# New NVMe All-Flash A250 System – A Technical Deep Dive

NetApp has introduced a new entry-level NVMe *All-Flash* system for small and medium-sized enterprises, called AFF A250. Doing that, NetApp has completed its NVMe offer, which now spans from A250 through the mid-range class A400 system to the robust A800. The big news, as well, is that the latest version of ONTAP that comes with A250, besides the support for SAN and NAS protocols, will provide support for the S3 protocol (object storage). That makes NetApp a steady leader in the field of UNIFIED FAS storage systems.

A250 Hardware

New A250 comes in a 2U internal chassis, which contains two controllers and two redundant power supplies. On the frontend of the device there is a space, provisioned to store up to 24 internal NVMe SSDs. The device supports expansion with one additional external NS224 NVMe shelf. It is also possible to expand it with one DS224C or DS2246 SAS shelf, but only to transfer existing ones. When purchasing A250 devices, it is not possible to purchase additional DS224C shelves. To conclude, it is possible to expand the device with a single NVMe or SAS shelf.
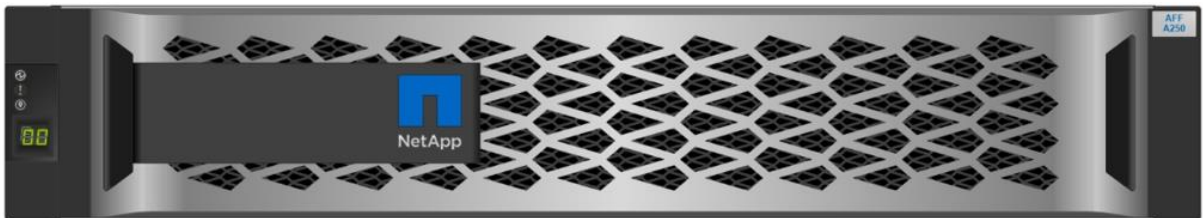
The controllers on A250, work in *Active/Active* mode, as usual. It means that the controllers work independently, each serving their workloads.
In cases of failure of a single controller, the other one, in addition to its own, takes over the tasks of the controller that is out of function (takeover occurs in the interval from 2 to 15 seconds).
This process is transparent, meaning that the applications do not stop working during the *failover* operation.
On the occasion of a sudden power failure, all unregistered data will be copied to the controller's internal *flash* memory.
The data will be saved, in cases of longer failures (battery-assisted NVRAM memory).

AFF A250 has 2x64-bit 12-core Skylake-D CPUs and 128GB RAM.

Scale-out

Besides the expansion with one more shelf, A250 can be extended with the additional NetApp FAS/ AFF systems (scale-out) - with up to 24 nodes in NAS or up to 12 nodes in SAN environment. The nodes added to a scale-out cluster can be of different performance and different generations. To check the compatibility list, please visit NetApp's Hardware Universe portal https://hwu.netapp.com or contact your local Partner.
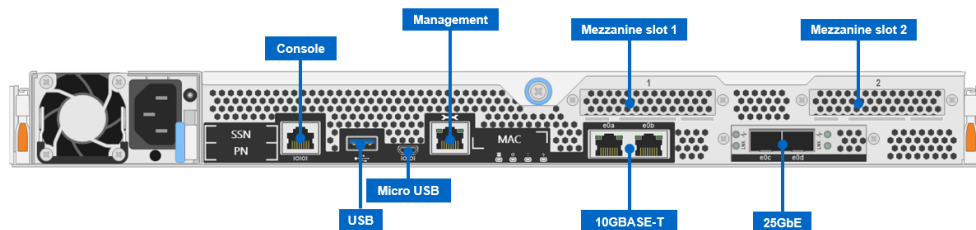
NVMe

NetApp A250 is an *end-to-end* NVMe system, meaning it uses both the NVMe protocol to access SSDs on the storage system and the *NVMe over Fabrics* transport protocol for host connections. This new protocol is designed especially for future-generation flash and SSDs, and it is significantly simpler and optimized to deliver maximum *throughput* and fastest response time. NVMe supports the execution of tens of thousands of parallel commands, can efficiently use not only SSDs, but also multi-core processors and gigabytes of memory, and parses and manages data successfully. By using the NVMe standard, it is possible to deliver 60% more performance, as well as to reduce delays by up to 50%

Connectivity

In terms of connectivity, there are two *mezzanine* slots per controller, which can accommodate the following types of cards:

- 4-port 32Gbps FC and 4-port 10 / 25Gbps Ethernet (RoCEv2) for host connectivity

- 2-port 100Gbps Ethernet (RoCEv2) for added NVMe shelf (slot 1 only) or 4-port 12Gb SAS for additional DS224C or DS2246 shelf.

- Each controller has two 10Gbps BASE-T ports for host connectivity, as well as 2x 25GbE ports for HA and Cluster Interconnect connection (for scale-out needs).

The figure below shows one of the two A250 controllers inside the chassis.

Besides the internal drives, AFF A250 supports up to 24 NVMe SSDs or 24 SAS SSDs, in an additional shelf. Find currently supported capacities on NetApp *Hardware Universe* portal: https://hwu.netapp.com or check it up with your local Partner.

Software Features included in the basic package (Core Bundle)

iSCSI/FC/CIFS/NFS/S3 - the device comes with all licensed protocols, as well as the necessary physical ports to connect to servers /clients. The support for the S3 protocol is a novelty.

Snapshot - NetApp functionality patented for the creation of a large number of *data images* in SAN and NAS environments. NetApp *Snapshot* can be created within seconds, regardless of the size of the data set, because this functionality does not overwrite/copy data while creating a *Snapshot*. For that reason, there can be thousands of *Snapshots* on the system without affecting performance.

SnapRestore - presents functionality that performs data recovery from Snapshot almost instantly, without copying data, both in NAS and SAN environment. Starting from ONTAP 9.8 release, SnapRestore is included in the basic software functionality package.

Inline zero **deduplication** — *Inline* removal of *all-zero* blocks before these blocks are stored on disks.

Inline/post-process **compression** — data compression performed on a "group of blocks" level, enabling significant space savings. *Inline* compression performs before the data is stored on disks. The data which requires more resources is subsequently compressed, using *post-process* compression.

Inline/postprocess **deduplication** - aggregate-level deduplication that takes a 4KB block as a unit, regardless of the data stores as a SAN or NAS. *Inline* deduplication performs before the data stores on disks. The data that requires more resources are subsequently deduplicated, using post-process deduplication, also at the block level.

Inline data compaction - This functionality is in charge of packing logic blocks smaller than 4KB into a single 4KB physical block on disks. The process takes place before the data stores on disks, and it produces a dramatic increase in effective space on *All-Flash* systems. It is enabled on FAS systems, as well.
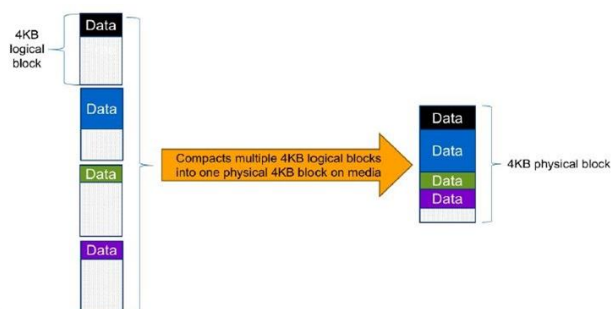
Take a look how this process works when all of these three functionalities are involved - deduplication, compression and compaction:

1. The first thing that happens, as the data comes to registration, is the detection of blocks that contain all zeros. These blocks do not store on disks. The metadata is recorded instead of them.
2. After that, an *inline deduplication* process starts. This functionality was first implemented in version 8.3.2, but the testings were performed only on the data that occurred in memory. With the current version 9, this functionality applies to both the data in memory and the data recently written to disks. In addition to this process, NetApp offers the *postprocess* deduplication option, where the data, not already deduplicated in the inline process (before writing to disks), is being checked and deduplicated subsequently.

Deduplication takes place at a system level, which means that all data on the storage system is considered as a single set that needs to be deduplicated.

3. The next step is *inline compression*. This process determines effectively whether a block can be compressed by more than 50%. The check occurs to prevent spent of CPU cycles on data, which compression could bring more losses than gains. In addition to this process, NetApp has also the ability of a *postprocess* compression, which can perform compression even after the data is written to disks.

4. After the completion of all the processes described above, but before the data is written to disks, a process called *data compaction* occurs (functionality presented in version 9 of the ONTAP OS). All the data that has gone through a compression process previously, or has not yet passed verification, are the candidates for *data compaction*. Before writing to physical blocks, small, uncompressed data blocks (less than 4KB), as well as the data with the compression ratio of more than 75%, are logically placed in a 4KB physical block, which is, in such a form, written to disks. As a rule, the smaller the documents are, and compression is higher, *inline compaction* has a greater synergistic effect.

NetApp All-Flash Efficiency Guaranteed 3:1

For all ONTAP-based All-Flash systems, NetApp guarantees entry of stored data in a 3: 1 ratio. It means that NetApp guarantees storage for three times more data in comparison with the amount of data specified as usable capacity. The conditions for this warranty application are the virtualized environment and the data that hasn't been compacted or encrypted previously. The procedure for receiving the warranty consists of the user's electronically signed confirmation that their environment meets these requirements.

In short, there are three conditions for fulfilling the guarantee:

- the user's environment is virtualized

- stored data has not been compressed previously

- stored data is not encrypted.

Effective data protection

NetApp Volume Encryption - software-based technology used for the encryption of data at rest, at a volume level. The encryption key is only available to the *storage* system and ensures that data can not be read in cases of disks' removal from the system. Both data and metadata are encrypted. The possibility of using the built-in *Onboard Key Manager* allows you to store keys on the system, and there is also the possibility of using an external *Key Manager*. NVE can be used on any type of aggregate (HDD, SDD, hybrid), as well as in conjunction with NSE disks, as the second level of encryption. The NVE license is free and comes with the "Encryption" software Bundle.

RAID-DP – RAID-DP is a NetApp default RAID type (RAID-Double Parity). It requires two *parity* disks per RAID group and protects data in case of a simultaneous failover of up to two disks. For RAID-DP, the recommended RAID group size is 12 to 20 HDDs and 20 to 28 SSDs, which also reduces the cost due to a smaller number of *parity* disks. RAID-DP also allows *on-line* expansion of RAID groups.

RAID-TEC - this patented NetApp technology is just an extension of technology introduced by NetApp a long time ago - RAID-DP (*no penalty* RAID 6). A novelty related to RAID-TEC is that this technology

allows **the simultaneous failure of any 3 disks within a RAID group**. If we take as an example a RAID group composed of 24 disks - even 21 of them are used for data storage, while data protection is provided by the remaining 3 disks. The ratio of data and parity disks, in this case, is 21: 3, or 7: 1.

This mechanism has application in cases of large-capacity disk usage (larger than 6TB), both rotating and SSD disks. With large capacity disks, the recovery time is very long, and the additional protection of the data set is necessary. NetApp is the first storage vendor to certify 15.3TB SSDs.

Thin Provisioning – functionality used for presenting to servers/clients significantly higher logical capacity than the physically occupied capacity on the storage system itself.

FlexVol - logical containers. Over *FlexVol* all other functionalities are performed: *Snapshot,* deduplication, compression, compaction, *thin provisioning, SnapMirror*, *SnapVault*, etc. *FlexVol* logical volumes can be dynamically expanded but also reduced without interruption of applications' operations.

FlexGroup - In large enterprises, a single *namespace* can require petabytes of space, highly exceeding FlexVol's capacity of 100TB. FlexGroup is a volume that supports up to 400 billion files, with 200 member volumes, working together to dynamically balance the load and allocate space equally across all members. No additional management or monitoring of FlexGoup is required. FlexGroup is easy to create and share with users. ONTAP does everything else.

Storage QoS - this functionality ensures that critical workload performance will not be degraded by other workloads competing for the same resources. It is possible to set *QoS Max* and *QoS Min* for each workload. *QoS Max* or *Qos ceiling* is the maximum *throughput* that a certain workload can "occupy", and it limits its impact on system resources. The *QoS Min* or *QoS floor* is adjusted at critical workloads, ensuring that it will receive the minimum necessary *throughput* values regardless of the other workload requirements. It is possible to set both *QoS Max* and *QoS Min* at the same workload. Adaptive QoS will automatically increase/decrease the QoS value with the increasing/ decreasing of the workload size, maintaining the IOPS/TBs (GBs) ratio.
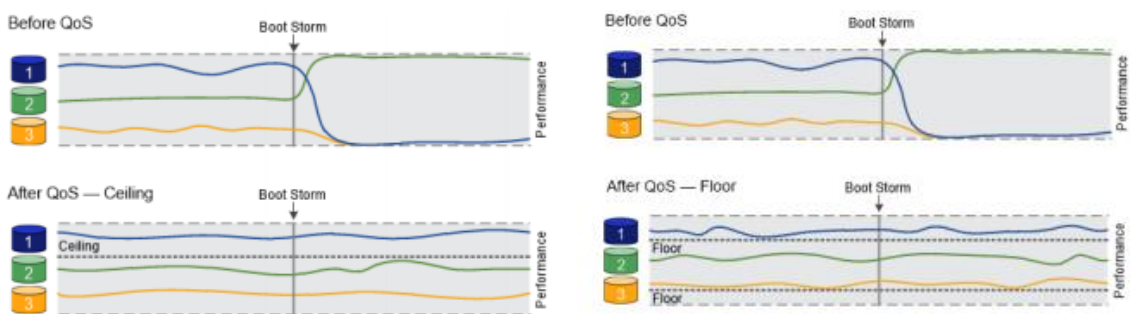
The throughput ceiling for workload 2 ensures that it does not "bully" workloads 1 and 3.

The throughput floors for workload 1 and workload 3 ensure that they meet minimum throughput targets, regardless of demand by workload 2.

System Manager - *OnCommand System Manager* is a graphical interface that allows users to manage NetApp storage systems and objects (disks, volumes, aggregates) and perform general tasks related to *storage* systems from a *web browser*. Cluster administrators can use *the System Manager* to administrate the entire cluster and its resources.

Unified Manager - *Active IQ Unified Manager* is a graphical product that provides extensive monitoring and management capabilities across multiple NetApp ONTAP systems (clusters) to enable management of the availability, capacity, security, and performance. Unified Manager can be installed on Linux and Windows servers or as a virtual appliance on a VMware host.

Virtual Storage Console, VASA Provider, and Storage Replication Adapter - VSC, VASA Provider, and SRA capabilities for the *VMware vSphere product suite* provided through a single virtual appliance. This toolkit enables vCenter Server plug-ins that provide end-to-end lifecycle management for virtual machines in environments that use NetApp storage systems.

ActiveIQ - NetApp Active IQ is a web-based service created for predictive analytics and proactive support to provide optimized operations. Active IQ works with AutoSupport and is used to present information and enable an easy analysis for modeling and optimizing storage infrastructure.

Disaster Recovery Solution - MetroCluster IP

MetroCluster functionality is realized by a combination of 2-point systems at two remote sites, which gives business continuity even in the events of a complete failure of one of the sites. Synchronous data replication is realized with synchronous replication (SyncMirror) to provide continuous data availability on both sites, up to a maximum distance of 700 km. The implementation of MetroCluster maintains the business continuity of critical applications and workloads, in the event of a disaster at one location. It enables a simple and transparent disaster recovery - without data loss, and with a simple configuration that doesn't require subsequent changes at a remote location.

MetroCluster-IP architecture supports 4-node configurations. It can use existing Cluster Interconnect switches, and it has a range of up to 700km distance.

MetroCluster-IP can work in *Active/Active* mode when both sites serve their customers and replicate data to each other, or it can work in *Active /Passive* mode when only one site serves users and data is replicated to the other (DR) location.

Metrocluster-IP has a *Mediator* software solution that comes in the form of a virtual appliance. It is installed in a third location and serves to automate a *switchover* in the event of a failover somewhere in the data center. It enables avoiding the so-called Split Brain scenario.

MetroCluster-IP software is not licensed, and it is a part of every ONTAP system.
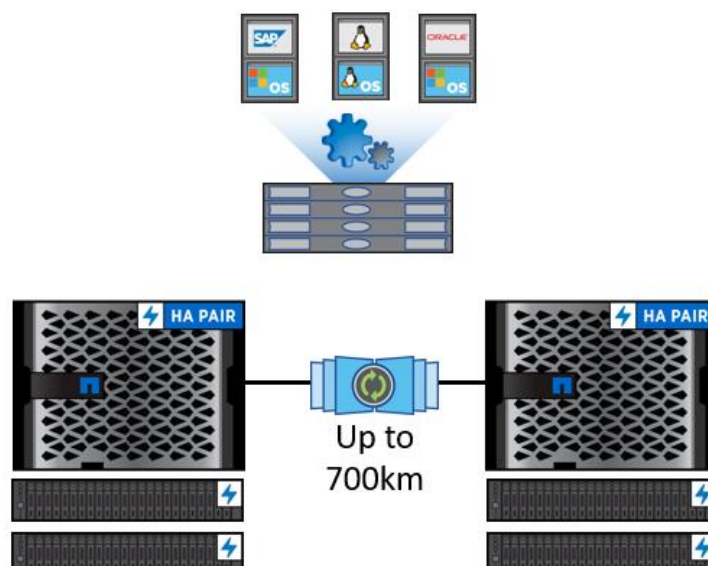
Additional software functionalities (DP Bundle)

FlexClone – Snapshot copy of production data that can be written over. It behaves like an independent copy, although there is no data copying here. Only changes are recorded at the level of a 4KB block. For example, there are dozens of developers who need a copy of a database to test. FlexClone functionality allows all developers to use the same database blocks, but only changes, each of them makes individually, are recorded. It drastically reduces disk space usage because there is no data copying.

SnapMirror - NetApp software that relies on *Snapshot* technology. It is used for replication within the same system, or on another system (location). SnapMirror works at the FlexVol volume level, and it is possible to choose between synchronous and asynchronous replication. The synchronous replication can be set as a:
1. *strict* replication - it will not send a confirmation to the host if the data is not entered in both sites (replication failed), or
2. *relaxed* replication – it will allow data to be written only to the first location and later automatically synchronize with the second location when replication between sites is re-established.

SnapMirror Cloud - a technology that enables the transfer of ONTAP Snapshots directly to object storage. It can be located on-premises or in the cloud. Switching is completely secure and efficient. It provides integrated data mobility and protection at the level of the entire Data Fabric.

SnapMirror Business Continuity - SM-BC is a solution that extends SnapMirror, by enabling automatic failover for business-critical applications without any manual intervention.

SnapVault - NetApp software used to backup data from one *storage* system to another, ie. from faster to slower disks (SSD to SATA / SAS), or to transfer backup data from the primary or multiple locations to the secondary location. SnapVault can be set at certain time intervals, so it allows sending consistent backups to a secondary location.

SnapCenter - A tool that provides centralized control and review, and gives users the ability to manage *application-specific* backups, restores, and clones. Using SnapCenter, database administrators, as well as administrators of virtualization platforms, use a single tool to manage backup, restore, and cloning operations for a variety of applications, databases, and virtual machines.

## Optional software functionalities (out of package)

SnapLock - Compliance solution of a high performance created for organizations that use a *write-once read many* (WORM) storages to keep critical files unchanged for regulatory reasons. The same license allows the use of a SnapLock in two modes. The first one is a more strict Compliance mode, which meets external (government) requirements, and the second one is a flexible *Enterprise* mode for internal regulations. SnapLock uses the *tamper-proof Compliance-Clock* to determine the retention period of WORM files. SnapLock licenses now come with „the Security and Compliance" software Bundle.

Multi-tenant key management - is a functionality that enables each tenant or each Storage Virtual Machine on the storage system to manage its keys through its separate key manager. This license comes with "the Security and Compliance" Software Bundle.

FabricPool - A hybrid *storage* solution that uses All-Flash aggregates as *a performance tier* and object storage as *a capacity tier*. FabricPool will store "hot" data on All-Flash, while it will move inactive "cold" data to object storage, which can be located both in the cloud or on-premises (NetApp StorageGRID).

Inactive data can take up to 80% of the total storage space, making FabricPool an extremely efficient solution, which significantly reduces storage costs without affecting performance, efficiency, or protection. It monitors user activities over each block of data and marks them as "hot" or "cold". Metadata always stays on SSDs, while "cold" data is packaged in 4MB objects and sent to object storage. The whole process is completely transparent for users and applications.

FabricPool requires an additional cluster-wide license, which is paid per TB for use with non-NetApp object systems. Now it comes with a "Hybrid Cloud" software Bundle. No FabricPool license is required when it's used with NetApp object storage.
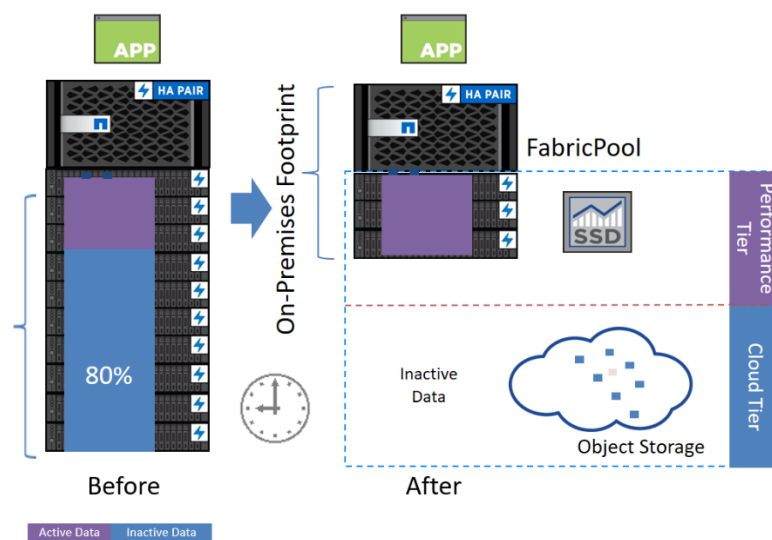
Hardware data encryption

NetApp Storage Encryption (NSE) - a support for special self-encrypting disks that encrypt data while they are written to disk. The data is not able to be read without the encryption key located on the disk. The encryption key, on the other hand, can only be accessed through authenticated nodes. Self-encrypting disks must be ordered when purchasing the device and cannot be mixed with non-encrypting disks on the same system.