REPORT REPRINT

Work smarter, work faster: Move the processing, not the data

MARCH 4 2019

By Tim Stammers

In the cloud world, the problem of data gravity is well known. But it also includes the performance-sapping micro-gravity suffered by IO data movements between storage and server processors. Pioneers of computational storage are sidestepping that problem by moving the processing to the data.

451 Research®

Summary

The concept behind 'in situ' processing or computational storage is simple: because it takes time and resources to move data, applications can be accelerated and systems made more efficient by abandoning the conventional approach of moving data to where the processing is done. Instead, move at least some of the processing to the data. The concept has also been called near-data processing or in-storage processing. Startups are pioneering the field, and one way to implement the concept is to allow application code to run on processors within flash drives. However, that is not the only approach they have taken. Target applications include but are not restricted to performance-sensitive applications such as machine learning, analytics and content delivery. Startups are pioneering the concept, and Samsung is currently the only major player in the sector. However, Lenovo, Micron, NetApp and Western Digital have already signaled their interest by joining a technical working group set up recently within the storage industry's cross-vendor body, the Storage Networking Industry Association (SNIA.) Other major suppliers have expressed interest in joining the group.

451 TAKE

We're sure that computational storage will become commonplace in at least one form or another. Over the last decade, the introduction of flash into enterprise storage systems has given the IT industry and the storage sector in particular something of a breathing space, because it has relieved infrastructure from the heavy performance constraints of disk. But this breather is showing signs of ending because new applications such as machine learning and analytics need even faster access to data than that provided by conventional storage systems – even those powered entirely by flash. This demand is set to be amplified by IoT or edge computing and the large volumes of data that will be generated in 5G networks. Computational storage looks to be an answer – and its benefits could be amplified by emerging storage-class memories (SCMs) such as Intel and Micron's 3D XPoint or Optane memory.

SNIA's working group

SNIA is tasked with developing vendor-neutral infrastructure architectures and standards. In November 2018 it announced the formation of a technical working group for computational storage. SNIA runs a total of 15 such groups covering other technical areas, and the computational storage group aims to promote device interoperability and to define interface standards for system deployment, provisioning, management and security. The group is co-chaired by NGD and SK Hynix. It says that over 20 companies are actively participating and include:

- Arm
- Eideticom
- Inspur
- Lenovo
- Micron Technology
- NetApp
- NGD Systems
- Nyriad
- Samsung Electronics
- Scaleflux

- SK Hynix
- Western Digital

Sidestepping even micro-gravity

Figuratively speaking, all data suffers from gravity, because it takes time and resources to move it from one location to another. At one end of the scale, petabyte-sized data being transferred into public clouds is simply too large and has too much gravity to be transported over a network, and must instead be shipped on physical media.

At the other end of the scale, far less gravity is involved when data is transferred in byte-level quantities at sub-system level; for example, between a flash drive and server memory, over an NVMe bus. This might be called micro-gravity and it significantly reduces overall application performance when it is suffered by millions of IOs occurring every second. Even in the current flash era, in many datacenters moving data between storage and processors takes more time and as much power as the processing itself, according to NGD.

Between those two extremes of PB-scale data transfers into clouds and individual IOs between storage and compute, there is edge computing. Edge computing is a macro-level version of in situ computing, because it involves the processing of IoT sensor data at the locations where it is generated, as an alternative to sending that data over a network for processing at a central location. Data gravity is the reason for doing that. Edge computing is also a strong candidate for a double implementation of in situ computing, because it often needs very high-performance processing. That makes it a strong candidate for the use of the products described in this report, resulting in a second, deeper implementation of the concept.

Storage-class memories are set to make computational storage even more attractive, because they will increase the potential boost to overall application performance, as result of their much greater write performance compared with flash.

A variety of approaches

Moving compute to the data can be done in multiple ways. One approach taken by three of the vendors profiled below – NGD Systems, Samsung and ScaleFlux – is to move processing into flash drives. The data still needs to be moved from the NAND flash chips to the processor in the drive, but that is completed using a Common Flash Interface (CFI), which has three to six times the bandwidth of the NVMe link that would be needed to move the same data into a server.

Another approach, taken by Eideticom, is to create a processor platform that does not provide any persistent data storage, but can access and process data from anywhere on an NVMe bus or an NVMe-oF network. This provides a processing resource that is quite separate to existing server compute resources, and is likely to suit tasks that are more compute-intensive than data-intensive. Nyriad has taken yet another route, by creating a storage system with unusual GPU controllers, bringing a new type of processing power into storage arrays.

Early players

451 Research is relatively early in its exploration of the topic of computational storage. Here we include a brief look at the companies in the field that we've spoken to. We'll be publishing a more detailed look at several of these companies and others in the coming months.

Eideticom

Founded in 2016 and based in Calgary, Canada, Eideticom is self-funded. Its founders and current leaders were previously engineering chiefs at chip designer PMC Sierra (now owned by Microsemi) and also founded RAD3, a communications specialist bought by PMC Sierra. Eideticom's product became GA late in 2018, and a number of companies are now trying out the technology. Those companies include two very large web-scale operators, alongside tier-one storage OEMs and organizations involved in high-performance computing, according to Eideticom.

The architecture of Eideticom's NoLoad product is unusual, even compared with other computational storage offerings. The company says it combines high performance with flexibility and disaggregates in situ compute and storage into independently scalable resources. Eideticom's NoLoad devices fit into PCIe-NVMe slots and are powered by FPGAs. Currently Eideticom is focusing on storage services, including data compression and de-duplication, RAID and erasure coding, in order to relieve the load on storage system controllers and hence boost their performance. Longer-term plans cover other areas such as analytics, machine learning and database acceleration.

The devices process data while transferring it from one set of NVMe flash drives to another. A key quality is the ability to complete these transfers in peer-to-peer mode, with no direct host CPU involvement. Using the NVMe-oF networked variant of NVMe, the NoLoad devices can also provide virtually the same level of performance when working with source and target drives located in different systems to the NoLoad devices themselves – hence allowing a very strong degree of flexibility and not just logical but even physically remote disaggregation of compute and storage resources.

NGD Systems

NGD was founded in 2013 and is based in Irvine, California. 451 Research has profiled the company and its product. NGD raised $28m from two VC rounds and has won three SBIR grants from the US National Science Foundation. Its co-founders have a strong track record in flash drive engineering and worked together at STEC, a successful pioneer of datacenter flash drives that was bought by Western Digital in 2013.

NGD shipped its first NVMe drives in 2017. In late 2018, it made a major product update when it began shipping versions of its cards powered by ASICs with embedded ARM cores, which it said reduces costs and improves scale compared with its previous devices that were powered by FPGAs. NGD isn't naming customers, but says the number of proofs of concept in its pipeline increased significantly when it launched its ASIC-powered drives. To our knowledge, NGD is the first to make this move into ASIC-powered computational storage.

Previously, NGD said its POCs included two projects at Microsoft, one involving image similarity and the other involving IoT or edge computing, investigation of the use of NGD's drive for encryption key and data stream management in content-distribution networks, development of autonomous cars, and in performance acceleration at a European online bookings company. NGD's drives have a second string to their bow, which is that when used as conventional, non-computational storage devices, they offer what NGD says is the industry's highest physical density, coupled with low power consumption.

Nyriad

Cambridge, New Zealand-based Nyriad's technology was originally developed for use with the Square Kilometre Array (SKA), an international project to build the world's largest radio telescope. The SKA says the telescope will generate more data traffic than the entire internet, with the first phase reportedly processing data at an enormous 160TB/s. Nyriad's founders developed the company's NSULATE technology to handle that workload while also protecting the precious data against hardware failures. NSULATE uses NVIDIA GPUs to complete erasure-coding (EC) calculations. Nyriad says this not only allows customers to set very deep EC parity calculations for high levels of data protection, but also provides very high performance. The company says that simultaneously with erasure coding, the GPUs can be used for other workloads such as machine learning and blockchain.

The company's partners include memory specialist Netlist, HPC systems vendor Boston Limited and HPC filesystem vendor ThinkParQ. Nyriad and Netlist have demonstrated a server-based system combining NSULATE with Netlist NVDIMMs, while Boston has developed an NSULATE-based system that also uses NVDIMMs, but from Micron. ThinkParQ sells the BeeGFS file system, which it describes as an alternative to Lustre, and has also developed a storage server that incorporates NSULATE. Oregon State University has adopted that system for use in computational biology.

Samsung

In October 2018, Samsung announced what it called a SmartSSD, which it said would increase speed and efficiency. No shipping date was given. A Xilinx FPGA powers the device and Xilinx said the SmartSSD accelerates storage services such as data compression, de-dupe and encryption, and compute functions such as machine learning inferencing and big-data analytics. At SuperComputing 2018, Bigstream, a supplier of software for accelerating a range of big-data analytics and ML applications, demonstrated its software working with Samsung's SmartSSDs and Apache Spark. The demo showed a performance boost of about three-fold to five-fold, and Bigstream said the boost scales further when more SmartSSDs are deployed. Samsung appears to have been exploring this field for a while because it gave a public and technical presentation on the topic in 2013.

ScaleFlux

Founded in 2014, ScaleFlux has raised over $40m in two funding rounds from investors that include Shunwei Capital, Xilinx and 'multiple tier-1 corporations.' The company's headquarters are in San Jose and has operations in China and EMEA. ScaleFlux's founders have strong track records in solid-state storage at companies that include flash drive controller specialist SandForce, PCIe flash pioneer Fusion-IO, Seagate, Mellanox and LSI.

ScaleFlux began shipping its computational storage product in 2017 and says hyperscalers and cloud operators are using its devices in transactional processing, analytics, gaming, digital payments, fraud prevention and ad tech. Named customers and partners include digital payments operator PhonePe, cloud services platform AliCloud and web-security integrator Global Dots.

ScaleFlux says its combination of software and hardware provides about two to four times faster performance compared with standard NVMe flash drives for data-driven applications like Aerospike, Hadoop and Apache HBase. As an SSD with the ability to process workloads in situ, ScaleFlux's system is similar to that of NGD or Samsung; however, it differs in having its flash translation layer (FTL) run on the host server.