# BOSTON
## Servers | Storage | Solutions

# HPC BROCHURE

## ISSUE 2019, Q3

# CONTENTS

# SERVICES

## SERVICES

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

**BOSTON LABS**
Boston has, in excess of, $2m of hardware that's available to customers, either onsite or via remote access, to test out the latest technologies and architectures. Boston customers can run their own workloads, benchmarks and simulations on configured systems prior to purchase.

**TAILOR MADE SOLUTIONS**
Boston deliver bespoke solutions matched to customers' applications and requirements. Through the utilisation of hardware and software accelerators, our solutions are designed around upgraded performance and faster workflows.
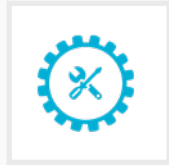
**PLANNING & INSTALLATION**
Our senior engineers understand the technical dependencies and requirements of your organisation, we will ensure a well thought out installation is managed and completed on schedule and with the utmost professionalism.

**QUALITY ASSURANCE**
Boston has numerous state-of-the-art build facilities, that are managed by our in-house engineering teams, to ensure all our solutions are built, and configured, to the highest standards for quality, stability and performance.
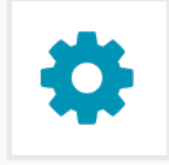
**ADDITIONAL MANAGED SERVICES & SUPPORT**
From tailor-made SLA solutions, warranty support and spares packages – Boston offers customised services level work packages for the ongoing support and management of your solutions.

# HPC CLUSTER ARCHITECTURE AND WORKFLOW

When it comes to the High Performance Computing (HPC) market, clusters are rapidly reshaping it, driven in the most part by price/performance and the open source Linux OS.

Even though there are endless opportunities that clusters offer, and clusters, comprised of commodity server hardware and software are gaining acceptance...getting a cluster running, and learning how to use it, requires skilled resources, services and time.

If not installed correctly, this can lead to badly planned clusters with software that hasn't been chosen appropriately.

The result? IT departments turn to costly SMP alternatives because of the standardisation shortcomings of cluster computing.

Boston understand the end-to-end process of building, implementing and managing clusters, so cost is minimised, as is the time to get a cluster fully operational.

Boston offers a range of both open source and commercial cluster management packages, that are selected in-line with our clients environment and workflow.

We are able to pre-engineer and test multiple variations of hardware and software packages in our Boston Lab facilities - making our Linux compute clusters easy to deploy, simple to use, consistent, transparent, turnkey and available.

01 **INGEST**

02 **PARALLELISE**

03 **CLUSTER MANAGEMENT**

04 **PRIVATE CLOUD**

05 **SCHEDULERS**

06 **SECURITY**

07 **CONTAINERISATION**

# WORLD-CLASS BOSTON LABS

The first Boston Labs facility was launched in 2005 just outside London, when it became apparent that our customers were not able to gain a full understanding of latest generation hardware and its application against their workloads/use cases.

Following the success of the UK Lab, further facilities were opened in Germany and India, with future sites earmarked for the future.

01 **LEVERAGE OUR EXPERTISE**

02 **EXPERT HELP EVERY STEP OF THE WAY**

03 **AT YOUR CONVENIENCE WITH ANYTIME REMOTE DIAL-IN ACCESS**

04 **BE THE 1ST TO ACCESS THE LATEST AND FUTURE TECHNOLOGIES**

05 **ON-SITE TESTING AVAILABLE**

Understandably, some customers are unable to visit our labs in person. Therefore, arrangements can be made for test systems to be set up in any of our labs with remote access via RDP, SSH or other means; enabling trials on Boston solutions from wherever you are in the world.

The Boston Labs facilities have grown in popularity to such an extent that even leading hardware manufacturers and industry journalists utilise them for benchmarking and cluster configuration.

To find out the latest technologies available for testing please visit our website.

# TAILOR MADE & CUSTOMISED

# PLANNING, INSTALLATION, TESTING & QA

As the chosen OEM manufacturing partner for numerous high-profile brands, we have a great deal of expertise in customising our solutions to meet your exact requirements.

Boston offers custom configuration of a server, workstation or solution needed by your clients, confident that all aspects of the design have been qualified by Boston for compatibility, cooling and expandability.

Many clients choose to have their solutions fully branded by us in order to hide the origins of the original hardware and to help identify their products as unique offerings within their target market sectors.

From simple customised stickers in with your logo, all the way through to fully custom-manufactured bezels, enclosures and packaging; we have the in-house expertise to suit your requirements.

- **CHASSIS ENCLOSURE DESIGN**

- **UNIQUE COLOURING OF CASES, BUTTONS AND SWITCHES**

- **BESPOKE ONE OFF "SHOW" DEMONSTRATION CHASSIS**

- **COMPANY BRANDED BEZELS AND SIDE PANELS**

- **CLIENT SPECIFIC BADGES AND STICKERS**

- **CLIENT BRANDED MANUALS & DOCUMENTATION**

- **APPLIANCE SPECIFIC BIOS BOOT SCREENS**

- **ON-SITE TESTING AVAILABLE**

Boston's in-house validation, build, test and QA procedures are second to none; however, we can augment these to reflect your specific needs; including full rack-scale design.

Depending on your circumstances, our expert solution architects can be available to work with you right from the early design phase, all the way through to the delivery and installation of your solution.

**01 CONSULT**

**02 DESIGN**

**03 ASSEMBLE**

**04 CONFIGURE**

**05 TEST**

**06 SHIP**

**07 INSTALL**

**08 SUPPORT**

With over 25 years in the industry, and access to leading-edge technologies such as immersed-computing; and leading software vendors, Boston are uniquely placed to offer a full rack-scale design that includes BIOS, firmware, IPMI, OS and customer imaging – including ongoing cluster management.

For more information on Boston's rack-scale design service visit bit.ly/BostonServices

# ADDITIONAL MANAGED SERVICES & SUPPORT

# GPU SOLUTIONS

Once you have received your new system, our after sales support team are on hand to answer any queries or problems that you may have.

Every support engineer is trained to deal with requests quickly and effectively, using escalation procedures where necessary to ensure maximum up-time. In addition to our standard warranty, we can offer fully tailored service solutions should this be required.

## GPU SOLUTIONS FOR HPC

### HPC CONFIGURATION CASE STUDY

**Challenges:**
The UK's Science and Technology Facilities Council (STFC) needed to upgrade the storage infrastructure for the JASMIN super data cluster. With demand for growth of up to 300PB in the next few years.

**Solution:**
BostonLabs liaised with STFC to understand their specific requirements before configuring a solution that ultilised Quobyte's Data Centre File System.

**Advantage:**
In addition to having S3 connectivity, the Quobyte system also afforded the JASMIN administrators the ability to scale their storage capacity (currently at 42PB) & performance linearly while providing ease of managment.

**01** 24/7 PHONE SUPPORT

**02** ONSITE WARRANTY SUPPORT

**03** NEXT DAY AND SAME DAY BUSINESS SERVICES

**04** GLOBAL WARRANTY COVERAGE AVAILABLE

### NVIDIA® DGX™ SYSTEMS
As the most accredited NVIDIA Elite Partner, Boston are pleased to offer NVIDIA DGX systems available for purchase, lease and testing along with training from the Deep Learning Institute.

### SUPERMICRO® WITH NVIDIA® TESLA® V100
As NVIDIA Tesla Partner of the Year, Boston are pleased to offer a comprehensive catalogue of products that are fully compatible with the full feature-set of NVIDIA Volta GPUs.

### IBM POWER AC922
The world's only server enabling NVIDIA NVLink between CPUs and GPUs, the AC922 delivers earlier prototypes and yields faster ROI.
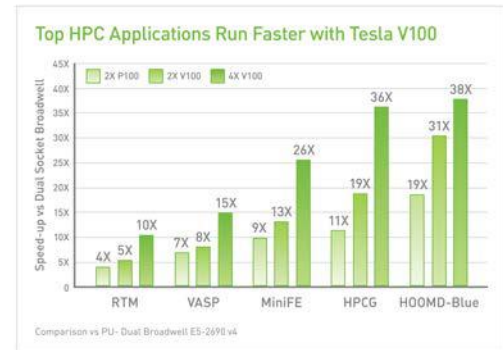
GPU SOLUTIONS

# NVIDIA® DGX™ SYSTEMS

NVIDIA® DGX™ Systems are built on the new, revolutionary NVIDIA® Volta™ GPU platform. Combined with innovative GPU-optimised software and simplified management, these fully integrated solutions deliver ground-break-ing performance and results.

NVIDIA® DGX™ Systems are designed to give data scientists the most powerful tools for AI exploration—tools that allow you to go from your desk to the datacentre to the cloud from the second you switch your system on.

GPU SOLUTIONS

# SUPERMICRO® WITH NVIDIA® TESLA® V100

Frequently first to market with solutions that utilise the latest generation of GPUs from NVIDIA®, Supermicro have a comprehensive range that enables the progression of AI from start-up sized organisations all the way to large HPC clusters.

Each solution also offers fantastic energy efficiency with many PSUs running at Titanium (96%) level. Boston are pleased to offer these solutions on their own or as part of a fully-coordinated cluster design and installation .

## KEY FEATURES

The NVIDIA® Tesla® V100 Tensor Core is the most advanced datacentre GPU ever built to accelerate AI, HPC, and graphics. It's powered by NVIDIA Volta architecture, comes in 16 and 32GB configurations, and offers the performance of up to 100 CPUs in a single GPU.



Top HPC Applications Run Faster with Tesla V100

## THE CASE FOR GPU POWERED HPC:

HPC datacentres need to support the ever-growing computing demands of scientists and researchers while staying within a tight budget. The old approach of deploying lots of commodity compute nodes substantially increases costs without proportionally increasing datacentres performance.

With over 550 HPC applications accelerated—including all of the top 15 —all HPC customers can now get a dramatic throughput boost for their workloads, while also saving money.

| DGX-1™ | DGX STATION™ | DGX-2™ | ANNA VOLTA | ANNA X16 | SYS-1019GP-TT |
|---|---|---|---|---|---|
| The essential instrument for AI research, designed to accelerate your datacentre and streamline your deep learning workflow. Experiment faster, train larger models, and get insights starting on day one. | Built on the same software stack that powers NVIDIA® DGX-1™, the NVIDIA® DGX Station™ is the only personal supercomputer for leading-edge AI development from your desk to the datacentre. | Break through the barriers to AI speed and scale with NVIDIA® DGX-2™, the first 2 petaFLOPS system that engages 16 fully interconnected GPUs for 10X the deep learning performance. | Housing up to 4x NVIDIA® Tesla® V100 SXM cards, but available part-populated, this dense yet compact server utilises NVIDIA® NVLINK™ for to 300 GB/s GPU to GPU communication. | Designed as a cloud platform for AI and HPC, combining 16 NVIDIA® Tesla® V100 32GB SXM3 GPUs connected via NVLink and NVSwitch to work as a unified 2 PetaFlop accelerator to deliver huge computational power. | A single socket server for your HPC cluster, with Intel® Xeon® Scalable CPUs, 6 Hot-swap 2.5" SAS/SATA drive bays along with support for up to two NVIDIA® Tesla® P100/P40/ P4/M60 GPUs. |

GPU SOLUTIONS

# IBM POWER AC922

# STORAGE SOLUTIONS

The IBM Power System AC922 is perfect for powering Enterprise AI initiatives throughout their lifecycle. With an optimised hardware and software stack from IBM, the AC922 delivers earlier prototypes and yields faster ROI than it's competitors. Coherent system and GPU memory capabilities enable faster processing of large models with fewer IO transfers. Power 9 architecture too enables direct NVLink 2.0 access from the processor to up to six NVIDIA V100 GPU accelerators – up to 5.6x* of the bandwidth provided by x86-based servers.

POWER9

SUPERMICRO

## STORAGE SOLUTIONS FOR HPC

### HYPER-CONVERGED, VIRTUALISED & CLOUD
We can provision all-flash, hyper-converged appliances at a fraction of the cost of our competitors, as well as a broad range of virtualised storage solutions.

### SCALE-OUT
Scale-out parallel file systems and high performance computing storage systems.

### NETWORK ATTACHED STORAGE (NAS)
A range of network attached file systems and scale out NAS solutions with all the features your enterprise would need.

### BURST BUFFER STORAGE
High bandwidth, low latency, all-flash tier for I/O intensive applications on your cluster.

### STORAGE CLUSTER MANAGEMENT
Automate and control your storage infrastructure under a single management interface.

### CUSTOM SOLUTIONS
Have a specific storage solution in mind? Get in touch and we will tailor make your solution to fit your exact requirements.

## KEY FEATURES

- The Power AC922 is uniquely capable of supporting larger models and data sets, by accessing system from GPU-based processes and algorithms, without PCIe bottlenecks.
- The Power AC922 delivers up to 3.8x** the AI performance, vs similarly-configured x86-based systems.

- The Power AC922 is the only server capable of delivering I/O performance between CPUs and GPUs, supporting the massive throughput required for HPC, deep learning and AI workloads.
- The Power AC922 includes a variety of next-generation I/O architectures, including: PCIe Gen4, CAPI 2.0, OpenCAPI and NVIDIA NVLINK.

| AC922 8335-GTC | AC922 8335-GTW |
|---|---|
| This air-cooled system features 2x POWER9 CPUs (16 or 20 cores) with NVLink 2.0 Technology: a link with up to 5.6X the performance (150 GB/sec) to each NVIDIA V100 with NVLink GPU. | This water-cooled system features 2x POWER9 CPUs (18 or 22 cores) with NVLink 2.0 Technology: a link with up to 5.6X the performance (100 GB/sec) to each NVIDIA V100 with NVLink GPU. |

* 1 5.6x more I/O bandwidth – tested results are based on IBM Internal Measurements running the CUDA H2D Bandwidth Test Hardware: Power AC922; 32 cores (2 x 16c chips), POWER9 with NVLink 2.0; 2.25 GHz, 1024 GB memory, 4xTesla V100 GPU; Ubuntu 16.04. S822LC for HPC; 20 cores (2 x 10c chips), POWER8 with NVLink; 2.86 GHz, 512 GB memory, Tesla P100 GPU Competitive HW: 2x Xeon E5-2640 v4; 20 cores (2 x 10c chips) / 40 threads; Intel Xeon E5-2640 v4; 2.4 GHz; 1024 GB memory, 4xTesla V100 GPU, Ubuntu 16.04
** Enables faster insights – 3.8x speedup based on comparing an AC922 with an Intel Xeon E5-2640 v4; 2.4 GHz; 1024 GB memory, 4xTesla V100 GPU running 1000 iterations of Enlarged GoogleNet model (mini-batch size=5) on Enlarged Imagenet Dataset (2240x2240) on Caffe

CLOUD SOLUTION

# MULTI-VERTICAL PRINT... PRIVATE CLOUD

## VSCALER
### A Private Cloud Appliance for any workload

AI & Deep Learning · Scientific Research · Broadcast & Media · Automotive

With vScaler, you can deploy your on-premise private cloud in minutes, add cloud-based services and applications, scale to public and hybrid cloud environments on-demand, all under a single management portal. vScaler simplifies datacentre infrastructure by integrating HPC, Big Data and Cloud resources into a converged platform that is capable of running any application at any scale. The platform enables agile sysadmin teams to quickly deploy scalable, production-ready private cloud environments, big data analytical platforms and/or traditional HPC clusters all under one management portal.

## KEY FEATURES

Finance, Government, Broadcast & Media, Scientific Research, Oil & Gas, Manufacturing and Bioinformatics are simply a few of the verticals that can leverage the vScaler cloud platform and services.

**70%** More cost effective than leading cloud providers

**40%** Faster than native OpenStack in LiINPACK Performance tests

**6x** Performance boost thanks to high performance vNICs

## USE-CASES

vScaler finely tuned cluster management software enables you to deploy clusters or clouds of any scale either on premise, in the cloud or across both in a hybrid model. Build and configure:

· HPC Clusters.
· Hadoop Clusters (Hortonworks or Cloudera)
· Clouds (Openstack)
· Parallel file systems (Lustre or BeeGFS)

## CASE STUDY

| FLEXIBLE INFRASTRUCTURE | SCALE-OUT STORAGE | UP TO 200 VMS PER APPLIANCE | CHALLENGE | SOLUTION | ADVANTAGE |
|---|---|---|---|---|---|
| Deploy what you need when you need it, mix & match different technologies and applications | Featuring a parallel file system for high performance IO. | Run up to 200VMs in a single Hyper-Converged 2U appliance | NxAARK's primary objective was to find a cost-effective cloud model that would enable it to quickly offer Hosted Enterprise Cloud to its customers at a competitive price point. | The vScaler HCI modular solution delivered a fully integrated, multi-location platform, hosted on-premise. The flexibility & fast learning curve aligned with NxAARK's build as-you-grow approach & enabled them to 'Go-Live' within weeks. | By deploying an out of the box, repeatable solution, NxAARK saved on the cost of setting up a dedicated team to build a cloud platform from scratch as well as time & effort to re-invent technology upgrades & enhancements. |

HYPERSCALE STORAGE PROCESSING

# GPU ALTERNATIVE TO RAID

**TRADITIONAL DATACENTRE** ≫ **NYRIAD-POWERED DATACENTRE**



◼ - COMPUTE    ◼ - STORAGE

The Boston Igloo Nebari, developed in partnership with NYRIAD features NSULATE, a Linux block device that functions as a software-defined alternative to RAID for configuring fast, reliable, larger scale storage solutions.

This solution utilises modern GPUs to perform storage-processing operations and can be configured as an alternative to RAID in the same environments that RAID solutions are typically used, in addition to enabling many new processing solutions.

NYRIAD    SUPERMICRO

## KEY FEATURES

### NSULATE

- Enables real-time hyperscale erasure coding up to 255 parity
- Cryptographic checksums and real-time corruption recovery

- Create highly parallel arrays with hundreds of devices
- High performance, even with massive degradation
- Compatible with all Linux filesystems and applications

## USE-CASES

**HPC**    **CLOUD**    **BIG DATA**

## WHY WOULD YOU USE A VIDEO CARD TO RUN A FILE SYSTEM?

| STORAGE DENSITY | PERFORMANCE | ENERGY USAGE | CHALLENGE | SOLUTION | ADVANTAGE |
|---|---|---|---|---|---|
| | | | The rapid advances in GPU computing performance has presented new challenges for storage technology. It is increasingly difficult to move data to and from storage fast enough to keep up with the processing speed of GPUs. | NSULATE had to meet several requirements to be a practical solution for hyperscale storage. Including being highly resilient to failure, extremely high in performance and scalability and provably able to detect and recover from nearly all data corruption. | NSULATE excels at real-time, high parity erasure coding, adding these capabilities automatically to file systems that do not have these features. |

SCALE-OUT

# PARALLEL STORAGE SOLUTION

## BENCHMARKS

### METADATA OPERATIONS

BeeGFS® was designed for extreme scalability. In a testbed with 20 servers and up to 640 client processes (32x the number of metadata servers), BeeGFS® delivers a sustained file creation rate of more than 500,000 creates per second, making it possible to create one billion files in as little time as about 30 minutes.
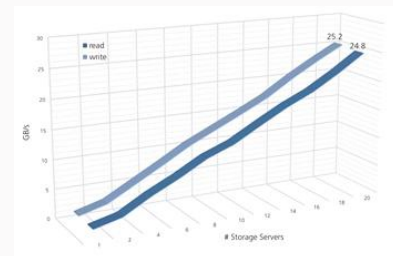
### THROUGHPUT SCALABILITY

In the same testbed system with 20 servers - each equipped with a single node local performance of 1332 MB/s (write) and 1317 MB/s (read) - and 160 client processes, BeeGFS® a sustained throughput of 25 GB/s - which is 94.7 percent of the maximum theoretical local write and 94.1 percent of the maximum theoretical local read throughput.

BeeGFS® is a pure software solution for scale-out parallel network-accessible storage, developed with a strong focus on performance and designed for very easy installation and management.

The flexibility, robustness, and outstanding performance of BeeGFS allows customers to increase productivity by delivering results faster and by enabling new data analysis methods that were not possible without the advantages of BeeGFS.
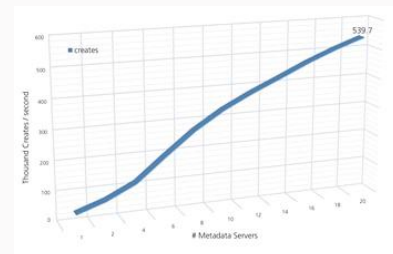
## KEY FEATURES

- Highly scalable parallel storage
- Uses a native Linux kernel module, so no kernel patching required
- Storage Pools enable transparent storage tiering

- Optional graphical interface for administration and monitoring systems
- Built in support for data replication through Buddy Mirroring to maximise data availability.

## USE-CASES

Create Parallel File Systems on demand and on the fly: BeeGFS® on-demand allows the creation of a BeeGFS® on a set of nodes with one single command line. Possible use cases for the tool are manyfold, a few include setting up a dedicated parallel file system for a cluster job across the compute nodes used for the job, cloud computing or for fast and easy temporary setups for testing purposes.

## CASE STUDY

| PROVEN BY CUSTOMERS WORLD-WIDE: | CHALLENGE | SOLUTION | ADVANTAGE |
|---|---|---|---|
| BeeGFS® is used all around the globe to provide extremely fast access to storage systems of all kinds and sizes, from small scale up to enterprise-class systems with thousands of hosts and including some of the fastest supercomputers in the world. | BeeGFS® needed to develop and deliver a high performance system with a very high fault tolerance. | The easy-to-install, parallel filesystem was perfect for the I/O intensive workloads required. As a result, the customer was able to transparently spread user data across multiple servers. | The customer benefited from the holistic approach of this fully customised turnkey solution. |

SCALE-OUT

# OBJECT STORAGE SOLUTION

90% of the data in the world has been generated in the last 2 years. Enterprises faced with designing storage infrastructure to handle this explosive growth can no longer rely on the scalability and accessibility of legacy file system technology. They need universal access, space efficiency, and highly available, robust architectures with the freedom to upgrade hardware as new technology becomes available. Caringo Swarm™ provides massively scalable, self-managing storage that unifies data silos and simultaneously handles mixed-use cases with a single deployment so you can extract the value of your data.

ANY DEVICE ACCESS    PUBLIC OR PRIVATE NETWORK    CONSOLIDATE OR DISTRIBUTE TO/FROM REMOTE SITES

Caringo Swarm
HASSLE-FREE, LIMITLESS STORAGE

HPC COMPUTE CLUSTER    LUSTRE OR GLUSTERFS    INSTRUMENTS & DEVICES    HOME DIRECTORIES    HADOOP

caringo    SUPERMICRO

## KEY FEATURES

- Built-in features to protect from accidental deletion
- 100% availability with no downtime upgrades
- Guaranteed data-integrity while continuously evolving hardware

- Choice of hardware, not locked into one vendor
- Single or multi-site deployment
- Simple, secure filetransfer with no storage silos

## USE-CASES

### LOWER STORAGE TCO

- Reduces CAPEX & OPEX
- Reduces strain on primary storage

### ELIMINATES RISK

- 100% business continuity
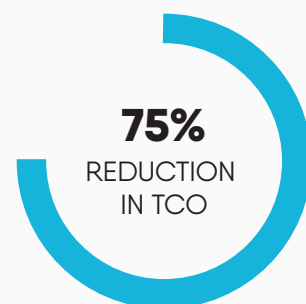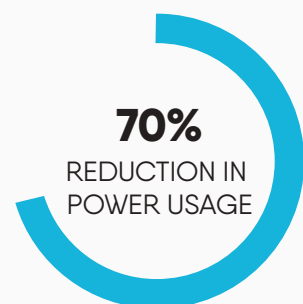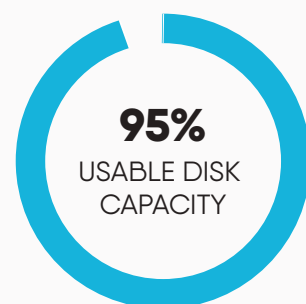- Enables regulatory compliance

### IMPROVES PRODUCTIVITY

- Manages & organises data at scale
- Unifies data silos

## CASE STUDY

### CARINGO SWARM BY THE NUMBERS

**95%** USABLE DISK CAPACITY

**70%** REDUCTION IN POWER USAGE

**75%** REDUCTION IN TCO

### CHALLENGE

A university needed an object storage platform; open source solutions were discarded because of their potential complexity in system management & tuning, while traditional storage systems were not cost competitive & lacked the flexibility needed.

### SOLUTION

Today, with the right technology, object storage can be adopted by organisations of all sizes. Caringo does exactly that, it has the right solution for allowing IT organisations to start with a small initial investment & grow as data & applications require.

### ADVANTAGE

Chosen for the ease of use and the affordable price point in both hardware, storage and networking as well as the software; alongside the expertise that came as part of the project.

SCALE-OUT

# FILE AND OBJECT STORAGE SOLUTION



Quobyte facilitates the entire HPC workflow no more silos, no more tedious and complex capacity planning, and better economics thanks to operational efficiency at scale.

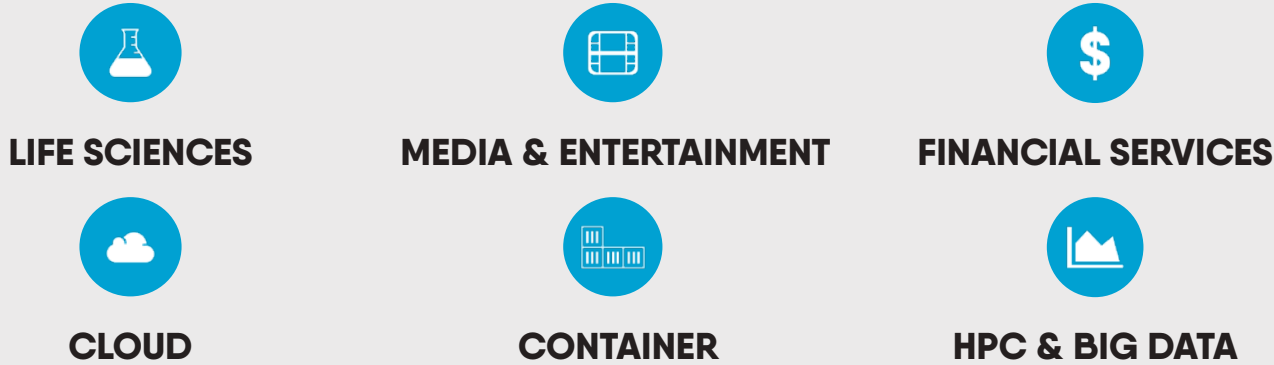Quobyte is a software defined storage solution with the capability to create multiple volumes,

providing both file and object storage. This makes Quobyte a highly versatile file system that can be utilised in several diverse scenarios, ranging from a standard Linux NFS server to S3 storage for cloud based solutions.



## KEY FEATURES

- Facilitates the entire HPC workflow – no more silos, no more tedious and complex capacity planning, and better economics thanks to operational efficiency at scale.
- High-performance – the distributed parallel file system delivers all the power you need

- Massively scalable – start with a few drives and scale linearly to hundreds of PBs and beyond
- All interfaces – access data through native clients for Linux, Windows, and macOS or use S3, NFS, SMB, and Hadoop

## USE-CASES

**LIFE SCIENCES**

**MEDIA & ENTERTAINMENT**

**FINANCIAL SERVICES**

**CLOUD**

**CONTAINER**

**HPC & BIG DATA**

## CASE STUDY

| QUOBYTE | | | | CHALLENGE | SOLUTION | ADVANTAGE |
|---|---|---|---|---|---|---|
| File, Block & Object Workloads within a Single System | Distributed File System | Linear Scalability & Performance | Data Safety, Redundancy Availablility, & Intergrity in software | A Datacentre File System was needed to manage the JASMIN Phase 4 super-data-cluster with an initial managed capacity of 42PB (petabytes). An average of 1–3PB of data is processed every day, which is expected to expand to 300PB by 2022. | Quobyte's Datacentre File System gives the JASMIN data facility the ability to unify their file, block, and object storage datasets in a centralised environment consisting of 11,500 cores on 600 nodes. | In addition to having S3 connectivity, the Quobyte system also afforded the JASMIN administrators the ability to scale their storage capacity (currently at 42PB) & performance linearly while providing ease of managment. |

SCALE-OUT

# NVME/BURST BUFFER SOLUTION

Solving one of the biggest challenges in computing today, Excelero enables low latency & high-performance access to NVMe volumes, delivered to remote clients with minimum processor overhead. This solution is available in disaggregated or converged configurations, and employs Mellanox technologies to keep processor overhead and latency low, whilst bandwidth is plentiful, making it a perfect fit for a myriad of different use cases. NVMesh is a Software-Defined Block Storage solution that features Elastic NVMe, a distributed block layer that allows unmodified applications to utilise pooled NVMe storage devices across a network at local speeds and latencies.
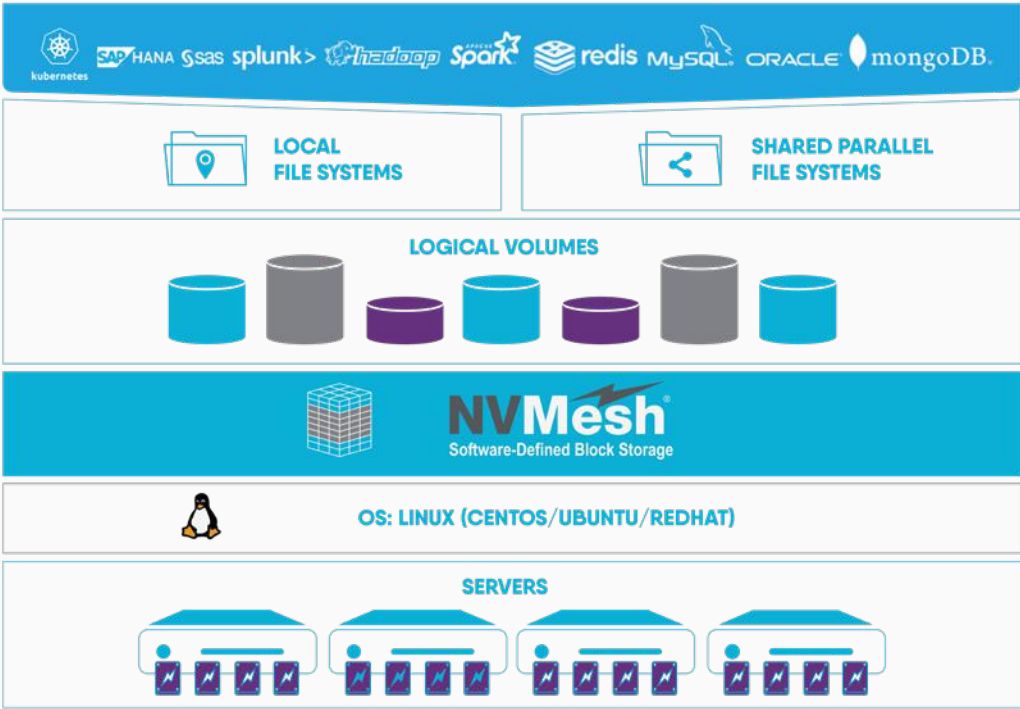


## KEY FEATURES

NVMESH

- 100% Server SAN
- Elastic NVME pools storage across a network at local speeds and latencies
- 0% CPU enables 100% converged infrastructure
- Virtual array

- Scalable
- Flexible
- Efficient
- Easy to use, manage & monitor
- Optimised for performance

## USE-CASES

**HIGH-FRAMERATE STORAGE FOR ANY-K VIDEO POST PRODUCTION**

**LOCAL BURST BUFFER**

**DATABASES**

**CONTAINERS**

## CASE STUDY

| SCALE & PERFORMANCE | EFFICIENCY | FLEXIBILITY | CHALLENGE | SOLUTION | ADVANTAGE |
|---|---|---|---|---|---|
| Local performance accross the network. Predictable application performance. Smart insights in utilisation. | Maximise the utilisation of your flash media. Reduce your capacity overhead. Easily manage & monitor. | Utilise any hardware. Use existing network infrastructure. Choose from multiple redundancy options. | InstaDeep™ decided to build an "AI as a Service" offering. Their key requirements for the data centre infrastructure were that it needed to be scalable, flexible and highly efficient to deliver high ROI. | InstaDeep chose Excelero's NVMesh® on Boston Flash-IO Talyn storage to provide GPUs with access to a scalable pool of high-performance NVMe to ensure full utilisation of the GPU processing power. | Excelero's NVMesh eliminates any compromise between performance & practicality, & allows GPU optimised servers to access scalable, high performance NVMe flash storage pools as if they were local flash, resulting in higher ROI & faster time to results. |

CLUSTER MANAGEMENT

# BRIGHT COMPUTING

Bright Computing automates the process of building and managing Linux clusters in your datacentre, the cloud, and at the edge. With Bright, you can:

- Provision 3 to 30,000+ nodes from bare metal in minutes
- Repurpose servers to accommodate fluctuating workloads on the fly

- Diagnose and resolve problems and performance issues quickly
- Deploy software updates across the entire cluster effortlessly
- Extend your on-premises environment to the cloud dynamically
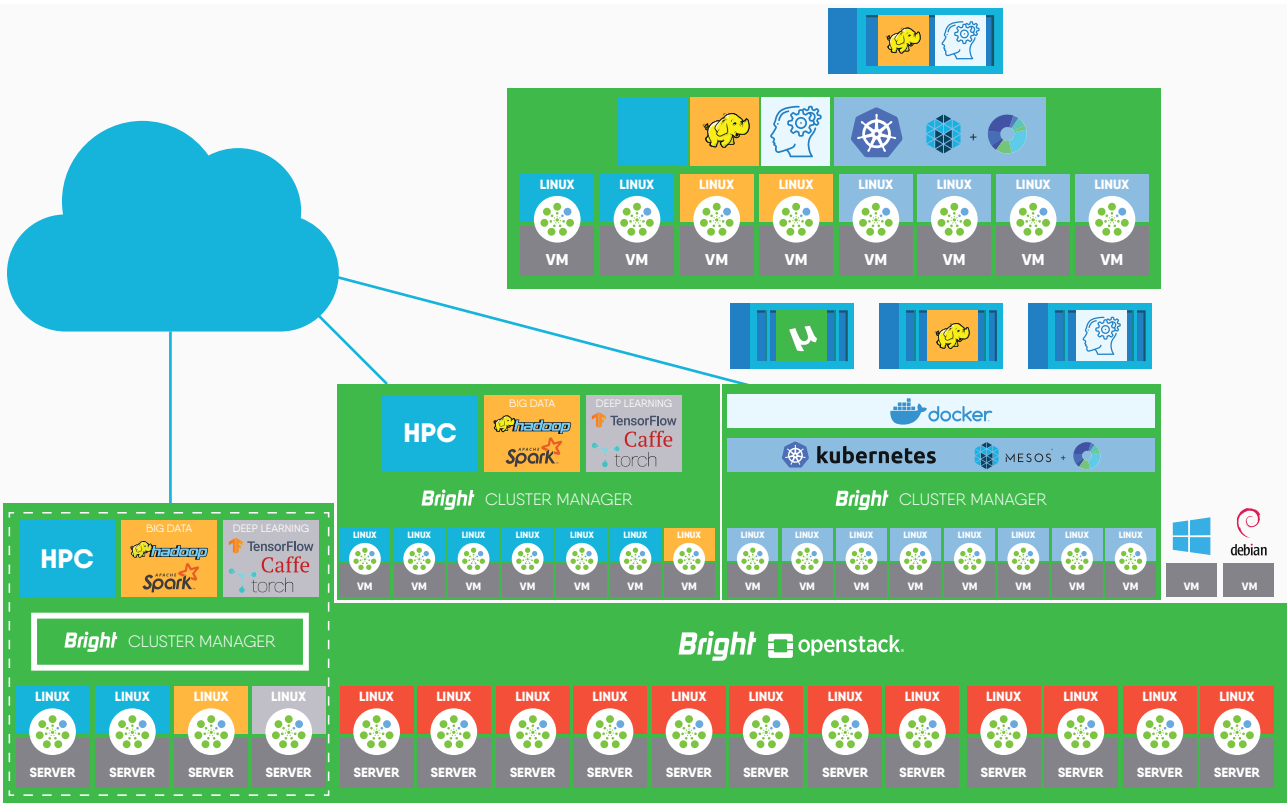
## KEY FEATURES

- Complete, Deploy, provision, monitor, manage, and scale HPC, Big Data, and OpenStack
- Elastic - Allocate compute resources dynamically, including into the cloud
- Intuitive - Easy to install, learn, and use

- Powerful - Manage complexity and ensure scalability of clusters and clouds
- Productive - Maximise throughput, minimise effort, optimise resources
- Proven - Join hundreds of customers who have already standardised on Bright

## USE-CASES

**HPC CLUSTERS**

**EDGE COMPUTING**

**CLUSTERS ON DEMAND**

**DEEP LEARNING & AI**

**MANAGE CONTAINERS**

**CLOUD BURSTING**

## CASE STUDY

| VERTICAL INDUSTRIES | | | | | CHALLENGE | SOLUTION | ADVANTAGE |
|---|---|---|---|---|---|---|---|
| Manufacturing and Energy | Life Sciences | Financial Services | Government | Academic and Research | MSU is recognised for it's research prominence, and expenditures typically exceed $100m annually. The IT Centre Research Cyberinfrastructure group was charged with building an affordable, scalable & easy to administer shared computing resource. | The Boston-built cluster sits on-premise, using virtualisation and cloud computing to provide a flexible & affordable solution, using Bright Cluster Manager to reconfigure nodes quickly and reliably. | The fully configured cluster was first soak-tested using Bright Computing at Boston Labs, this greatly reduced the installation time on-site and meant the system was ready to use almost from day 1. |

# BOSTON IMMERSED COMPUTING



Boston, in partnership with Asperitas, are making sustainable data centres possible anywhere they are needed, to facilitate emerging digital technologies, from high density compute to cloud. Immersed Computing® is a concept based on efficient liquid immersion technology and the basis for a unique infrastructure solution: the AIC24.

The AIC24 is the only immersion plug & play solution based on natural convection.
Immersed Computing® is enabling optimised and even customised server designs for record breaking high density platforms (CPU and GPU) and solutions for HPC without the datacentre complexity.
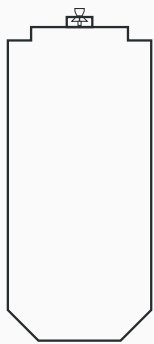
## 1. HIGH DENSITY SERVER SOLUTIONS OPTIMISED FOR IMMERSION
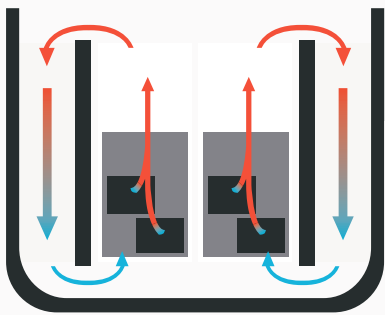
## 2. UNIQUE IMMERSED COMPUTING SOLUTION BASED ON NATURAL CONVECTION
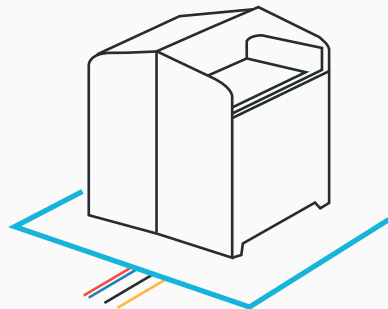
## 3. SIMPLIFIED MODULAR HIGH DENSITY DATA CENTRE DESIGN

## KEY FEATURES

### SUSTAINABILITY

- Circular design
- CO2 ↓ upto 40%
- 98% energy reuse

### FLEXIBILITY

- Modular system
- Plug & Play
- Anywhere

### EFFICIENCY

- TCO ↓ upto 40%
- Availability ↑
- Density ↑ x5

## USE-CASES

The opportunities for immersion computing in the datacentre are endless; with deployments already in HPC, Cloud, Enterprise, Life Sciences and Media & Entertainment – more and more businesses are recognising the value in what immersed computing offers.

## CASE STUDY

### INDUSTRY VALUE CHAIN EXAMPLE

| CLOUD VALUE CHAIN | SINGLE TENANT | HOSTING (IAAS/PAAS) | PRIVATE IN COLOCATION | SAAS IN COLOCATION |
|---|---|---|---|---|
| ENERGY EFFICIENCY | | | | |
| SOFTWARE EFFICIENCY | | | | |
| IT HARDWARE | | | | |
| DC OPERATIONS | | | | |
| DC FACILITIES | | | | |
| DC BUILD | | | | |

### CHALLENGE

The Department for Applied Bioinformatics in conjunction with the LOEWE Centre for Translational Biodiversity Genomics (LOEWE-TBG) was looking for a solution, which was highly energy efficient, sustainable and reliable.

### SOLUTION

With the latest AMD EPYC™ processors, the Asperitas Immersed Computing® solution does not require any raised floor, CRAC units or Chillers, drastically saving customer's CAPEX and OPEX.

### ADVANTAGE

Cost effective solution, able to support latest technologies & extremely low maintenance cost. Ability to extend and upgrade with only little effort.

# "LEADING-EDGE
# DISRUPTIVE TECHNOLOGY
# INTEGRATORS"

### vSCALER
PRIVATE CLOUD

### FLASH-IO TALYN
BURST BUFFER SOLUTION

### DGX-2™
GPU ACCELERATED SOLUTION

## SPEAK TO US ABOUT THE **NVIDIA® DEEP LEARNING INSTITUTE**

AMD
Partner Program
ELITE

NVIDIA
ELITE
PARTNER

intel | Technology
Provider
**Platinum 2019**

## WHY CHOOSE US?

**TAILOR-MADE SOLUTIONS**
Boston has the knowledge and expertise to tailor your ideal solution.

**BOSTON HPC LABS**
Remotely test and benchmark your technologies.

**LEADING-EDGE TECHNOLOGY**
Boston's R&D Labs facility offer the latest technology first.

**STORAGE**

**WORKSTATIONS**

**SERVERS**

**NETWORKING**

**SOLUTIONS**

**CLOUD SERVICES**

# NETWORKING

Thanks to our close partnership with Mellanox, a leading end-to-end network solution manufacturer, we have a vast array of solutions to suit your HPC requirements.

## NETWORKING

### SOFTWARE DEFINED NETWORKING (SDN) AND CUMULUS® LINUX®
SDN changes the fundamental way that networks work, the core concept of SDN is decoupling the two roles that networks must perform; the control plane and the data plane. The addition of Cumulus Linux means customers have the option to choose the leading Linux NOS on the market.

### INFINIBAND SWITCHES AND ADAPTERS
Mellanox combine industry standard InfiniBand technology, supporting up to 200Gb/s in their adapter and switch products, with integrated InfiniBand Router and InfiniBand to Ethernet gateways, Mellanox switches provide scalable fabric for powering the world's largest and fastest high-performance computing systems and next generation datacentres.

### ETHERNET SWITCHES AND ADAPTERS
Mellanox Ethernet products enable users to benefit from far more scalable, lower latency, and virtualised fabric with lower overall fabric costs and power consumption, greater efficiencies, and more simplified management than traditional Ethernet fabrics.

### CABLING
The Mellanox LinkX product family of cables and transceivers provides the industry's most complete line of 10, 25, 40, 50, 100 and 200Gb/s interconnect products. They are often used to link top-of-rack switches downwards to servers, storage & appliances and upwards in switch-to-switch applications. Products are available in both Ethernet and InfiniBand protocols and SFP & QSFP form factors.

NETWORKING SOLUTION

# MELLANOX QUANTUM™ HDR SWITCH QM8700

NETWORKING SOLUTION

# MELLANOX SPECTRUM SWITCH SN2010

Built with Mellanox's Quantum™ InfiniBand switch device, the QM8700 provides up to forty 200Gb/s ports, with full bi-directional bandwidth per port. The QM8700 is the world's smartest network switch, designed to enable in-network computing through the Co-Design SHARP (Scalable Hierarchical Aggregation and Reduction Protocol) technology. The Quantum™ switch improves the performance of selected collective operations by processing the data as it traverses the network, eliminating the need to send data multiple times between end-points.

The SN2010 switch is the ideal top of rack (ToR) solution for hyper-converged and storage deployments. Packed with 18 ports of 10/25GbE and 4 splittable ports of 40/100GbE, the SN2010 can deliver up to 1.7Tb/s aggregate throughput.

With its optimisation for RoCE, full buffer utilisation, and zero packet loss combined into a small form factor, the SN2010 is driving forward the world's most innovative datacentre infrastructures.

## KEY FEATURES:

**PERFORMANCE**
40 X HDR 200Gb/s ports in a 1U switch
80 X HDR100 100Gb/s ports (using splitter cables)
16Tb/s aggregate switch throughput
Sub-90ns switch latency

**ADVANCED DESIGN**
Adaptive Routing
Congestion Control
Collective offloads (SHARP)
VL mapping (VL2VL)

**OPTIMISED DESIGN**
1+1 Redundant & hot-swappable power
N+1 Redundant & hot-swappable fans
80 Plus Gold and Energy Star certified power supplies
x86 ComEx Broadwell CPU

## KEY FEATURES:

**THROUGHPUT**
1.7Tb/s
2.52B packets-per-second

**LOWEST POWER**
57W (ATIS)
power consumption

**HALF-WIDTH FORM FACTOR**

**LOWEST LATENCY**
300nsec for 100GbE port-to-port
Consistently low latency regardless of packet size, or L2 vs L3 forwarding
Side by side configuration
Small port count

## INTERESTED?

The QM8700 together with Mellanox ConnectX®-6 adapter card support HDR100. By utilising two pairs of two lanes per port, the QM8700 can support up to 80 ports of 100G to create the densest TOR switch available in the market.

## INTERESTED?

Boston are official distributors of Mellanox and are fantastically placed to advice you on the best solution for your requirements.

Our team are ready to answer any questions you may have about networking.

# HPC NODES

# HPC NODES

## COMPUTE



### 6029TP-HTR (PER NODE)

| | |
|---|---|
| **Form Factor:** | 2U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 16x DDR-4 Registered ECC 2666MHz |
| **Expansion:** | 2x PCIe 3.0 x16 LP Slots |
| **Disk:** | 3x 3.5" Hot swap SATA3 bays |

### 2029TP-HTR (PER NODE)

| | |
|---|---|
| **Form Factor:** | 2U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 16x DDR-4 Registered ECC 2666MHz |
| **Expansion:** | 2x PCIe 3.0 x16 LP Slots |
| **Disk:** | 6x 2.5" Hot swap SATA3 bays |

### 1029TP-DTR (PER NODE)

| | |
|---|---|
| **Form Factor:** | 1U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 16x DDR-4 Registered ECC 2666MHz |
| **Expansion:** | 2x PCIe 3.0 x16 LP Slots |
| **Disk:** | 4x 2.5" Hot swap SATA3 bays |

## AMD EPYC™

### AS-1114S-WTRT

| | |
|---|---|
| **Form Factor:** | 1U Rackmount |
| **CPU:** | Single AMD EPYC™ 7002 Series Processor |
| **Memory:** | 8x DDR-4 Registered ECC 3200MHz |
| **Expansion:** | 2x PCIe 4.0 x16 FHFL Slots, 1x PCIe 4.0 x16 LP Slot* |
| **Disk:** | 10x 2.5" Hot swap SATA3 |

*\*On some PCI-E 4.0 devices, it will only run at PCI-E 3.0 speed.*

**COMING SOON**

### AS -2124BT-HTR (PER NODE)

| | |
|---|---|
| **Form Factor:** | 2U Rackmount |
| **CPU:** | Dual AMD EPYC™ 7002 Series Processors |
| **Memory:** | 16x DDR-4 Registered ECC 3200MHz |
| **Network:** | 2x PCIe 4.0 x16 LP Slots |
| **Disk:** | 6x 2.5" Hot swap SATA3 |

### AS-2014TP-HTR  (PER NODE)

| | |
|---|---|
| **Form Factor:** | 2U Rackmount |
| **CPU:** | Single AMD EPYC™ 7002 Series Processor |
| **Memory:** | 8x DDR-4 Registered ECC 3200MHz |
| **Expansion:** | 2x PCIe 4.0 x16 LP Slots |
| **Disk:** | 3x 3.5" Hot swap SATA3 |

## GPU



### ANNA

| | |
|---|---|
| **Form Factor:** | 1U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **GPU:** | 4x NVIDIA® Tesla® V100 SXM2 |
| **Memory:** | 12x DDR-4 Registered ECC 2666MHz |
| **Disk:** | 4x 2.5"  SATA/SAS bays |

### ANNA XL

| | |
|---|---|
| **Form Factor:** | 4U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **GPU:** | 8x NVIDIA® Tesla® V100 PCIe |
| **Memory:** | 24x DDR-4 Registered ECC 2666MHz |
| **Disk:** | 24x 3.5" Hot swap SATA/SAS bays |

### ANNA XL X16

| | |
|---|---|
| **Form Factor:** | 10U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **GPU:** | 16x NVIDIA® Tesla® V100 SXM2 |
| **Memory:** | 24x DDR-4 Registered ECC 2933MHz |
| **Disk:** | 6x 2.5" SATA drives,16x U.2 NVMe SSD 2x M.2 NVMe drives |

## GPU

### DGX STATION™

| | |
|---|---|
| **Form Factor:** | Tower |
| **CPU:** | Dual Intel® Xeon® E5-2698 v4 |
| **GPU:** | 4x NVIDIA® Tesla® V100 SXM2 |
| **Memory:** | 256GB DDR-4 Registered ECC |
| **Disk:** | 3x 1.92TB SSD RAID 0 |

### DGX-1™

| | |
|---|---|
| **Form Factor:** | 3U Rackmount |
| **CPU:** | Dual Intel® Xeon® E5-2698 v4 |
| **GPU:** | 8x NVIDIA® Tesla® V100 SXM2 |
| **Memory:** | 512GB DDR-4 Registered ECC |
| **Disk:** | 4x 1.92TB SSD RAID 0 |

### DGX-2™

| | |
|---|---|
| **Form Factor:** | 10U Rackmount |
| **CPU:** | Dual Intel® Xeon® Platinum 8168 |
| **GPU:** | 16x NVIDIA® Tesla® V100 PCIe |
| **Memory:** | 1.5TB DDR-4 Registered ECC |
| **Disk:** | 8x 3.84TB NVMe SSD |

## STORAGE



### 1029P-NR32R

| | |
|---|---|
| **Form Factor:** | 1U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 24x DDR-4 Registered ECC 2666MHz |
| **Network:** | 2x 10GBase-T LAN ports via Intel® X540 |
| **Disk:** | 32x Hot swap NVMe Ruler SSDs |

### 1029UZ-TN20R25M

| | |
|---|---|
| **Form Factor:** | 1U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 24x DDR-4 Registered ECC 2666MHz |
| **Network:** | 2x 25GbE SFP28 ports |
| **Disk:** | 20x 2.5" Hot swap drive bays |

### 6029P-E1CR12H

| | |
|---|---|
| **Form Factor:** | 2U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 16x DDR-4 Registered ECC 2666MHz |
| **Network:** | 2x 10GBase-T LAN ports with Intel X722 + PHY Intel X557 |
| **Disk:** | 12x 3.5" Hot swap SATA/SAS bays |

## STORAGE

### 6049P-E1CR36H

| | |
|---|---|
| **Form Factor:** | 4U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 16x DDR-4 Registered ECC 2666MHz |
| **Network:** | 2x 10GBase-T LAN ports with Intel X722 + PHY Intel X557 |
| **Disk:** | 36x 3.5" Hot swap SATA/SAS bays |

### 6049P-E1CR60H

| | |
|---|---|
| **Form Factor:** | 4U Rackmount |
| **CPU:** | Dual Intel® Xeon® Scalable Family |
| **Memory:** | 24x DDR-4 Registered ECC 2666MHz |
| **Network:** | 2-ports of 25GbE SFP2 |
| **Disk:** | 60x 3.5" Hot swap SATA/SAS bays |

### 6048R-E1CR90L

| | |
|---|---|
| **Form Factor:** | 4U Rackmount |
| **CPU:** | Dual Intel® Xeon® E5-2600 |
| **Memory:** | 8x DDR-4 Registered ECC 2400MHz |
| **Network:** | Intel i350 Dual Port Gigabit Ethernet |
| **Disk:** | 90x 3.5" Hot swap SATA/SAS |

# TECH TALK

## BOSTON FINDS (AMD) ROME HAVING ALREADY VISITED NAPLES

### WHERE WE HAVE COME FROM AND WHERE WE ARE GOING TO WITH AMD?

AMD shook up the CPU industry when it released the first EPYC generation 7000 series (codename Naples) of processors and is looking to do the same with its second generation of EPYC 7002 series (codenamed Rome).

Some people in the marketplace will look at the first generation and have deja vu - but to understand the potential userbase, we must look at the history of AMD and their main competitor Intel. AMD brought out the Athlon 64 and Opteron 64bit (K8) processors in 2003 which changed the processor landscape at the time; going from 32bit to 64bit, and bringing dual core processors, at a time when neither were available from Intel. Unfortunately for AMD this lead did not last, and Intel brought out the Intel Core microarchitecture. This started in late 2005/early 2006 with the Pentium M, Core Solo and Core Duo, Core 2 followed in late 2006/early 2007, introducing 64-bit extensions.

### FEATURE COMPARISON BETWEEN THE 7001 AND 7002 SERIES

| | SOCKET SP3 EPYC NAPLES CPU (2017) | SOCKET SP3 EPYC ROME CPU (2019) |
|---|---|---|
| CPU TDP | 120W – **180W** | 120W – **225W** |
| SOCKET | SP3 (LGA 4094) | SP3 (LGA 4094) |
| NO OF SOCKETS SUPPORTED | 1S and 2S | 1S and 2S |
| NO OF CORES | Up to **32C** / **64** Threads | Up to **64C** / **128** Threads |
| CACHE | L2: 512KB per core (16MB total) L3: **64MB shared cache** **(8MB per 4 cores)** | L2: 512KB per core (16MB total) L3: **256MB shared cache** **(16MB per 4 cores)** |
| MEMORY | 8 DDR4 Channels at **2666MHz** | 8 DDR4 Channels at **3200MHz** |
| MEMORY CAPACITY | Up to **2TB** | Up to **4TB** |
| PCIₑ LANES | 128 PCIe **Gen3.0** | 128 PCIe **Gen4.0 + 2 PCIe 2.0 lanes** |
| LITHOGRAPHY PROCESS | **14nm** | **7nm** |

# TECH TALK

AMD responded with the release of Phenom, but arguably didn't match the pace of Intel... that was until June 2017 when AMD EPYC 7000 (Naples) series launched. EPYC provided a lower price point and higher performance level than the Intel equivalent in some workloads.

AMD's EPYC 7002 Series to EPYC generation (Rome) builds upon the progress made by AMD in this market, and looks to blow away any reminders of the previous pattern seen. This is a big part of why Rome is so important, since it proves that AMD can do multiple generations of performance and technology enhancements while showing a full roadmap for the future.

### CLOSER LOOK AT NEW CCX DESIGN DESIGNED FOR ZEN2

At first glance you would assume that all the dies seen below are CPU cores with SOC, as was the case with the 1st generation, however, as of the second generation the CPU cores are now split.

### 1ST GENERATION OF EPYC WITH THE HEAT SPREADER REMOVED:
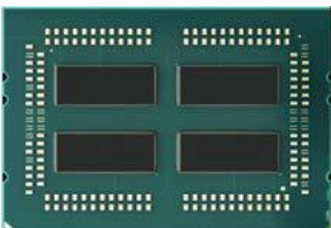
In the first-generation Naples utilised four System on chip (SoC) dies each with 8 cores and these are interconnected with AMD infinity fabric.
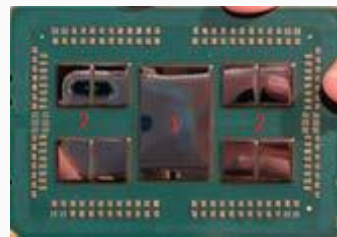
### "ROME" –
### UNIQUE 2 SOCKET PLATFORM: 3-XGMI-2, UP TO 160 PCIE LANES, UP TO 32 SATA
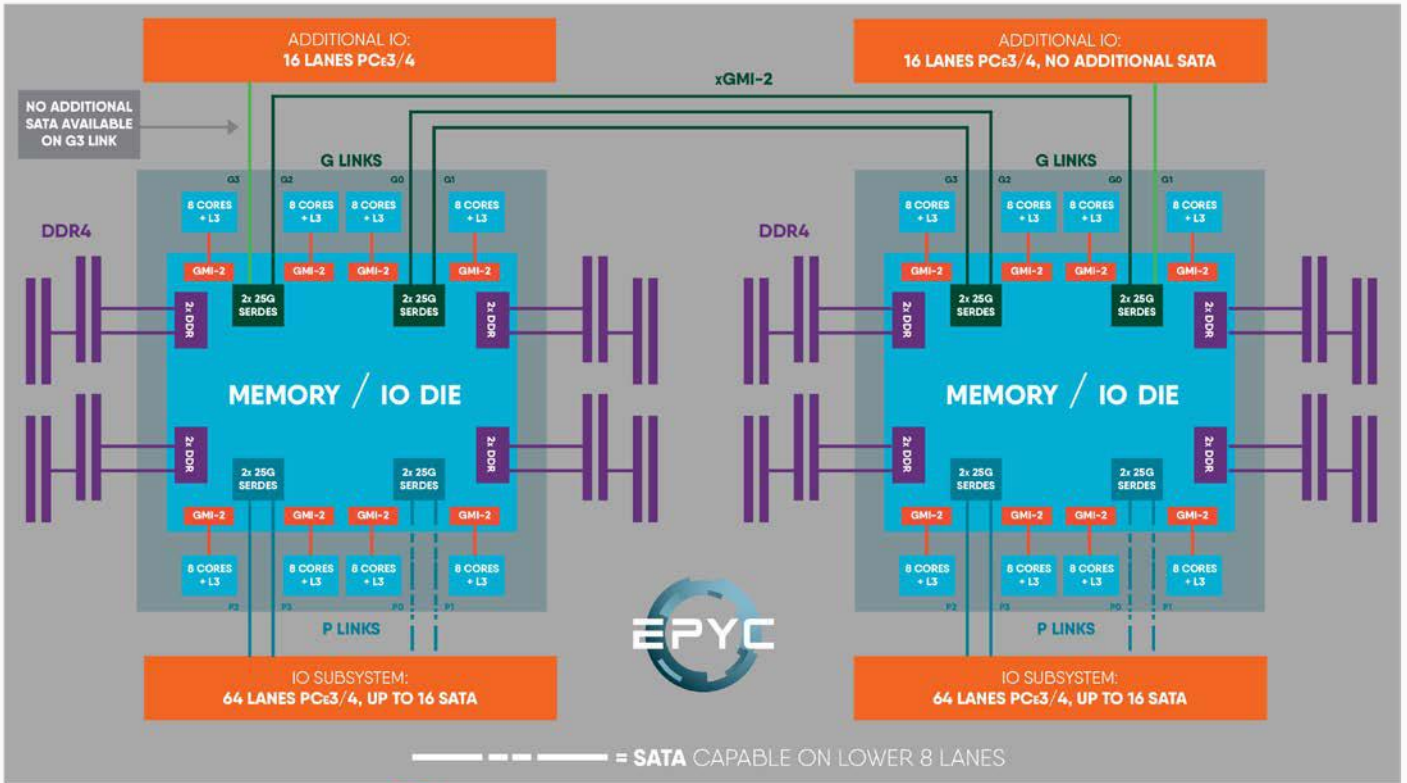


**128 PCIE DUAL SOCKET DESIGNS**

## 2ND GENERATION OF EPYC WITH THE HEAT SPREADER REMOVED:

In the second-generation, things have changed from the original EPYC design. While previously you had SOC`s, this has now been separated out for improved scale out architecture - AMD calls this the CCX and memory I/O die. The largest die shown is a memory/IO die (#1) which provided all the functionality of the SOC apart from the CPU cores. These are the smaller CCX to the left and right of the big Memory/IO die in the middle (#2).

## MEMORY/IO DIE ARCHITECTURE DIAGRAM

This is a CPU topology diagram for a 2 socket AMD 7002 series system and, as can be seen, there are many changes from the previous 7001 generation of EPYC.

Memory control is now centralised per socket since AMD have done away with the SoC design. EYPC 7001 had 4 NUMA nodes per socket for 8 in total with a dual processor system. NUMA topology can have a negative performance impact if the application is not NUMA aware. More importantly for AMD, Intel's competing Xeon scalable products had two NUMA nodes per socket which is what most operating systems and applications will have been optimised for. AMD's EYPC 7002 product now supports 2 NUMA nodes per socket which is a resulting benefit of the Memory / IO Die shown above. The CPU cores themselves are separate and

each CCX is 8 cores connected to 16MB of on-die Level 3 cache.

PCIe is now implemented with Up to 64 + 32 + 1 lanes per socket and up to 16 PCIe lanes can be enumerated as SATA per socket.

Inter-communication between CPU sockets is handled by Inter-chip Global Memory Interconnect 2 (XGMI-2) links. There are 4 XGMI-2 links shown in the diagram above: three of these are used by the CPU's to "speak" to each other and the remaining 1 link per socket "feeds" 16 PCI-E lanes. Only one socket uses the 4th XGMI-2 link to break out into SATA.

## SUMMARY OF KEY CPU EXTENSIONS WITH GENERATION 1 AND 2 OF EPYC

While each CPU generation is all about cores, threads and frequency, when you dig deeper there are further enhancements that come in the form of extensions. The purpose of an extension within the CPU architecture is to extend the functionality outside of the default instruction set and the functionality will be customised towards solving the targeted workload more efficiently than can be done otherwise.

Improving efficiency is not the only use case and in some scenarios extensions will enable the CPU to process workloads that it could not otherwise do, or to enable a feature that the CPU could not previously support.

AVX2 (Advanced Vector Extensions 2) is the second generation of AVX and both generations are supported on AMD EPYC. AVX additionally performs the instructions introduced with SSE extensions. SSE extensions were, in turn, an

evolution of MMX, which could do two 32-bit floating point operations in one instruction. SSE doubles this to 4 and AVX offers up to 8 FP 32bit operations. AVX2 brought further improvements expanding the number of vector integer SSE and AVX instructions to 256 bits. While all these mathematical operations may sound impressive, they may not mean much to the average user other than a box to tick.

In the CPU industry, increasing frequency, cores and threads can only get you so much performance as there are inherent issues in scaling these factors, frequency in particular. Adding these workload extensions allows for more operations per instruction. Each CPU core will have a certain level of IPC (instructions per cycle), sometimes known as Instructions per clock; this is what ultimately determines what is commonly known to most as single-threaded performance (and then scaled up with cores/threads to give multi-threaded performance).

However, it is not all down to having these supported in hardware, since these instructions must be called or written into the software to see any benefit. Not only must Operating Systems have support for the extension, applications will need this support too. Many well-known software packages already support AVX2 and, whilst this is not an exhaustive list, here are some examples of supported software platforms (the version listed or higher) that support AVX2 (as of the date of this article): Cinema 4D, Pixar's Renderman 22.5 (onwards), Google's Tensorflow, Ansys Fluent 18.1 and Gaussian 16. Previous software versions did not necessarily implement the AVX2 extension and this is partially why performance can change significantly between program versions.

## CPU EXTENSIONS ENABLED WITHIN NAPLES AND NOW ROME

- Simultaneous Multi-Threading
- AVX, BMI, F16C, FMA, SSE2/3/4
- AVX2, BMI2, MOVBE, RDRAND
- ADX, RDSEED
- Virtualization Acceleration IOMMU
- Advanced Interrupt Virtualization
- Nested Virtualization

## CPU EXTENSIONS ENABLED WITHIN ROME

- CLWB = Cache Line Write Back
- GMET = Guest Mode Execution Trap
- PQE/PQM Cache Allocation Technology and Platform QoS Monitoring
- VIOMMU = Virtualized IOMMU
- WBINVD = Cache Line Write Back and Invalidate
- WBOINVD = Cache Line Write Back without Invalidate
- X2APIC

## SECURITY FEATURES AND VULNERABILITIES

AMD have several technologies and updates baked into EPYC generation 1 and 2: Secure Encrypted Virtualization (SEV), AMD Secure Memory Encryption (SME), AMD Secure Encrypted Virtualization-Encrypted State (SEV-ES) and mitigation of side channel attacks. All with the express goal to improve overall system security whilst ensuring dependable system performance.

# TECH TALK

### UPGRADE PATH AND CONSIDERATIONS GOING FROM NAPLES TOO ROME WITH SUPERMICRO

Generally, between incremental generations of CPUs which are 'Pin to Pin compatible', a BIOS update to support the new CPU will be necessary. For the Rome upgrade from Naples there are several caveats that are covered in this section.

Naples to Rome is 'pin to pin compatible' meaning that you can install a EPYC Generation 2 CPU into a EPYC Generation 1 CPU era motherboard mechanically however, due to the updated BIOS ROM size increasing between the two generations, memory speed increasing and with the arrival of PCI-E generation 4.0, various hardware changes are required.

Check out our full guide of supported features for EPYC Gen 2 and the respective Supermicro motherboard generations below.

If you have any concern or confusion about your upgrade path, system architecture, your required features or what revision of hardware you currently have, we are happy to help. When purchasing systems with us here at Boston, let us know your requirements and we can direct you towards the system that matches.

If you would like more information or design and architecture help around AMD's 2nd Generation of EPYC processors, then we'd be keen to hear from you. You can get in touch with us today at **sales@boston.co.uk** or call us on **01727 876 100**.

### SUPERMICRO MOTHERBOARD REVISIONS – EPYC 2ND GENERATION SUPPORT

| | H11 (Rev 1.x) | H11 (Rev 2.x) | H12 |
|---|---|---|---|
| **EPYC Generation 1 (Naples)** | Supported | Supported | Supported |
| **EPYC Generation 2 (Rome)** | Not Supported | Supported | Supported |
| **PCI-E 4.0 Support** | Not Supported | Not Supported | Supported |
| **DDR4 3200Mhz** | Not Supported | Supported | Supported |
| **DDR4 2933Mhz** | Not Supported | Supported | Supported |

# TECH TALK

## INSTADEEP™ POWERS AI AS A SERVICE WITH SHARED NVME

### NVMESH™ FEEDS UNLIMITED STREAMS OF DATA TO GPU–BASED SYSTEMS WITH LOCAL PERFORMANCE
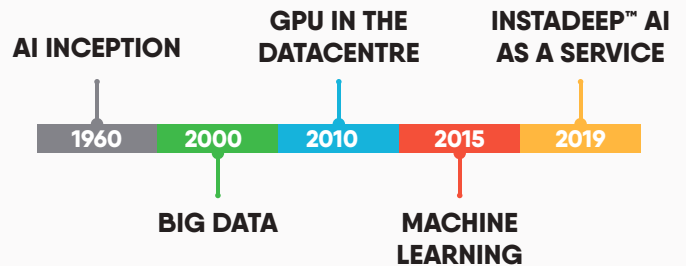
InstaDeep™ Ltd. is a global AI innovator, headquartered in London with offices in Paris, Tunis, Nairobi and Lagos. The company delivers AI-powered decision-making systems for customers across a wide array of industries, including Logistics, Manufacturing, Energy and Mobility. With expertise in both machine intelligence research and concrete business deployments, InstaDeep™ provides competitive advantages to customers in an AI-first world: the company's AI solutions allow enterprises to unlock data insights, realise value, increasing efficiency and speed across organisations.

### UNCAGING GPUS AND NVME PERFORMANCE FOR AI AND MACHINE LEARNING

AI and ML use has exploded over the past few years as four key technology evolutions have made it far easier to capture, store and process data into insights that can help enterprises outsmart the competition:

1) New sensor technologies have proliferated that capture images, temperature, heartrate, and more – adding even more data volumes.

2) Big Data analytics and Data Lakes for storing these massive volumes of data arose, so that teams could analyse and mine that data to turn it into valuable business or research insight.

3) The rise of powerful GPU technologies that lower the cost of massive compute on those data have made parallel processing faster and much more powerful.

4) Next-gen storage options such as NVMe flash media have swept the storage industry and are well-suited to these new computational engines, although they harken back in time to the days when direct attached storage (DAS) models were new. DAS is fast, but often underutilised and hence costly.

**AI INCEPTION** | **GPU IN THE DATACENTRE** | **INSTADEEP™ AI AS A SERVICE**

1960 — 2000 — 2010 — 2015 — 2019

**BIG DATA** | **MACHINE LEARNING**

### INSTADEEP™ AI AS A SERVICE

When InstaDeep™ decided to build an AI as a Service offering, it had a few key requirements in mind. The data centre infrastructure needed to scale modularly, as the company began this cutting-edge service with a few key customers but planned to quickly expand to its entire global customer base. The infrastructure had to be flexible to meet performance requirements for a wide range of workloads. Today the infrastructure is used by multiple scientists who run workloads for many different customers. Finally, for the service to be attractive for customers, and a sound business move for InstaDeep, the infrastructure needed to be highly efficient – the GPUs especially would need to deliver the highest return on investment (ROI).

InstaDeep chose the Boston Flash-IO Talyn storage with Excelero's NVMesh® to provide GPUs with access to a scalable pool of high-performance NVMe to ensure full utilisation of

the GPU processing power. The Talyn system included a 2U Boston Flash-IO Talyn server with Micron NVMe flash and Excelero NVMesh software that provides access to up to 100TB external high-performance storage. Leveraging the Mellanox 100GB Infiniband network cards in the DGX, the GPUs use the NVMe storage with local performance. The ability to choose any file system to run on NVMesh was an immense benefit. Early tests quickly showed that external NVMe storage with Excelero gives equal or better performance than local cache in the DGX.

### FEEDING THE GPU BEAST

The biggest advantage of modern GPU computing is also creating its biggest challenge: GPUs have an amazing appetite for data. Current GPUs can process up to 16GB of data per second. The latest NVIDIA® DGX-2™ system has as many as 16 GPUs, but by far not enough local storage. The DGX-1 has a theoretical limit of 7.8GB/s bandwidth, but with only 4 SATA SSDs it is limited to about 2.2GB/s. Theoretically it can process 2 million random IOPs but local storage only provides 400K IOPs. The latest NVIDIA® DGX-2™ has 30TB (8 x 3.84TB) local NVMe but is not optimised to use it efficiently. Other brand GPU servers typically feature few PCIe lanes for local flash (NVMe or other), meaning even the lowest latency option for these servers is a severe bottleneck or is simply too little capacity for the GPUs. Starving the GPUs with slow storage or wasting time copying data wastes expensive GPU resources and affects ROI.

### OPTIMAL APPROACH FOR SPECIFIC AI, ML USES

NVMe flash offers great benefits for specific AI use cases like training a machine learning model, and checkpoints. Machine learning involves two phases - training a model based on what is learned from the dataset and running the model. Training of a model is the most resource hungry stage. Hardware used for this phase, incorporating high-end GPUs or specialised system-on-chips (SoCs, is expensive to buy and operate so it should be always busy for best ROI.

### SUMMARY

The capability of GPUs and the rise of affordable compute power challenges IT teams to think at data centre scale – leveraging the ability to apply AI, ML and deep learning techniques to large data pools, while making sure the entire system is scale-out, highly performant, and efficient. The only storage that is fast enough to keep up with GPUs is local NVMe flash, since GPUs, networking and NVMe are all competing for valuable PCIe connectivity, one of them must compromise and settle for less. The Boston Flash-IO Talyn with Excelero NVMesh eliminates any compromise between performance and practicality, and allows GPU optimised servers to access scalable, high performance NVMe flash storage pools as if they were local flash. This technique ensures efficient use of both the GPUs themselves and the associated NVMe flash. The result is higher ROI, easier workflow management and faster time to results.

# BOSTON
Servers | Storage | Solutions
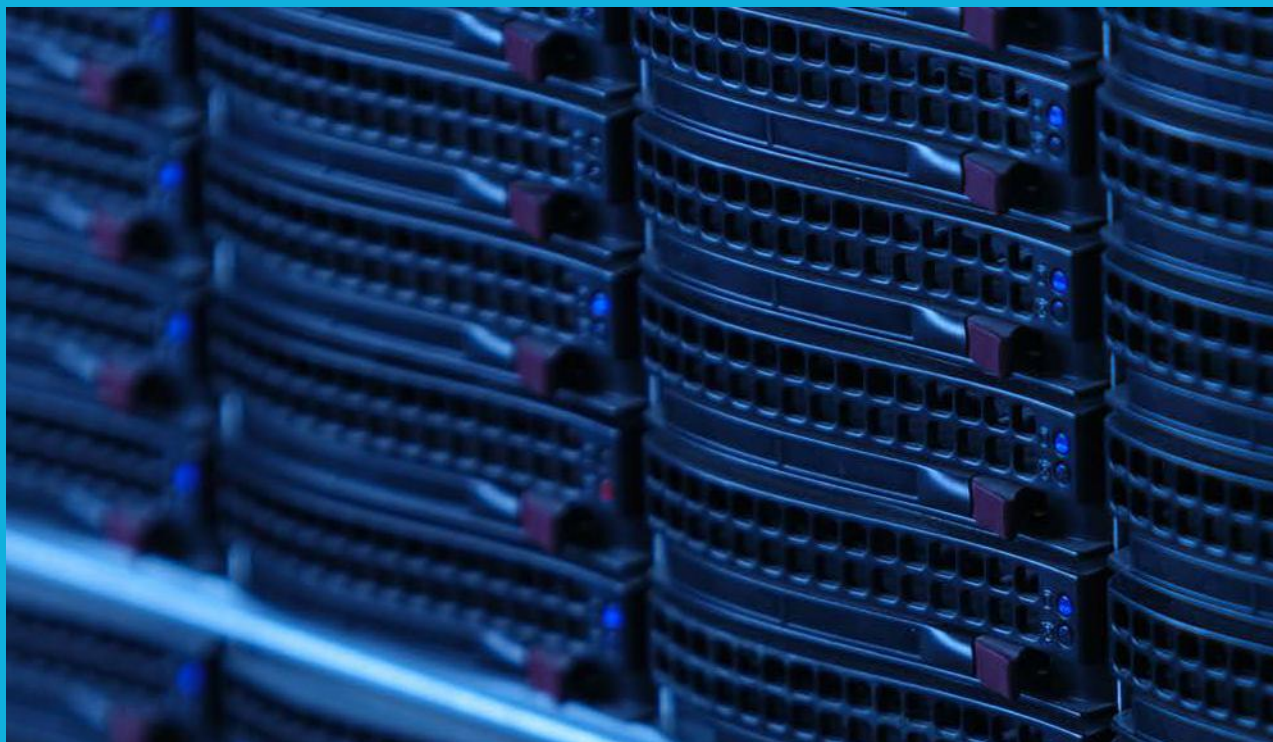


## UK (Headquarters)
+44 (0) 1727 876 100
sales@boston.co.uk
www.boston.co.uk

## INDIA
+91 22 5002 3262 (Mumbai office)
+91 80 4308 4000 (Bangalore office)
sales@bostonindia.in
www.bostonindia.in

## AFRICA
+27 826 486 485
sales@boston.co.uk
www.boston.co.uk

## GERMANY
+49 (0) 89 9090 1993
sales@boston-it.de
www.boston-it.de

## SWITZERLAND
+47 (0) 71 554 22 75
sales@boston-it.ch
www.boston-it.ch