Excelero

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL CHOOSES ELASTIC NVMe STORAGE TO POWER GPU SERVERS FOR MACHINE LEARNING AND AI

Reducing machine learning training time three-to-four days to under an hour[1]

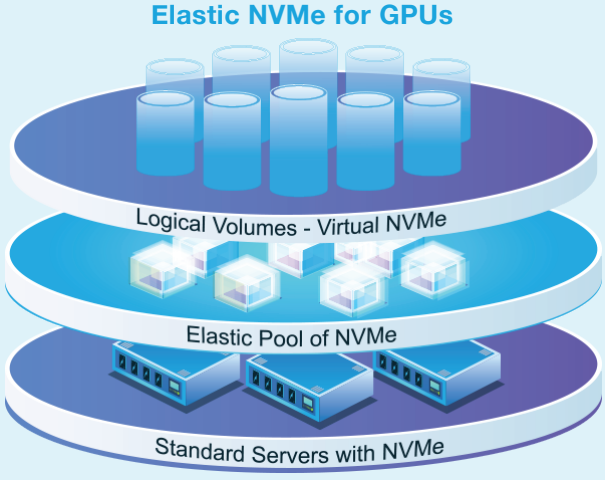
CASE STUDY

Science & Technology Facilities Council | BOSTON Servers | Storage | Solutions | Mellanox TECHNOLOGIES | NVIDIA

The Science and Technology Facilities Council (STFC) supports pioneering scientific and engineering research by over 1,700 academic researchers worldwide on space, environmental, materials and life sciences, nuclear physics and much more. Research involves a wide variety of data-rich analyses and machine learning on data from various beam-lines of the Diamond synchroton, cryo-electron microscopy, ISIS Neutron and Muon Source, RALSpace, the Centre for Environmental Analysis (CEDA), and other techniques. Computational workloads at STFC are massive: processing hundreds of TB's of data requiring both fast compute and fast storage.

To enable faster analyses, STFC often utilizes deep neural networks running on state of the art NVIDIA® DGX-2™ GPU computing systems, which were commissioned with the support of the Alan Turing Institute. To ensure full utilization of all the GPUs, STFC has now deployed Excelero's NVMesh® storage software on Boston Ltd. Flash-IO Talyn®. The elastic NVMe solution will provide GPUs with access to a scalable pool of high-performance NVMe, enabling machine learning workloads to process more data, much faster. Excelero's NVMesh delivers low-latency (5µs), high bandwidth distributed block storage for AI and ML workloads. NVMesh enables shared NVMe across any network and supports local or distributed file systems. GPU-based systems benefit from the performance of local NVMe flash with the convenience of centralized storage while avoiding proprietary hardware lock-in and maximizing the overall GPU ROI.

Benefits of Elastic NVMe for Machine Learning and AI:



- **Access remote NVMe at local speed**
- **Elastic NVMe storage infrastructure**
- **Share NVMe resources across multiple GPU servers**
- **Exceed the performance limits of local flash on GPU servers**
- **Eliminate the need to copy data locally**
- **Datasets can be larger than what can fit inside the GPU Server**
- **Full CPU offload on both ends**

[1]Source mention: some content in this document was kindly borrowed from several pages, blogs and press releases on the STFC website: https://stfc.ukri.org/

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL CHOOSES ELASTIC NVMe STORAGE TO POWER GPU SERVERS FOR MACHINE LEARNING AND AI

Reducing machine learning training time three-to-four days to under an hour

MACHINE LEARNING AND AI TECHNOLOGIES TO HELP SCIENTISTS ANALYZE VAST AMOUNTS OF DATA

The Science and Technology Facilities Council (STFC) is a UK government agency for scientific research. It is one of Europe's largest multidisciplinary research organizations supporting scientists and engineers worldwide. The agency funds and supports university-based research in areas such as particle physics, nuclear physics and astronomy, and provides access to world-leading, large-scale facilities and campuses to promote academic and industrial collaboration.

STFC's Scientific Computing Department (SCD) manages high performance computing (HPC) facilities, services and infrastructure that allows processing huge amounts of data. They provide the computational expertise, services and products that help the scientific community make vital discoveries and deliver progress. One of the better known STFC HPC initiatives is JASMIN, the UK's leading, and globally unique, environmental science supercomputer and data facility, jointly managed by SCD and CEDA, which is part of RAL Space.

Research to save the planet

Much of the research done at STFC is focused on one of the greatest challenges of our time: saving our planet. STFC's contribution comes from collecting and analyzing data about our planet by space satellites and ground based monitoring. Among others, STFC's RAL Space carries out an exciting range of world-class space research and technology development. CEDA supports researchers, via JASMIN and other facilities, with science ranging from sea surface temperatures to greenhouse gas concentrations and the movement of volcanic ash.

Source: https://stfc.ukri.org/

STFC's Scientific Machine Learning (SciML) Group was established with the aim of enabling scientists to analyze large amounts of data, with the group bringing machine learning and AI expertise. The user community served by the SciML Group and the Alan Turing Institute, in conjunction with the scientific community, routinely utilizes deep neural networks running on state of the art NVIDIA® DGX-2™ GPU computing systems located at the Scientific Data Centre at its Rutherford Appleton Laboratory site in Oxfordshire. As most of the analyses become image-centric, the use of GPU-based workstations needed to be extended to support the high throughput and low latency required for end-user response times.

Adding NVIDIA DGX-2 servers offered better computational support, yet lacked the storage performance required to leverage the full GPU processing capacity. High performance computing solutions provider, Boston Ltd., worked with STFC to evaluate all-flash arrays and open systems-based storage options, and commissioned a benchmark of Excelero's NVMesh for share NVMe Flash storage at local performance.

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL CHOOSES ELASTIC NVMe STORAGE TO POWER GPU SERVERS FOR MACHINE LEARNING AND AI

Reducing machine learning training time three-to-four days to under an hour

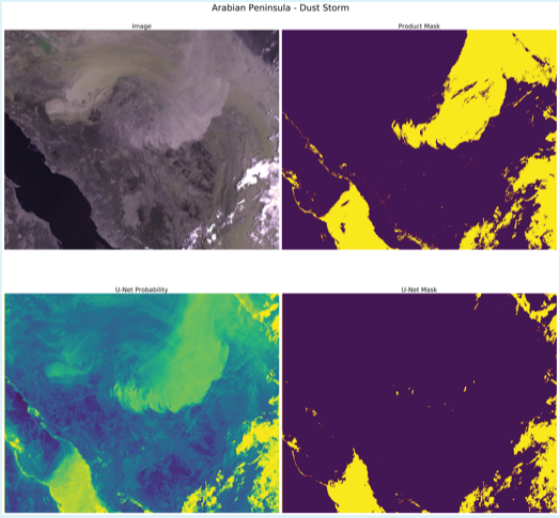
UNCAGING GPUs PERFORMANCE WITH NVMe FOR AI AND MACHINE LEARNING

Use of AI and ML-based systems use has exploded over the past few years as technology evolutions have made it far easier to capture, store and process data into insights:

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

• The rise of powerful GPU technologies that lower the cost of massive compute on those datasets has made parallel processing faster and much more powerful.

• 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 underutilized and hence costly.



Source: https://stfc.ukri.org/

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. NVIDIA's DGX-2 has 30TB (8 x 3.84TB) local NVMe but is not optimized to use it efficiently. Starving the GPUs with slow storage or wasting time copying data wastes expensive GPU resources and affects the ROI.

Fortunately, NVIDIA's DGX nodes also have massive network connectivity. They can ingest as much as 48GB/s of bandwidth via 4-8 x 100Gb ports – playing a key part in the solution: Excelero's NVMesh enables customers to maximize the utilization of their GPUs leveraging the massive network connectivity of the DGXs and the low-latency and high IOPs/BW benefits of NVMe in a distributed and linearly scalable architecture.

Excelero's NVMesh eliminates any compromise between performance and practicality, and allows GPU optimized 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 end result is higher ROI, easier workflow management and faster time to results.

When data scientists at STFC are training machine learning models, they literally process hundreds of terabytes of data and they need to do so in the shortest amount of time. A single training run, for example, for analyzing a few thousand satellite images to identify regions of clouds, which is also referred to as an epoch, is an important metric toward training time. Training machine learning models requires epochs to be run a number of times to find anomalies, detect noise etc. The faster the epochs are run, the more they can be obtained in a shorter time period. Leveraging elastic NVMe with GPUs drastically increases the number of epochs that can be run, and enables STFC data scientists to train more and better models.

SCIENCE AND TECHNOLOGY FACILITIES COUNCIL CHOOSES ELASTIC NVMe STORAGE TO POWER GPU SERVERS FOR MACHINE LEARNING AND AI

Reducing machine learning training time three-to-four days to under an hour

STFC MACHINE LEARNING ARCHITECTURE: ELASTIC NVMe FOR GPUs

STFC's storage architecture now includes two Boston Flash-IO Talyn® systems built on Supermicro® building blocks, networked via a Mellanox® 100G InfiniBand network to two NVIDIA DGX-2 computing systems, each with 16 NVIDIA 32GB V100 SXM modules.

With the BeeGFS file system providing a single name space to simplify management and virtualization, and the low latency and high throughput of its NVMesh system, STFC now has a GPU computing architecture where storage no longer presents a bottleneck, even with its complex research needs.

BENCHMARK: REDUCING MACHINE LEARNING TRAINING TIME FROM DAYS TO HOURS

Boston Ltd. worked with STFC to evaluate all-flash arrays and open systems-based storage options, then commissioned a benchmark of Excelero's NVMesh for share NVMe Flash storage at local performance.

Boston Ltd.'s benchmark results showed the proposed STFC architecture delivered an average latency of 70 microseconds – nearly one-quarter of the typical 250 microsecond latency of traditional controller-based enterprise storage when running NVIDIA validation tests on each NVIDIA DGX-2 system. The combined NVMesh and BeeGFS deployment therefore showed potential for meeting STFC's high throughput, low latency demands.

NVMesh features for GPU

- **NVMesh unifies remote NVMe devices into a logical block pool that performs the same as local NVMe flash**
- **NVMesh allows full utilization of the IOPs and bandwidth capabilities of NVMe drives across a network**
- **DGX-1 and 2 can use their massive network connectivity to access remote NVMe logical volumes, with redundancy if desired!**
  - **MUCH faster than local SATA SSDs**
  - **Larger shared pools than possible within the platform**
- **Other GPU optimized systems can access remote NVMe at local latencies and**

Operational since July 2019, STFC's storage architecture enabled its Scientific Machine Learning group to run training sets that formerly took three to four days, in under an hour.

Backed by STFC's new deployment, the user communities, including the Alan Turing Institute, are now able to carry out machine learning research functions across mulltiple fields, including environmental, material and life sciences, and astronomy.