How does the ADS1000 differ from traditional All Flash Arrays?

Apeiron realized in 2013 that the storage industry had reached an inflection point; customers began to demand more real-time reporting against larger and more complex data sets. These "Big Data" applications were typically deployed across multiple general purpose servers, with internal Flash or HDDs storing the data. In many environments, the speed of deployment and significant cost savings associated with internal storage were "good enough".

To accommodate the shift to scale-out, applications began to include functionality typically reserved for the proprietary storage controllers. Functionality such as snapshots, clones, RAID and distance replication were now managed by the application software. The need for powerful integrated storage controllers had diminished as scale-out grew in popularity.

As applications focused on Operational Intelligence began to grow and prove their business value, users demanded more complex queries spanning ever growing data sets. The scale-out architecture, compelling for its simplicity and cost, was beginning to show serious deficiencies when scaled to 10's or 100's of TeraBytes. Performance plummets and complexity grows as these environments try to keep up with demand. A new approach was needed. Apeiron realized that a storage networking platform built specifically for NVMe SSDs would be the answer. NVMe provides such a significant leap in performance that the traditional "store and forward" arrays connected to a legacy SAN would now be the primary bottleneck.

Apeiron saw this very early, and set out to design an all NVMe storage system, without the expensive and complex controller architecture. This design would eliminate a significant source of latency and cost, and would enable the application to manage storage resources in a manner best suited for it. Apeiron found the answer in 2013 when it combined the industry's latest Field Programmable Gate Array (FPGA) chips, and high bandwidth 40GbE switching components directly in to the drive enclosure. Running "hardened Layer-2 Ethernet" as the transport between drive and server means Apeiron accomplishes 100Gb InfiniBand performance with standard Ethernet components.

A fully networked, scalable NVMe system which looked and acted like internal storage to the server! With the ability to network thousands of NVMe SSDs from any supplier, the system would present a massive pool of ultra high performance SSDs to the application without the complexity and latency of external switching and proprietary storage controllers. To the server, the network would be imperceptible, eliminating the issues scale-out environments encounter when growing.

In 2016, Apeiron announced the General Availability of the ADS1000. A product which has shattered the published performance and scale of all other AFA's. Keep reading to learn how the system has accomplished this...

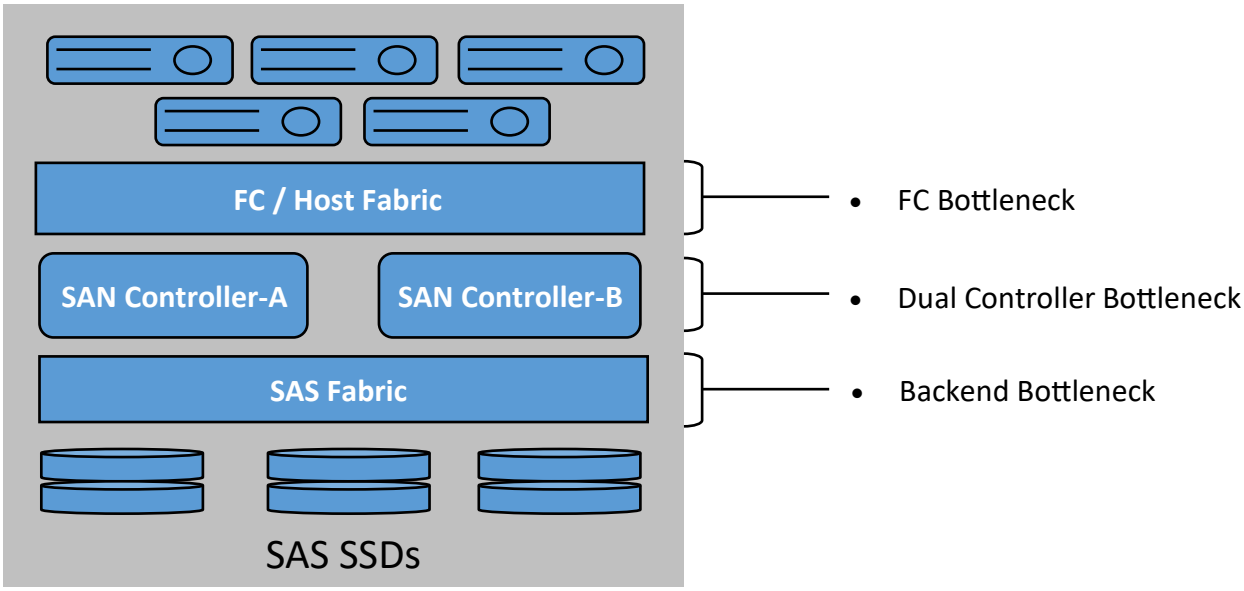
ADS1000 Versus Traditional Architectures

The ADS1000 is based upon two primary components; a server based HBA and 2U storage enclosures. The server side component is installed in the PCIe slot of any x86 server. Each Apeiron HBA contains a single Altera FPGA, which is responsible for encapsulating the NVMe command and passing it to the integrated 40GbE switching complex. This encapsulation protocol adds only 4-Bytes of meta data to every packet. This is an extremely small amount of overhead when compared to RDMA/InfiniBand based solutions. Apeiron's technology provides an extremely lightweight protocol running over a non-blocking switch architecture. The ADS1000 is so efficient that a total latency of only 1.35 micro seconds is added to the transaction time (2.7 micro seconds round trip). Given that NVMe SSD latency is on the order of 90 to 100µS, the ADS1000 is imperceptible to the application or CPU, the ADS1000 looks like internal disk to the server. In addition, this means that NVMe products such as Intel's Optane drive will pass 100% of their performance to the application; Apeiron has removed the #1 bottleneck from the storage complex.

When the packet reaches the integrated switch, it is de-encapsulated by additional FPGA's and passed to the NVMe SSD in its native format. Apeiron can leverage any NVMe SSD without modification, this enables the customer to deploy the proper NVMe drive profile for the applications' needs.

In addition to providing non-blocking switching infrastructure, the same ADS1000 network is used for extremely fast, low latency server to server communications. Dedicated server communication channels ensure there is not any "noise" from the server environment impacting the read/write performance of the storage. The traditional SAN

Figure-1 Traditional SAN dual-controller architecture

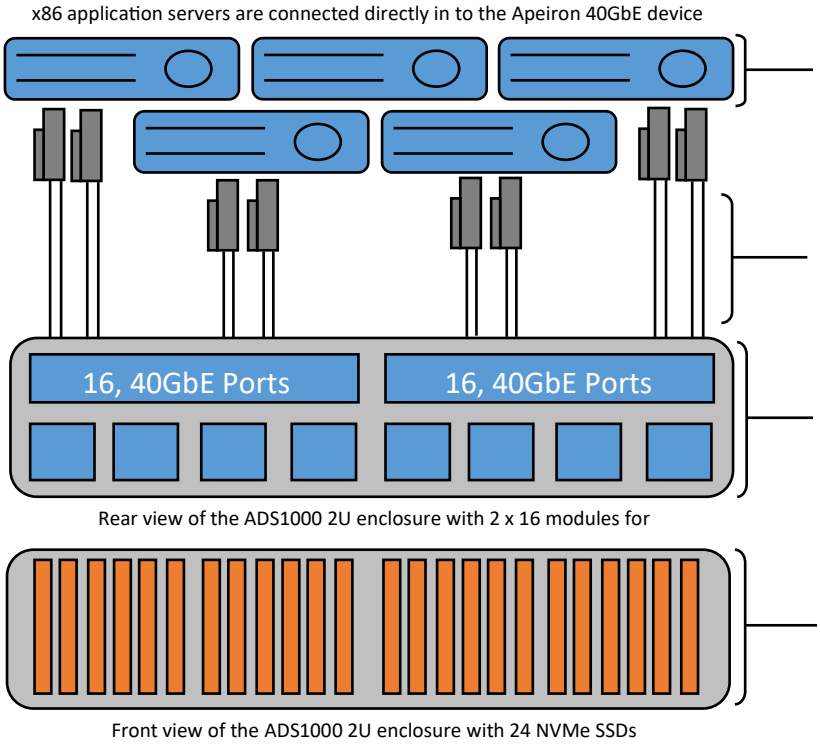


ADS1000 NVMe over Ethernet (NoE) Architecture

By leveraging the power of FPGA technology and integrated 40GbE switching, Apeiron has eliminated the traditional bottlenecks from the storage environment. The performance has proven to be faster than anything available today, and actually improves server performance by 70-150% by eliminating the storage controller. The ADS1000 places the bottlenecks where they belong; at the drive and at the CPU. This is the optimum balance for heavy workloads on massive data sets. Because the compute and storage are disaggregated, servers can be dynamically added to the environment when more compute is necessary. The same is true for storage. When the controller is eliminated from the storage data path, system capacity is no longer limited by the constraints of a complex compute environment embedded in the array.

The Apeiron architecture below removes the traditional bottlenecks induced by external switching protocols and limited controller bandwidth. By leveraging integrated Ethernet components, and moving the network protocol overhead to embedded hardware, the ADS1000 provides over 95GB/s of throughput in each ADS unit. Performance is only limited by the NVMe SSD itself, which typically provides close to 800 k IOPs (4k reads). This equates to 18.4M IOPS per 2U enclosures which scales in a perfectly linear manner with each ADS unit added to the net-

Figure-2 Apeiron ADS1000 NVMe over Ethernet Architecture



Apeiron's technology is compatible with any x86 platform. Server efficiency is improved by 70-150% when storage protocol processing is moved to the application

Each Apeiron HBA is PCIe connected. Using FPGA technology to encapsulate each NVMe command and MOVE the PCIe TLP to the enclosure

Each 2U device has 32, 40GbE ports integrated directly in to the enclosure. The packet is de-encapsulated and routed to proper drive based on its unique MAC address

Each enclosure can accommodate 24 NVMe SSDs from any supplier and of any capacity. This includes Intel Optane Technology

ADS1000 Summary of Benefits

The ADS1000 has solved the scalability and performance issues inherent to captive storage (scale out). The combination of native NVMe networking and a ultra-low latency network protocol provide the industry's best performance and throughput available today. For Big Data and other scale-out workloads where real-time queries on massive data sets is desired, the ADS1000 can scale linearly to meet the business needs.

The system has eliminated the legacy controllers and external switching infrastructure required when the application is not storage aware. The use of common FPGA's remove the performance limiting controller from the architecture completely. This means the application is now free to process and query at the full capability of the NVMe SSD and Server CPU complex. The "middle" infrastructure of external switches and storage controllers are no longer choking your data path.

The removal of expensive controllers and external switches reduce CAPEX and OPEX over traditional architectures. The environment is simplified, and performance is increased by many factors. Apeiron now makes it technically and economically feasible to eliminate complex storage tiering procedures, and have all data reside on ultra fast, high capacity NVMe SSDs. Real-time queries on years of data. The graphics below demonstrate the ADS value in Splunk and Hadoop environments.

ADS1000 Performance (Per 2U)	
Capacity	38/76/184/244/384TB
Latency	100µs (NAND INDUCED LATENCY)
Protocol Overhead	<3µs (roundtrip)
Bandwidth sustained	72 GB/s
Random 4K reads	**18.4 M IOPS PER 2U ENCLOSURE**

ADS1000 performance, capacity and throughput testing prove the absolute best performance NVMe system available. This level of performance and scale translates to queries in years of data instead of days, and a footprint reduction of at least 80%

Type	Records / Second	Total Time (in seconds)
Bare Metal	12,388	4,834
Virtualized Environment	10,528	5,596
Apeiron ADS1000	**86,341**	**620**

Audited Splunk testing left proves at least a 7x query advantage over even bare-metal deployments with captive SSDs. Query was run on ADS while ingesting 8TB/Day with ES

TestDFSIO 1TB Test (512 2GB files)	HDD	SATA SSD[3]	Apeiron NVMe SSD[2]	Apeiron Advantage
Servers[1]	7	7	4	
Datanodes	6	6	3	50% consolidation with a 9x performance increase over scale-out
Disks	12	12	12	Internal SSD vs. external NVMe SSD
Read [MB/s]	722	4,087	35,764	8x-9x IO Performance
Write [MB/s]	218	952	2,526	3x IO Performance with 3x Replication

Cloudera Hadoop TestDSFIO testing vs. HDDs, captive SSDs and Apeiron ADS1000. A 9x performance advantage with a 50% reduction in physical servers