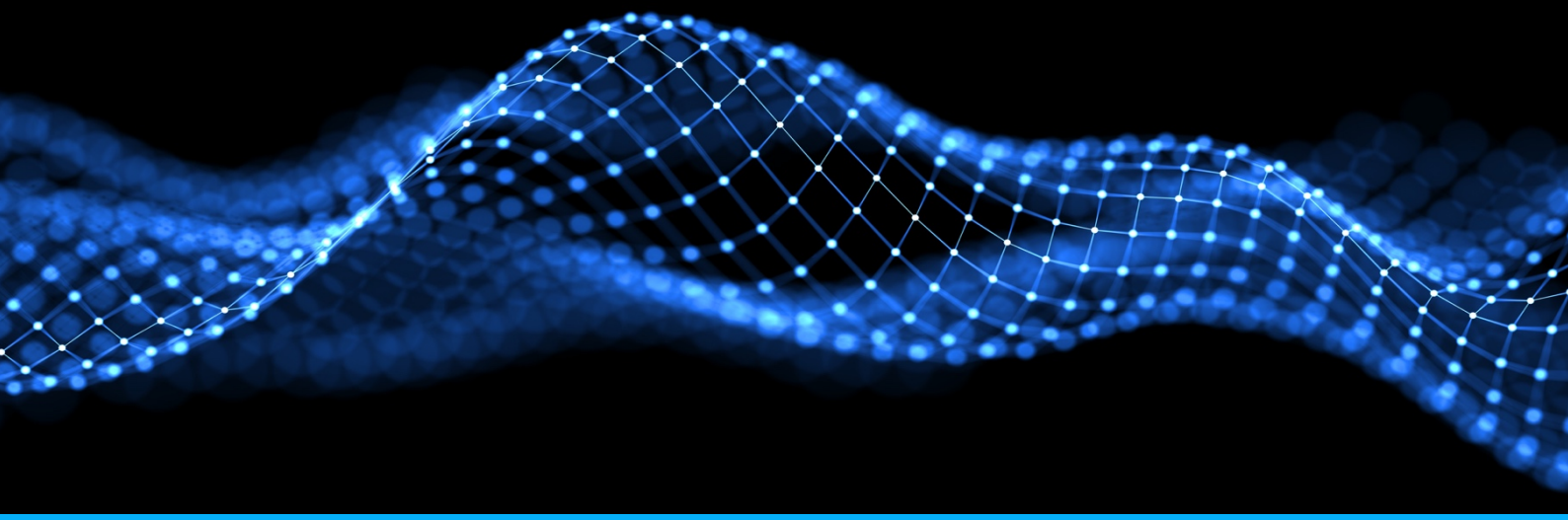
Vexata VX-100



PRODUCT BRIEF

The Architect's View

VX-100 is a new shared storage platform from Vexata Inc. The technology has been designed to exploit the benefits of NVMe media, dividing out the control and data paths of traditional I/O. By accelerating the data path using hardware components like FPGAs, Vexata has been able to develop a solution that adds a mere 5-10µs to I/O through the platform, fully exploiting the benefits of NVMe NAND and SCM technology.

Background

NAND-based NVMe SSDs offer significant performance improvements over traditional flash media. NVMe SSDs also include new solid-state persistent memory technologies such as Intel Optane. Both types of NVMe devices are pushing performance levels close to one million IOPS, tens of gigabytes of throughput and values as low as 10µs of latency.

In order to fully exploit these technologies, new storage architectures are needed that minimise as much as possible the data path between host and media. Traditional storage protocols in shared storage introduce too much latency into host I/O, which due to the capabilities of new media is an increasingly large part of the end-to-end response time experienced by host servers.

The NVMe protocol introduces features that optimise the I/O stack to persistent media. In addition, NVMe increases the amount of parallel I/O that can be performed to a single device by supporting up to 64K I/O request queues and 64 elements per queue. Storage arrays need to be capable of using this parallel capability, as a means of scaling out both capacity and front and back-end host performance.

Vexata VX-100

Vexata has developed a new storage architecture specifically to exploit the benefits of NVMe. The VX-OS (Vexata Operating System) is a combination of hardware and software that is divided into components which separate data and control paths. The data path is accelerated at the front-end of the architecture using FPGAs in a component called VX-Router. At the back-end, SSDs are built into a dedicated storage module called an ESM (Enterprise Storage Module). This houses four drives with a dedicated MIPS processor to handle local metadata and the offload of some intensive data management tasks.

Joining the two sides of the architecture is a scale-out redundant Ethernet midplane. This provides up to 256GB/s of internal bandwidth in the current hardware implementation, however none of the components is limited by design and can all be scaled independently.

Today Vexata has two offerings. The VX-100F, based on NAND flash, scales to 435TB and can deliver 7 million IOPS and 70GB/s of bandwidth at 220µs of latency. The VX-100M uses Intel Optane media, scales to 30TB and can deliver 80GB/s of bandwidth and 7 million IOPS at 40µs latency. Both products ship as a 6U array, with dual front-end controllers, dual redundant Ethernet midplanes and up to 16 ESMs.

Architectural Choices

Vexata has chosen to address the main issue seen with traditional dual-controller architectures, where data services in the data path cause an increase in host I/O response times. Specific design choices have enabled the platform to deliver more performance from the underlying media than has been previously possible.

Some of the changes involve using hardware acceleration, for example FPGAs in the VX-OS Router component and a MIPS processor in each ESM. Using some discrete hardware in this way enables the platform to offload traditionally compute intensive activities (such as snapshot consolidation and RAID rebuilds) and minimise or eliminate the impact on host latency.

Vexata has also chosen to initially support existing Fibre Channel environments, which won't deliver the lowest end-to-end latency compared to using new NVMe over Fabrics (NVMe-oF) solutions. However, the architecture is designed to cater for NVMe-oF in the future. This provides customers with an upgrade path to use existing infrastructure (e.g. storage area networks based on Fibre Channel) and upgrade to faster Ethernet networks when appropriate., without having to perform a "rip and replace" of networking infrastructure.

Market Positioning

The VX-100 platforms are targeted at existing enterprises that need to deploy a solution with high levels of throughput at consistent low latency. This could include accelerating traditional monolithic applications such as databases, but increasingly will tackle performance issues of ML/AI. Machine learning activities need high degrees of parallelism and consistent low latency. Support of Fibre Channel allows the technology to be introduced into existing environments, then progressively upgrade to take advantage of increased performance as it is needed.

Caveats

As a new vendor, Vexata needs to demonstrate the capability of their platform against a range of workloads. Competition in the all-flash market is tough and new solutions need to be able to demonstrate significant benefit over the incumbent vendors. As yet, Vexata doesn't have a high degree of data efficiency features – the platform doesn't support data deduplication, for example. In the wider all-flash market, dedupe can cause performance issues, so the company may be holding off until the feature can be successfully integrated, without impacting performance.

NVMe over Fabrics is still in the early development phase. Features such as multi-pathing do not exist in Ethernet implementations, which means early adopters need to be aware of the risks. The full benefits of VX-OS and the VX-100 platform will be realised with NVMe-oF implementations that will take some time to be delivered.

However, all that said, the approach Vexata is taking makes the technology look very interesting for the future and is one example of new architectures that existing vendors will have to find solutions to match.

Reference Information

- Vexata website - https://www.vexata.com/
- Vexata VX-100 product page - https://www.vexata.com/product/vx-100-scalable-storage-systems/
- Architecting IT VX-100 White Paper – Vexata VX-OS Architecture (PDF) https://tsa.io/brkwp0107

Further details on Vexata and new architectures can be found in the following Architecting IT blog posts:

- Avoiding the Storage Performance Bathtub Curve (22 June 2018) - https://tsa.io/brk-post-905c
- Avoiding All-Flash Missing IOPS (7 June 2018) - https://tsa.io/brk-post-5ef9
- NVMe over Fabrics – Caveat Emptor? (15 June 2018) - https://tsa.io/brk-post-482b
- The Race towards End-to-End NVMe in the Data Centre (4 June 2018) - https://tsa.io/brk-post-1cb1