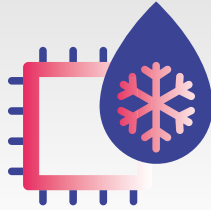
APPLICATIONS

- AI and Machine Learning Training
- High-Performance Computing
- Data Analytics and Financial Modelling

Each model can be tailored to specific use cases, from high-density tasks to intensive AI applications, ensuring that you find the ideal solution for your data centre's needs.



CASTROL



CASTROL OIL & IMMERSION COOLING TECHNOLOGY

At the core of our liquid cooling solutions is Castrol's pioneering immersion cooling technology. Castrol's specially engineered cooling fluids provide superior thermal management, ensuring efficient heat transfer even during high loads. This results in a lower energy footprint, helping organisations minimise operational costs while extending the lifespan of their equipment.

Castrol's immersion cooling fluids are engineered for superior heat transfer, optimising the cooling process and enabling data centres to operate more efficiently, this also minimises the need for traditional air conditioning which reduces energy consumption, supporting Boston's mission to deliver sustainable technology options to our customers.

KEY FEATURES:

- **Purpose-Designed Submersion Fluid:** Specially engineered to deliver exceptional thermal conductivity and cooling efficiency for high-performance computing environments
- **Environmentally Safe and Sustainable:** Non-toxic and eco-friendly formulation supports green initiatives and sustainability goals in modern data centres
- **100% Heat Encapsulation:** Ensures complete thermal containment for maximum cooling effectiveness and minimal energy loss

BENEFITS:

- **Significant Cost Reductions:** Immersion cooling dramatically lowers operational and cooling expenses, contributing to long-term financial savings
- **Extended Component Lifespan:** Hardware operates in a controlled environment, improving reliability and reducing the need for maintenance or early replacements
- **Air-Free Cooling Architecture:** Minimises or eliminates reliance on traditional air cooling systems, boosting energy efficiency and enabling a more sustainable infrastructure



LIQUID COOLING WITH BOSTON LIMITED

As data demands grow, so does the need for innovative cooling solutions that support high performance, efficiency, and sustainability. At Boston Limited, we're leading this evolution through our partnership with Castrol, a global leader in thermal management and immersion cooling fluids.

By combining Castrol's advanced liquid submersion technology with Boston's expertise in high-performance computing, we deliver cutting-edge cooling solutions that encapsulate 100% of the system's heat output. This approach significantly reduces or eliminates the need for air conditioning, lowering energy consumption and operational costs.

Additionally, cooler operating components experience improved reliability and a longer lifespan, leading to fewer failures, reduced maintenance requirements, and a lower total cost of ownership (TCO). With Boston and Castrol, businesses can achieve unparalleled efficiency while meeting their environmental sustainability goals.

WHY LIQUID COOLING?

Liquid cooling dissipates heat more efficiently than traditional air cooling, making it ideal for data centres running intensive workloads. By immersing components directly in a specialised cooling fluid, heat is removed at the source, ensuring consistent thermal management even under peak loads.

SUSTAINABILITY & COST-EFFICIENCY

With reduced reliance on air conditioning, liquid cooling significantly lowers energy consumption, helping businesses meet sustainability targets. Additionally, cooler operating temperatures improve hardware reliability and lifespan, leading to fewer failures, reduced maintenance needs and a lower total cost of ownership (TCO).

Our partnership with Castrol gives Boston customers access to customised, high-efficiency cooling solutions that go beyond traditional methods. Designed for demanding environments, these systems enhance reliability, reduce operational costs and support sustainable business growth.



SUBMER



Submer is a leading provider of single-phase liquid immersion cooling tanks, designed to make data centre operation and construction more efficient and sustainable. Their technology allows hyperscalers and large industries to improve energy efficiency, reduce costs and optimise space utilisation. Submer's approach goes beyond just providing cooling solutions—they develop and refine their products, APIs and installation processes to create future-proof infrastructures. Their solutions tackle key challenges such as heat reuse, net-zero water consumption and site selection, helping businesses meet both current and future sustainability goals.

Their single-phase immersion cooling technology is highly scalable, allowing data centres to increase density without compromising efficiency. Whether for edge, enterprise or hyperscale deployments, Submer's technology offers a flexible, space- and energy-efficient approach to high-performance computing.

KEY FEATURES:

- **Ultra-Low PUE Efficiency:** Achieve a power usage effectiveness (PUE) as low as 1.03, dramatically improving energy efficiency and lowering operational costs
- **High-Density Compute Support:** Accommodates up to 100kW of compute per Pod, enabling high-performance workloads and scalable growth within a compact footprint
- **Modular Pod Architecture:** Prefabricated building blocks reduce deployment times and simplify installation, accelerating time-to-value for new deployments

BENEFITS:

- **Extended Hardware Lifespan:** Consistent thermal management helps maintain optimal operating temperatures, improving long-term component reliability
- **Reduced Failure Rates:** Enhanced cooling performance contributes to lower hardware failure rates, minimising downtime and replacement cycles
- **Lower Infrastructure CAPEX:** Simplified infrastructure design cuts construction costs by up to 50%, making it ideal for rapid and cost-effective expansion
- **Sustainable Cooling Strategy:** Designed to minimise water consumption, supporting environmentally responsible data centre operations



SUPERMICRO



Supermicro's liquid-cooled rack solution delivers superior performance and efficiency for HPC, large-scale AI and cloud-scale compute infrastructure. As modern applications demand high-performance CPUs and GPUs, power consumption and heat generation have increased significantly. To address this, Supermicro provides a complete, end-to-end liquid cooling ecosystem, covering everything from servers and Coolant Distribution Units (CDUs) to manifolds and even cooling towers.

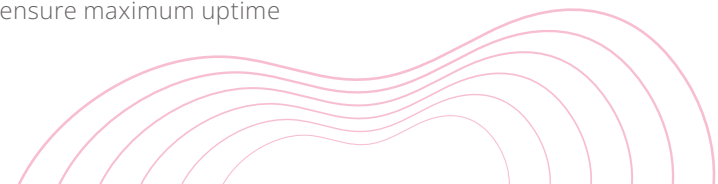
Supermicro offers a wide range of liquid-cooled servers, including 1U, multi-node, blade and GPU-accelerated solutions, supporting the latest AMD, Intel and NVIDIA processors and GPUs. Each component is designed in-house for maximum performance, reliability and seamless integration, delivering a fully integrated, rack-level solution.

KEY FEATURES:

- **Comprehensive End-to-End Solution:** From server to cooling tower, Supermicro provides a fully integrated liquid cooling infrastructure
- **Range of CPU and GPU-Accelerated Servers:** Supporting HPC, AI and analytics workloads across 1U, multi-node, blade and GPU-accelerated architectures
- **Enterprise-Grade Reliability:** Features redundant cooling pumps, power supplies, leak-proof connectors and leak detection to ensure maximum uptime

BENEFITS:

- **Full Turn-Key Single Source Solution:** A one-stop-shop experience, with all components designed to work together seamlessly
- **Accelerated Lead Times:** Benefit from in-stock inventory, reducing procurement and deployment delays
- **Up to 51% Reduction in Data Centre Electricity Costs:** Lower cooling requirements translate to significant operational cost savings





ZUTACORE

For organisations running compute- and data-intensive applications, optimal server performance and longevity depend on efficient cooling solutions. As each new generation of processors generates more heat and requires tighter thermal tolerances, traditional cooling methods struggle to keep up.

ZutaCore's two-phase dielectric liquid cooling technology offers a breakthrough solution. Unlike single-phase cooling, ZutaCore's system boils the coolant at the heat source, rapidly removing thermal energy from the processor. This phase change process enables CPUs to maintain high utilisation rates with zero throttling, improving performance and efficiency. Additionally, as a waterless system, ZutaCore eliminates the risks associated with traditional liquid cooling.

KEY FEATURES:

- **50% Reduction in Energy & Space Usage:** Optimises data centre footprint and power consumption
- **3x Processing Capacity:** Higher cooling efficiency allows for greater compute density
- **50% CapEx Reduction:** Lower infrastructure costs by reducing cooling complexity and footprint

BENEFITS:

- **Two-Phase Direct-to-Chip Cooling:** Boiling coolant extracts heat at the source, ensuring maximum thermal efficiency
- **<1000W of Cooling Per Chip:** Supports the highest-performance CPUs and GPUs without throttling
- **Waterless Dielectric Liquid Cooling:** Non-conductive coolant enhances safety and reliability



NVIDIA NETWORKING



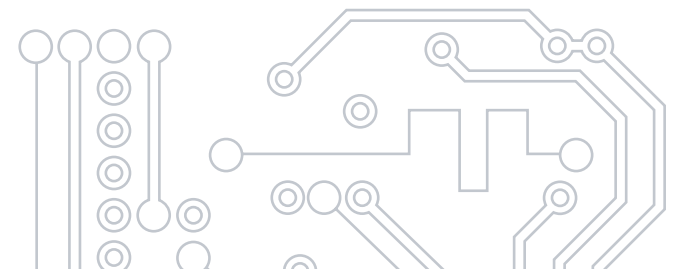
NVIDIA Networking Ethernet products provide scalable, low-latency and efficient fabric with simplified management, reducing overall costs and power consumption compared to traditional Ethernet.



NVIDIA Networking InfiniBand switches and adapters support up to 400Gb/s, enabling scalable fabric for high-performance computing and data centres.



The NVIDIA Spectrum™-X networking platform is the first Ethernet platform designed specifically to improve the performance and efficiency of Ethernet-based AI clouds.





NVIDIA ETHERNET SPECTRUM®-4 PLATFORM

The world's first 400Gb/s end-to-end networking platform, NVIDIA Spectrum®-4 provides 4x higher switching throughput than previous generations, with 51.2 Tb/s per second.

It consists of the NVIDIA Spectrum®-4 switch family, NVIDIA ConnectX®-7 SmartNIC, NVIDIA BlueField®-3 DPU and the DOCA™ data centre infrastructure software to supercharge cloud-native applications at scale. Built for AI, NVIDIA Spectrum®-4 Switch arrives as data centres are growing exponentially and demanding extreme performance.

NVIDIA Spectrum®-4 features a fully-shared and monolithic packet buffer that's dynamically available to all ports. This provides excellent microburst absorption with true, port-to-port, cut-through latency.

KEY FEATURES:

- **400Gb/s End-to-End Networking:** Delivers ultra-fast, high-bandwidth connectivity to meet the demands of modern AI, HPC and cloud workloads
- **51.2Tb/s Switching Capacity:** Ensures ultra-low latency and high-performance data flow, supporting data-intensive and latency-sensitive applications
- **4x Higher Throughput:** Offers four times the switching throughput of previous generations, enabling exceptional scalability and system responsiveness

BENEFITS:

- **Extreme Performance:** Designed to power next-gen workloads with unmatched speed, ensuring smooth data transfer across large-scale deployments
- **Enhanced Efficiency:** Optimised networking architecture improves energy and resource utilisation, contributing to more efficient data centre operations
- **Built for the Future:** Scalable, high-speed infrastructure future-proofs your network for growing workloads and evolving performance requirements



NVIDIA INFINIBAND QUANTUM™-2 PLATFORM

The NVIDIA Quantum™-2 based QM9700 and QM9790 switch systems deliver an unprecedented 64 ports of NDR 400Gb/s InfiniBand per port in a 1U standard chassis design.

A single switch carries an aggregated bidirectional throughput of 51.2Tb/s per second (Tb/s), with a landmark of more than 66.5 billion packets per second (BPPS) capacity.

Supporting the latest NDR technology, NVIDIA Quantum™-2 brings a highspeed, extremely low-latency and scalable solution that incorporates state-of-the-art technologies.

KEY FEATURES:

- **Low Latency, High-Speed Connectivity:** Built for high-speed, low-latency performance, perfect for environments where speed and responsiveness are critical
- **Massive Throughput:** Provides an aggregated bidirectional throughput of 51.2Tb/s, handling massive data flows with ease
- **64 Ports of NDR 400Gb/s:** Delivers 64 ports of NDR 400Gb/s, enabling ultra-fast connectivity for large-scale, data-intensive applications

BENEFITS:

- **Advanced Technology Integration:** Utilises state-of-the-art technologies for optimal performance, ensuring future-proof scalability and cutting-edge capabilities
- **High Performance Capacity:** Supports a 66.5 BPPS capacity, delivering robust performance for high-demand workloads like AI and HPC
- **Compact 1U Design:** Offers a 1U standard chassis for efficient use of space, ideal for dense, high-performance deployments





NVIDIA SPECTRUM™-X

Spectrum™-X is a revolutionary solution created by NVIDIA for building multi-tenant, hyperscale AI clouds and AI Factories with Ethernet. With it, organisations can significantly improve the performance and power efficiency of AI clouds and gain higher predictability and consistency, resulting in faster time to market and a greater competitive edge.

NVIDIA Spectrum™-X enhances performance and power efficiency by up to 1.7X, accelerating the processing, analysis and execution of AI workloads and, in turn, the development and deployment of AI solutions. Spectrum™-X is fully standards-based Ethernet with support for open Ethernet stacks (SONiC) at cloud scale, and is tuned and validated across the full stack of NVIDIA hardware and software, creating an unmatched Ethernet solution for AI clouds.

KEY FEATURES:

- **Near-Perfect Bandwidth at Scale:** Delivers exceptional bandwidth across large-scale environments
- **Extremely Low Latency:** Provides ultra-low latency for fast, real-time data communication
- **Advanced RoCE Extensions:** Enables scalable AI communications with Remote Direct Memory Access (RoCE) extensions
- **Deterministic Performance & Isolation:** Guarantees predictable performance and effective isolation
- **Open Network OS Support:** Fully compatible with open network operating systems like SONiC and Cumulus

BENEFITS:

- **Enhanced AI Cloud Performance:** Maximises AI cloud performance by ensuring efficient data handling and reducing bottlenecks
- **Reduced Total Cost of Ownership (TCO):** Lowers overall infrastructure and operational costs through optimised networking and performance
- **Faster Time to Results:** Accelerates AI and data processing workflows, enabling quicker insights and decision-making
- **Full-Stack Optimisation:** Streamlines the entire network stack for improved efficiency, providing a unified, optimised platform

SOFTWARE



DKube's software is an AI and machine learning platform that streamlines the end-to-end data science workflow, from data preparation and model training to deployment and monitoring.



NVIDIA Base Command™ is the software backbone of the NVIDIA DGX™ platform, delivering a complete stack for managing AI development and deployment, combining robust workflow orchestration, powerful cluster management tools, accelerated libraries for compute, storage and networking, and system software fine-tuned for demanding AI tasks.



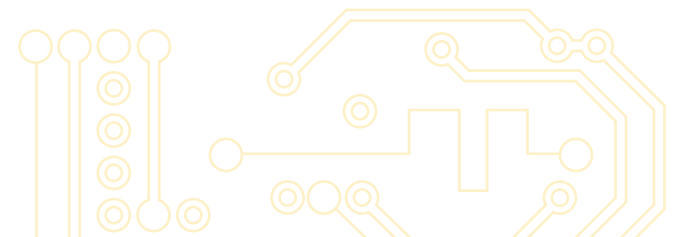
NVIDIA Mission Control™ is a next-generation orchestration platform built to manage every layer of your AI factory, from developer workloads to infrastructure and facilities. Supporting advanced architectures like Grace Blackwell, it delivers hyperscale agility with full-stack intelligence, workload scheduling via NVIDIA Run:ai, and autonomous recovery for rapid issue resolution.



OpenHPC is a modular, open-source framework designed for deploying and managing HPC clusters. It simplifies cluster management by providing pre-packaged, integrated software components, including tools for provisioning and resource management.



TrinityX is a cluster management solution focused on HPC and AI workloads. It enables efficient orchestration, monitoring and scaling of HPC clusters.





DKUBE

DKube is a powerful data integration and analytics platform designed to streamline how organisations manage and analyse their data. It offers a comprehensive set of tools for aggregating data from various sources, performing advanced analyses, and visualising results through customisable dashboards.

With its capability for real-time data processing and scalable architecture, DKube supports businesses in making informed decisions and staying competitive in a dynamic market.

KEY FEATURES:

- **Seamless Data Integration:** Integrates effortlessly with various data sources and formats, enabling efficient aggregation and analysis for diverse needs
- **Powerful Analytics Tools:** Equipped with advanced analytics capabilities to derive actionable insights from complex data sets, facilitating smarter decisions
- **Customisable Dashboards:** Users can design personalised dashboards to visualise data in ways that meet specific business requirements
- **Real-Time Data Processing:** Supports real-time data processing, ensuring up-to-date information for faster, more accurate decision-making
- **Scalable Platform:** Designed to grow alongside expanding data and user demands, ensuring long-term usability and adaptability

BENEFITS:

- **Informed Decision-Making:** Provides real-time, comprehensive insights to help businesses make data-driven, informed decisions
- **Streamlined Data Management:** Reduces the time and effort spent on data integration and analytics, allowing teams to focus on higher-value tasks
- **Enhanced Data Visualisation:** Customisable dashboards and visualisation tools improve data accessibility and user understanding, making complex data more digestible
- **Market Insight Advantage:** Leverages advanced analytics to gain deeper insights into trends and opportunities, helping businesses stay ahead of the competition





NVIDIA BASE COMMAND MANAGER™

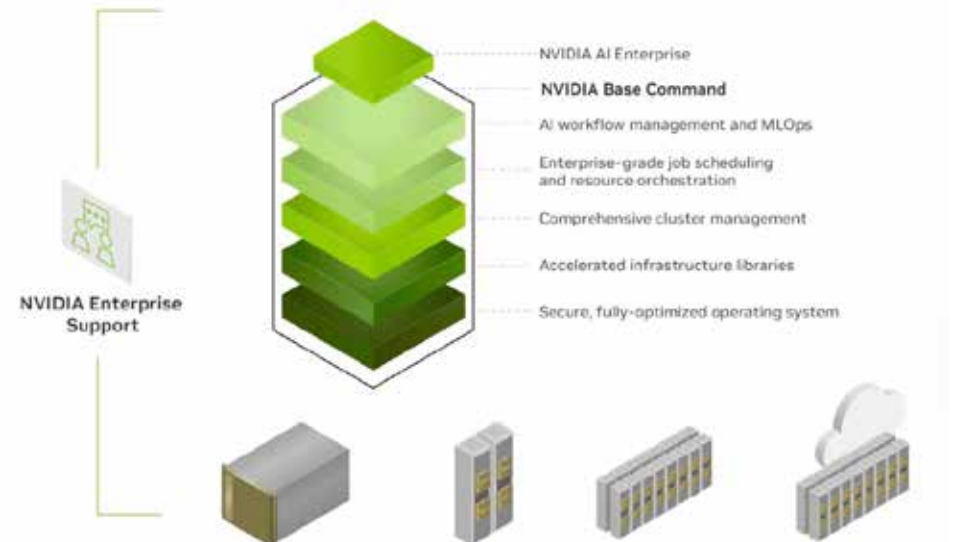
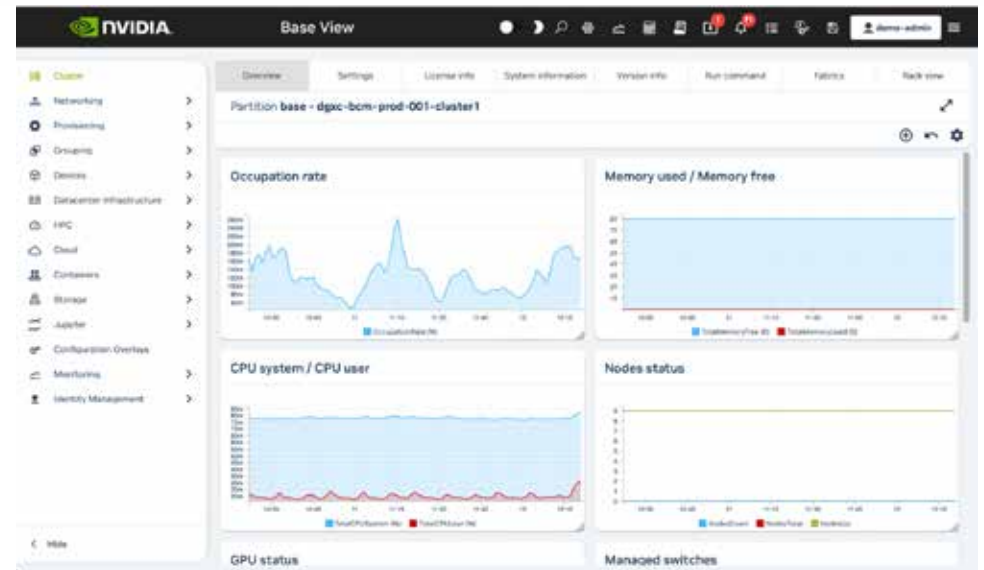
NVIDIA Base Command Manager is an enterprise-grade cluster management solution designed to streamline the deployment, operation and scaling of AI and HPC infrastructure. It automates provisioning, orchestrates workloads and provides comprehensive monitoring tools, enabling organisations to efficiently manage GPU-accelerated clusters across on-premises, edge and hybrid cloud environments.

KEY FEATURES:

- **Automated Cluster Provisioning:** Simplifies the setup of heterogeneous clusters, supporting a range from a few nodes to hundreds of thousands, and integrates seamlessly with Kubernetes for workload orchestration
- **Comprehensive Monitoring and Management:** Offers real-time health checks, system diagnostics and performance metrics through a unified interface, ensuring optimal operation of AI and HPC workloads
- **Integration with NVIDIA Ecosystem:** Fully compatible with NVIDIA DGX systems, BlueField DPUs and Spectrum switches, providing a cohesive environment for accelerated computing tasks

BENEFITS:

- **Enhanced Operational Efficiency:** Reduces administrative overhead by automating routine tasks, allowing IT teams to focus on strategic initiatives
- **Scalability:** Supports growth from small-scale deployments to large, complex infrastructures without compromising performance or manageability
- **Improved Resource Utilisation:** Ensures optimal use of GPU resources, leading to faster time-to-insight for AI and HPC applications





NVIDIA MISSION CONTROL™

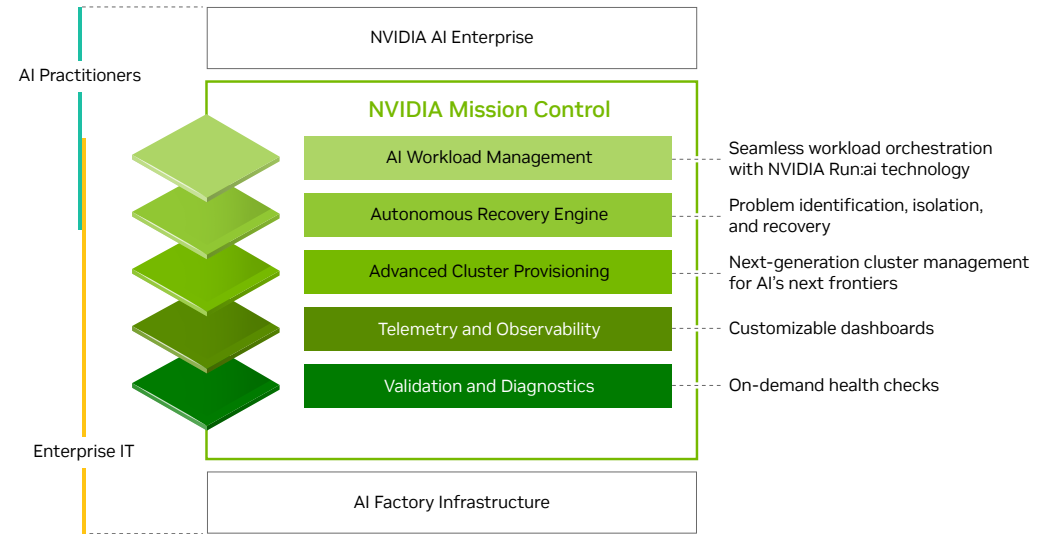
NVIDIA Mission Control is a next-generation operations and orchestration platform built to run every layer of your AI factory - from developer workloads to infrastructure and facilities. It delivers the agility and efficiency of a hyperscaler while removing the complexity and overhead traditionally placed on IT teams. Designed to support the most advanced NVIDIA architectures, including Blackwell and Grace Blackwell systems, Mission Control brings instant responsiveness to AI workloads and enables infrastructure resiliency through full-stack intelligence.

With built-in capabilities like NVIDIA Run:ai for optimised workload scheduling and autonomous recovery for rapid issue resolution, Mission Control transforms how enterprises deploy and manage AI. It boosts resource utilisation by up to 5x, accelerates AI experimentation with 10x faster recovery from failures and ensures enterprise-grade support is always on hand.

KEY FEATURES:

BENEFITS:

- **Unified AI Operations:** Automates and orchestrates workloads, infrastructure and facilities from a single platform
- **Dynamic Workload Scheduling:** Integrates with NVIDIA Run:ai to optimise resource allocation across training and inference
- **Autonomous Fault Recovery:** Detects, isolates and resolves hardware and job failures up to 10x faster
- **Hyperscale Agility:** Delivers cloud-like flexibility and responsiveness across AI data centres
- **Maximised Resource Efficiency:** Increases infrastructure utilisation by up to 5x for cost-effective scaling
- **Accelerated AI Development:** Reduces downtime and simplifies experimentation for faster time-to-insight





OPENHPC

Boston offer a complete solution built around OpenHPC's comprehensive stack of management tools, Our team of experienced HPC engineers, can take you through the journey of designing and implementing a balanced, functional and cost effective cluster management solution.

OpenHPC is a collaborative, community-driven project that provides a comprehensive collection of open-source software components designed to simplify the deployment and management of HPC Linux clusters. It offers a modular framework that includes provisioning tools, resource management, I/O clients, development tools and a variety of scientific libraries, all pre-built and tested for HPC environments.

KEY FEATURES:

- **Comprehensive Software Stack:** OpenHPC aggregates a wide range of pre-packaged software components essential for HPC systems, facilitating easier installation and management
- **Modular Architecture:** Its modular design allows users to select and integrate only the components that meet their specific needs, providing flexibility in cluster configuration
- **Community Collaboration:** As a community-driven initiative, OpenHPC benefits from collective expertise, continuous updates and shared best practices, ensuring a robust and evolving ecosystem

BENEFITS:

- **Simplified Deployment:** By offering pre-built packages and standardised configurations, OpenHPC reduces the complexity involved in setting up HPC clusters, enabling faster deployment
- **Flexibility:** Users can tailor their HPC environments by selecting components that align with their specific requirements, allowing for customised solutions
- **Scalability:** OpenHPC's framework supports the scaling of resources to meet the demands of growing computational workloads



TRINITYX

TrinityX is an advanced HPC and AI cluster management solution that provides 24/7 oversight with SLA-backed reliability. Designed to simplify cluster administration, TrinityX handles the complexities of Linux, SLURM, CUDA, InfiniBand, Lustre and Open OnDemand, allowing you to focus entirely on your research and innovation.

Built by the co-founder of Bright Computing, TrinityX brings decades of industry expertise to your infrastructure—combining proven best practices with modern innovation for maximum reliability and performance. Modern AI and HPC environments require efficiency, scalability and ease of management. TrinityX optimises cluster operations while offering seamless integration with Kubernetes, Singularity and other leading technologies in an intuitive, graphical and command-line interface—suitable for users of all levels.

It supports both AMD and NVIDIA accelerators, and is designed for on-premise, cloud and hybrid deployments, giving you full control over where and how your workloads are managed.

KEY FEATURES:

- **Rapid & Secure Cluster Deployment:** Intuitive setup streamlines cluster provisioning, ensuring fast, secure deployment
- **Centralised Monitoring & Alerts:** Proactive system monitoring with real-time alerts to prevent disruptions and maintain system health
- **Intelligent Resource Management:** Dynamic power scaling ensures optimal efficiency, adapting to fluctuating workloads

BENEFITS:

- **Operational Agility:** Open-source flexibility and a modular architecture let you adapt to evolving workloads with ease
- **Enterprise-Grade Reliability:** Designed for high availability and stability, allowing users to focus on research rather than infrastructure
- **Fully Managed Support:** SLA-driven Remote System Administration, ensuring proactive and hassle-free cluster management

BOSTON SERVICES

At Boston, we bring together a diverse range of expertise and services under one roof to meet all your technology needs. Whether you're looking for cutting-edge solutions, in-depth training, or reliable support, we offer a complete suite of services to support your organisation's success. From our global team of technical and HPC experts who benchmark and optimise new technologies, to our Boston Training Academy that nurtures talent and education and our Boston Labs that provide custom tech solutions and testing facilities—we cover every aspect of your technology journey.



BOSTON TECHNOLOGY CONSULTING

Our technical and HPC experts, based around the world, are constantly benchmarking new technologies, the results of which we share with our customers, to deliver fully optimised solutions.



BOSTON TRAINING ACADEMY

The mission of the Boston Training Academy (BTA) is to become a renowned developmental ground for talent engagement, education and solutions across a variety of disciplines.



BOSTON LABS

Boston provides tech facilities to test emerging technologies and enterprise-grade systems from leading manufacturers. Boston Labs is customer-centric, customised solutions with off-the-shelf components to enhance business outcomes.



PROFESSIONAL SERVICES

Our senior engineers understand the technical dependencies and requirements of your organisation, we will ensure a well thought out installation is managed and completed on schedule and with the utmost professionalism.



SUPPORT AND ONSITE WARRANTY

From tailor-made SLA solutions, warranty support and spares packages – Boston offers customised services level work packages for ongoing support.

BOSTON TECHNOLOGY CONSULTING



Our technology consultancy service helps organisations to effectively create, optimise and grow their business in today's data and IT focused market. Our consultants mastermind solutions for the most complex business needs in an effort to achieve their goals, while balancing costs and performance. Our core values are leadership, integrity, collaboration, accountability and simplicity.

Our global reach enables us to provide agile and scalable solutions to everyone from the smallest players to industry heavy hitters. Our main objective is to solve current problems while minimising risks of future issues. Such all-encompassing approach to problem solving has been utilised in the banking, financial services and insurance (BFSI), government, manufacturing, oil and gas, logistics, telecoms, pharma and retail.

Our consultancy service is aimed at enterprises striving to utilise specialised services to optimise their solutions and enhance efficiency. With our deep industry knowledge, we provide valuable guidance to unlock the full potential of your Infrastructure.

- **Cross-domain Expertise:** We stand out through our wide-ranging expertise across multiple technical domains, allowing us to provide holistic solutions that address complex challenges spanning various technologies.
- **Innovative Problem Solving:** We excel in innovative problem-solving, employing a team of seasoned consultants who approach challenges with creativity and resourcefulness, leading to unique solutions that drive customer success.
- **End-to-end Project Management:** We offer comprehensive project management services, overseeing projects from inception to completion. This end-to-end approach streamlines processes, reduces customer workload and ensures successful project outcomes.



BOSTON TRAINING ACADEMY



Boston Training Academy provides structured training by world-class trainers who deliver tailored courses based on the current knowledge of the attendees. Our program content is designed by industry experts and covers data science skills, exposure to analytical tools and corporate learning and consulting on AI projects, to name a few.

The AI Centre of Excellence (CoE-AI) at our Training Academy brings together experts from industry academia and Government to provide the latest infrastructure, smart tools, skilled resources and leadership for AI exploration. The CoE-AI aims to create the most sophisticated ecosystem for AI, Data Sciences, Machine Learning and Deep Learning to enable breakthrough innovations.

We also offer online starter courses for complete beginners to gain working knowledge on the fundamentals of NLP, computer vision, deep learning and Python for data science. Boston also host NVIDIA Deep Learning Institute (DLI) workshops. Participants receive hands-on training sessions, with help from developers, data scientists and researchers. Attendees learn how to approach challenges using deep learning techniques like building transformer-based NLP, building conversational AI applications, building intelligent recommender systems, among other approaches.

We provide practical training to universities and businesses on algorithm creation, hardware development and the latest technologies. Our expert instructors empower organisations to stay competitive in a rapidly evolving digital world.

- **Expert-led Training:** We offer expert-led training programs delivered by industry professionals, ensuring participants gain practical insights and skills that are immediately applicable in their professional roles.
- **Tailored Learning Paths:** We excel in providing personalised learning paths, allowing participants to choose courses that match their skill levels and career aspirations. This tailored approach ensures that individuals receive relevant and effective training.
- **Hands-on Lab Experience:** With a strong emphasis on practical learning, we offer hands-on lab experiences that enable participants to engage with real-world scenarios, fostering better understanding and confidence in their abilities.



BOSTON LABS



DISCOVER THE FUTURE WITH BOSTON LABS: YOUR TESTING PARTNER

Are you navigating the ever-evolving landscape of new and enhanced technologies? We understand the challenge of making informed decisions for your upcoming projects. The task becomes even more daunting when you can't experience and assess the hardware firsthand prior to making a commitment.

Through our Boston Labs, we bridge this gap for you. Our state-of-the-art, remote accessible, on-site test facilities offer the perfect environment to explore and evaluate emerging technologies. We recognise that sometimes, the true fit of a product or solution can only be gauged through hands-on experience.

Our dedicated technical team is poised to assemble comprehensive solution stacks using cloud-based lab systems or even loaned hardware that can be seamlessly integrated into your environment. This ensures that your investment aligns perfectly with your requirements. Immerse yourself in our incredible R&D lab, Boston Labs, where an array of the latest hardware and software solutions await your evaluation. Whether you

choose remote assessment or prefer to collaborate closely with our seasoned field application engineers, the choice is yours.

Our veteran team excels in guiding you through the setup and optimisation of both hardware and software components, crafting solutions that harmonise performance, power efficiency and cost-effectiveness. When you reserve a session at Boston Labs, you gain access to expert advice on product selection and fine-tuning, helping you achieve unparalleled outcomes.

EXPERTS IN AI AND DATA

Boston Training Academy and Boston Technical Consulting are our in-house initiatives which represent the pinnacle of technological advancement, driving forward groundbreaking solutions that harness the power of artificial intelligence.



PROFESSIONAL SERVICES



PRESALES CONSULTATION AND SERVICES

Our Presales Consultation and Services aim to provide optimal solutions for your business needs. Our team of experts is dedicated to understanding your requirements, offering personalised consultations and delivering our wide range of technology solutions. We go beyond product knowledge, bringing a deep understanding of industry trends and emerging technologies to help you stay ahead of the competition.

SERVICES INCLUDE:

- Consultancy and Solution Design
- Inventory and Site Survey
- Technical Account Management
- Boston Labs and POC Services
- Bespoke System Design and OEM Manufacturing
- Product and Technology updates
- Environmental and Regulatory Testing

POST SALES AND IMPLEMENTATION

Our Post Sales and Implementation team ensures a smooth and successful deployment of your technology solutions. With skilled project management expertise, thorough testing, seamless integration and comprehensive training, we prioritise your success from purchase to implementation.

SERVICES INCLUDE:

- System Integration and Burn in Testing
- Racking and Installation
- Project Management
- Global delivery
- Turnkey Solutions and Rack and Roll
- Training and Education

DATA CENTRE AND HOSTING SERVICES

Your business relies on a secure, reliable and high-performance infrastructure to thrive in the digital landscape. Boston offers state-of-the-art data centre and hosting solutions designed to meet your specific needs. Our industry expertise delivers robust and scalable data centre solutions, including dedicated servers, virtual private servers (VPS), cloud hosting and colocation services. With enterprise-grade infrastructure, customised solutions, 24/7 support and scalability, we empower your digital infrastructure to reach its full potential. Choose Boston's Data centre and Hosting Services for exceptional performance and reliability.

SERVICES INCLUDE:

- Hosting
- Leasing
- Managed Services and Cloud
- Data centre and Rack Design
- Rack and Roll

END OF LIFE AND EXTENDED LIFE CARE

As technology rapidly evolves, effectively managing end-of-life products and systems becomes crucial. Boston Limited's End of Life and Extended Life Care services provide comprehensive solutions to navigate this critical phase of your technology lifecycle. Our expert team understands the challenges associated with retiring outdated systems and mitigating risks related to unsupported or vulnerable hardware. With our services, we ensure a seamless transition to newer technology while maximising the value of your existing investments.

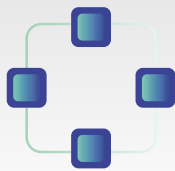
SERVICES INCLUDE:

- Decommissioning, Recycling and Disposal
- Out of Warranty Repair and Refurbishment

Please contact us if you are interested in discussing any of our additional service offerings.



SUPPORT AND ONSITE WARRANTY



At Boston, exceptional customer support is the foundation of our successful partnerships. Our Support and Onsite Warranty services offer peace of mind, knowing that our dedicated team of experts is ready to assist you whenever you need help.

With prompt and reliable assistance available 24/7, our experienced professionals with deep expertise across various technologies are just a phone call away. We provide onsite support to minimise downtime and ensure smooth operations, coupled with warranty coverage for eligible hardware, ensuring swift replacement or repair.

WARRANTY SERVICES	RETURN TO BASE (RMA)	ADVANCED WARRANTY	ONSITE WARRANTY			
			PEWTER	BRONZE	SILVER	GOLD
TECHNICAL SUPPORT DESK	YES	YES	YES	YES	YES	YES
ADVANCED PARTS SHIPPING	NO	YES	NO	YES	YES	YES
REMOTE ISSUE TRIAGE	YES	YES	YES	YES	YES	YES
ONSITE SUPPORT/DIAGNOSIS	NO	NO	NO	NO	NO	NO
ONSITE BREAK/FIX LABOUR	NO	NO	YES	YES	YES	YES
ENGINEER TO SITE TARGET (POST DIAGNOSIS)	N/A	N/A	NBD	NBD	SBD	4 HOURS
ACCESS TO SERVICES 24 HOURS A DAY, WEEKENDS & UK BANK HOLIDAYS	NO	NO	NO	NO	NO	YES



