



Technical FAQ

Advanced Drive Partitioning Technical FAQ

Mohammad Jawwad Memon, NetApp

January 2015

Overview

This document addresses the most frequently asked question about ONTAP's Advanced Drive Partitioning functionality which includes HDD partitioning, SSD Flash Pool partitioning and AFF use case.

TABLE OF CONTENTS

1	Terminology	3
2	Advanced Drive Partitioning Overview	3
	Why Advanced Drive Partitioning?	3
	What is Advanced Drive Partitioning?	4
3	Root-data Partitioning for HDDs	4
	What is Root-data partitioning for HDDs?	4
	What are the design goals of root-data partitioning?	5
	How do we configure an active/passive configuration on a FAS2520 with root-data partitioning?	5
	How do we configure an active/active configuration on a FAS2520 with root-data partitioning?	6
	Can we add an external shelf to a system that is running Data ONTAP 8.3 with Advanced Drive Partitioning?	7
	What is the usable capacity profile for various drives using root-data partition?	8
	How do I upgrade my system to DATA ONTAP 8.3 to enable Advanced Drive Partitioning?	8
4	Flash Pool SSD Partitioning	8
	What is Flash Pool SSD Partitioning?	8
	What is an SSD Storage Pool?	9
	What is an Allocation Unit (AU)?	9
	Is there a concept of ownership in an SSD Storage Pool?	10
	Can I use system manager to provision cache using Flash Pool SSD Partitioing?	10
	How do I provision SSD cache in a Flash Pool aggregate?	10
5	Root-data partitioning for SSDs (AFF)	11
	What is root-data partitioning for SSDs?	11
	How many SSDs are partitioned with root-data partitioning in an AFF system?	11
6	Supported Configurations	11
	What platforms are supported with Advanced Drive Partitioning?	11
	What drives are being supported with Advanced Drive Partitioning?	12

1 Terminology

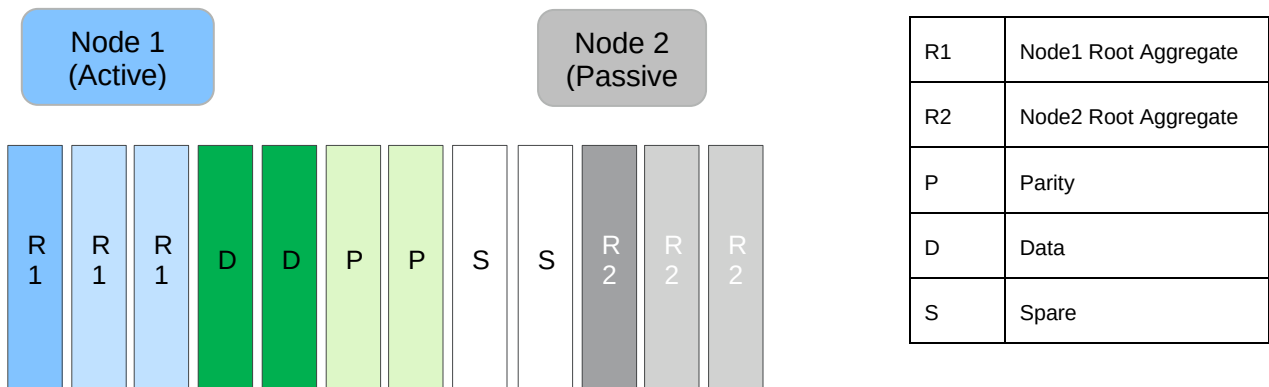
- **Physical disk:** Refers to a full disk (non-partitioned) as we know it today.
- **Shared disk:** Refers to a partitioned disk shared between aggregates and an HA pair.
- **Disk Partition:** Disk Partition is a chunk of a physical HDD or an SSD.
- **Root Partition:** Refers to a smaller partition used by root aggregate
- **Data Partition:** Refers to a larger partition used by data aggregate
- **Storage Pool:** Storage Pool is a collection of partitioned SSDs.
- **MUA:** Minimum Unit of Allocation that can be added to a Flash Pool aggregate.
- **Allocation Unit:** A Minimum Unit of Allocation (MUA) that can be added to a FlashPool aggregate.

2 Advanced Drive Partitioning Overview

Why Advanced Drive Partitioning?

Currently, cDot HA configuration require dedicated root aggregate without user data for each controller. Even if root aggregates are configured as RAID-4 (which is not recommended), this results in serious storage efficiency challenges. Not only does this act as a competitive disadvantage, but is also a significant hurdle to increasing cDot adoption.

FAS2240 HA pair, Active/Passive Configuration running cDOT (without drive partitioning)



Following is the configuration:

- 1) Two dedicated cDOT Root Aggregates:
 - a. Root Aggregate 1: 1D + 2P
 - b. Root Aggregate 2: 1D + 2P
- 2) One Data Aggregate:
 - a. Data Aggregate: 2D + 2P

This configuration shows the serious storage efficiency challenges we are facing in the entry segment especially with cDOT deployments. Having two root aggregates, with Raid-dp and allocating two drives for hot spare leaves us with four drives to create the data aggregate. However, the storage efficiency on 7-mode active-passive configuration is much better since a dedicated root aggregate is not required – although it is not recommended. Still, there is a storage wastage with dedicated root aggregate on the passive node, which becomes a challenge in the extremely cost-capacity sensitive entry space.

What is Advanced Drive Partitioning?

Advanced Drive Partitioning is a new feature, which is part of the Data ONTAP 8.3 feature payload. It addresses the storage efficiency challenges in the entry space, increases cache allocation flexibility, and reduces overall \$/GB across AFF platforms . It is divided into three use cases:

1. Root-data Partitioning for HDDs
2. Flash Pool SSD Partitioning
3. Root-data Partitioning for SSDs (AFF use case)

3 Root-data Partitioning for HDDs

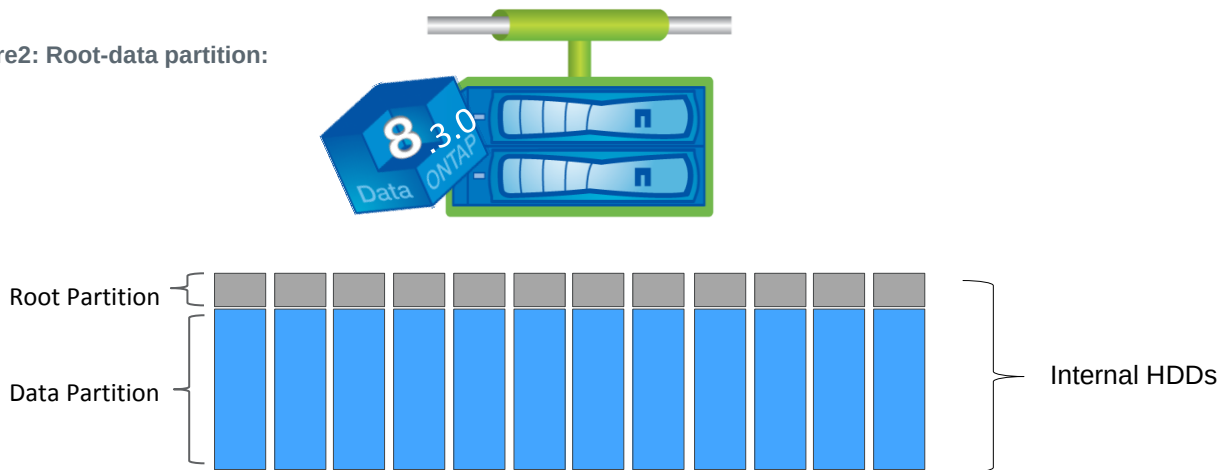
What is Root-data partitioning for HDDs?

This use case is addressing the usable capacity challenges in the entry space—by significantly reducing the root aggregate space allocation. Shared HDDs enables root aggregate to use less space, leaving more space to provision data aggregates. Therefore, it improves storage utilization and increase usable as well as effective capacity from your entry systems.

Each internal HDD is shared by two partitions:

1. **Root Partition:** Root partition is a smaller partition and it is used to create root aggregates.
2. **Data Partition:** Data partition is a larger partition and it is used to create data aggregates.

Figure2: Root-data partition:



It also allows you to create active-active configuration in a 12 drive entry system, which was previously not possible with Clustered DATA ONTAP (in pre-8.3 versions) due to capacity constraints.

In addition, Clustered Data ONTAP requires the use of a dedicated root aggregate, which also plays a significant role in creating the capacity challenges in the platforms where the number of HDDs are limited.

However, the primary reason for using a dedicated root aggregate is that hosting user data volumes on the same aggregate as the root volume can increase the risk of the aggregate becoming inconsistent. (you can find more information on this topic in [Storage Subsystem Technical FAQ](#))

Sharing HDDs into root and data partitioning lets you manage spare capacity efficiently. Users can assign partitions for spare-capacity instead of dedicating the whole drive. Especially in an active/passive configurations where passive node is not serving the data and it is only hosting the root aggregate. Since, root aggregates are provisioned using the smaller root-partitions, you can assign smaller root-partitions as a spare partition, instead of wasting the whole drive. Thus, reducing spare-capacity overhead.

Each partitioned HDD can be shared with up to two nodes and two aggregates (a data aggregate and a root aggregate) within an HA pair. There are some trade-offs to sharing an HDD such as, lack of fault isolation and performance isolation. For example, an internal drive failure can effect the root aggregate as well as the data aggregate. Therefore, Raid-dp is enforced by design to increase resiliency and enable best practices. Moreover, you can have raid-dp on both the root aggregates as well data aggregates. Also, with more capacity available in the system makes it easier to assign spare drive for each node to make sure the system is resilient.

Root-data partitioning is a very compelling feature which makes our entry systems extremely competitive. This is the reason why we say that our next generation of entry system are truly cDOT ready!

What are the design goals of root-data partitioning?

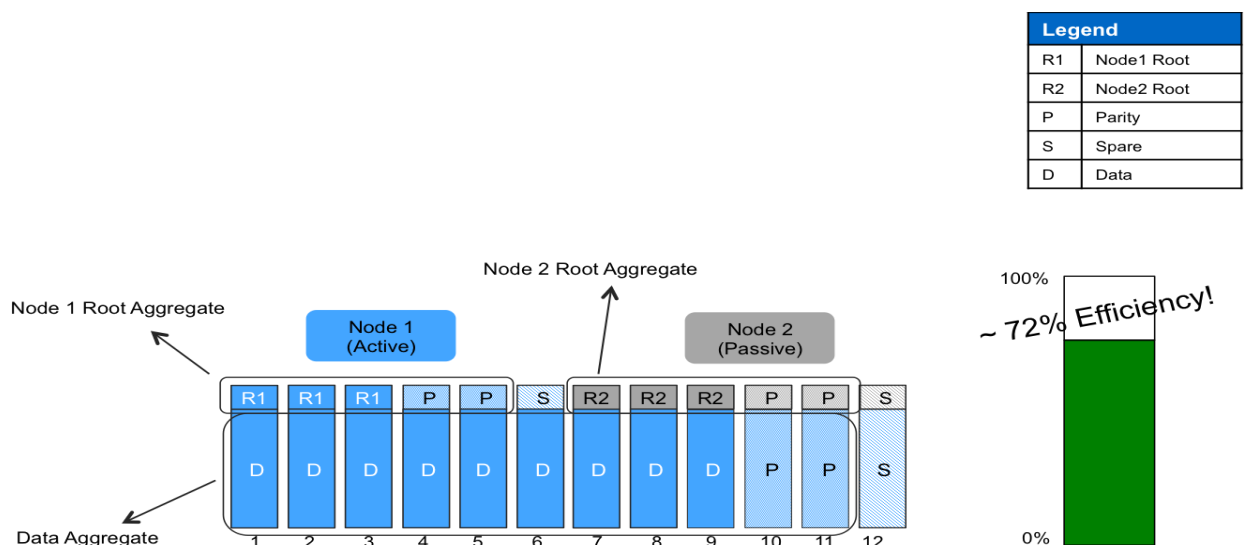
Simplicity, traditional aggregate provisioning model and auto-partitioning.

As we introduce this new root-data partitioning capability, we want to make sure, we do not add any complexities to the system. In fact, we want to make sure we retain the simplicity of the system, which is part of the design goals of root-data partitioning. Therefore, our system admins won't have to learn any new commands to provision aggregates. They won't have to make any choices about partition sizes, which disks gets partitioned or any other details about partitioning. Hence, we are preserving the traditional aggregate provisioning model.

One thing to note here is that partitioning happens automatically during system initialization time. Data ONTAP automatically creates root-data partitioning on each internal HDD when system is zeroing the drives.

How do we configure an active/passive configuration on a FAS2520 with root-data partitioning?

Figure 3: FAS2520 Active/passive configuration



- Two dedicated cDOT root aggregates:

- R1: 3D + 2P + 1S
- R2: 3D + 2P + 1S

- One data aggregate:

- D1: 9D + 2P + 1S

In the above diagram, ONTAP created root aggregates using smaller root-partitions which opens 12 data partitions to create a data aggregate. Note that this configuration is using raid-dp on both the root aggregates as well as the data aggregate. At the same time, spare capacity is allocated based on each aggregate.

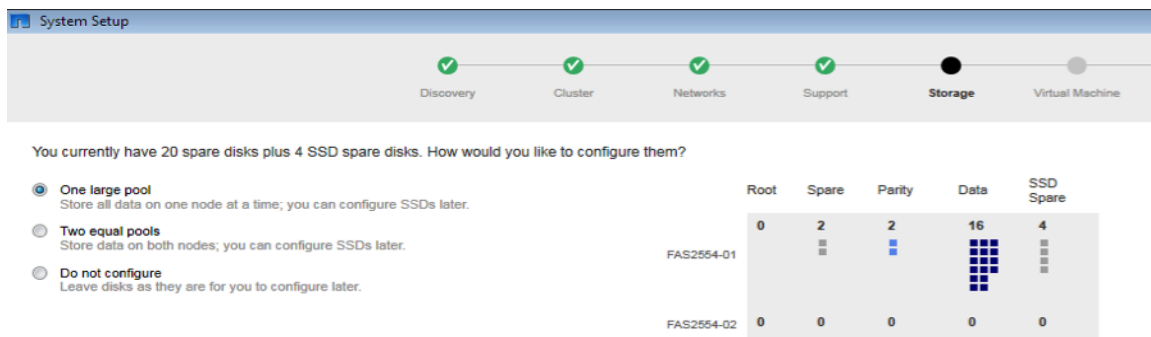
System Setup allows you to create active/passive and active/active configuration during system setup. It automates the process of assigning drives to an active system by letting you select from the following three options:

1. Create one large pool (active/passive)
2. Create two equal pools (active/active)
3. Do not configure.

Below is the snapshot of a system setup creating an active/passive configuration. You can also create an active/passive configuration from command line using the following CLIs:

- storage disk removeowner -data true <disk>
- storage disk assign -data true -disk <disk> -owner <node_name>

Figure 4: System Setup Active/Passive configuration

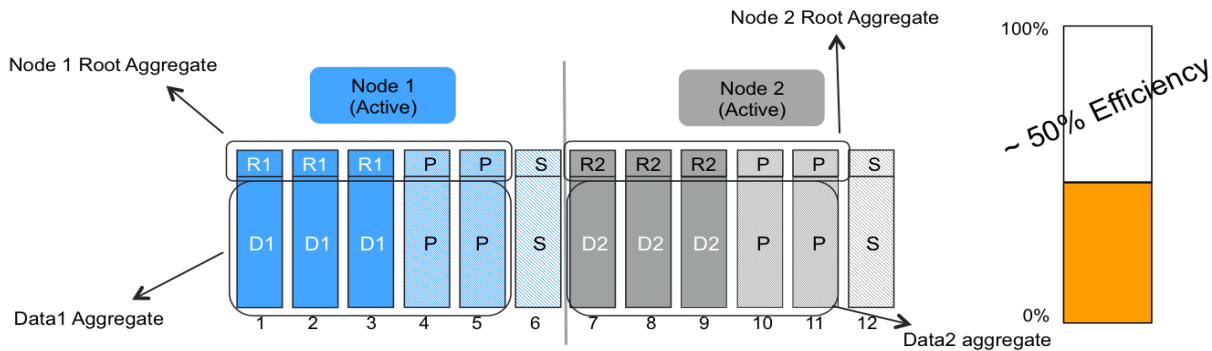


How do we configure an active/active configuration on a FAS2520 with root-data partitioning?

The diagram below shows how root-data partitioning is configured using active/active configuration on a FAS2520 system. This configuration was previously not possible in the pre-8.3 releases. One can take advantage of workload isolation and multi-tenancy using active/active configuration from 12 drive entry system. Overall, storage efficiency is slightly low in this configuration because of two data aggregates which adds to the parity and spare capacity overhead.

Figure 5: FAS2520 active/active configuration

Legend	
R1	Node1 Root
R2	Node2 Root
P	Parity
S	Spare
D	Data



“Figure 4: System Setup Active/Passive configuration” shows the snapshot of a system setup creating an active/active configuration. You can also create an active/active configuration from command line using the following CLIs:

- `storage disk removeowner -data true <disk>`
- `storage disk assign -data true -disk <disk> -owner <node_name>`

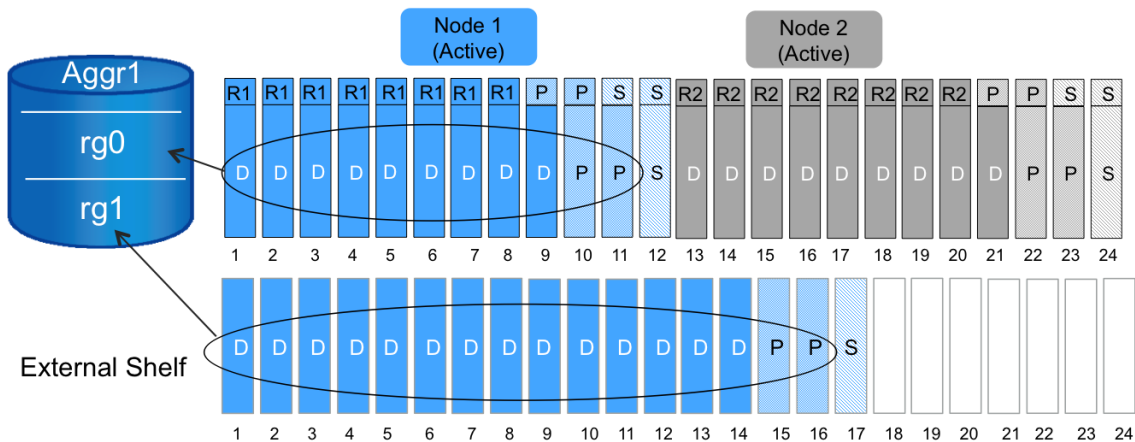
System Setup allows you to create active/passive and active/active configuration during system setup. It automates the process of assigning drives to an active system by letting you select from the following three options:

- Create one large pool (active/passive)
- Create two equal pools (active/active)
- Do no configure.

Can we add an external shelf to a system that is running Data ONTAP 8.3 with Advanced Drive Partitioning?

Yes, you can attach an external shelf to an entry system that is running advanced drive partitioning. If you are looking to grow your existing aggregates which is using internal partitioned drives, you can attach an external shelf and grow the existing aggregate by creating a separate raid group using external drives. Following is an example of mixing partitioned drives with non-partitioned drives in the same aggregate.

Figure 6: Attaching an external shelf with root-data partitioning



In this example, Aggr1 consists of two raid groups, rg0 and rg1; rg0 consists of internal partitioned drives and rg1 consists of external no-partitioned drives. However, if you decide to add the external non-partitioned drives on to the same raid group, say, rg0 ONTAP will then right size (or partition) the external drives.

What is the usable capacity profile for various drives using root-data partition?

Please see Synergy for usable capacity values for all the different drive types and configuration. More information can be found on this spreadsheet: [Usable Capacity with Advanced Drive Partitioning](#)

How do I upgrade my system to DATA ONTAP 8.3 to enable Advanced Drive Partitioning?

There are currently three procedures available for upgrading to 8.3 to enable root-data partitioning on an entry system:

1. [Disruptive procedure to enable Advanced Drive Partitioning with cDOT 8.3](#)
2. Non-disruptive procedure to enable Advanced Drive Partitioning with cDOT 8.3
3. [KB article to upgrade the newly ordered entry systems running cDOT 8.2 to cDOT 8.3 to enable Advanced Drive Partitioning](#)

Can we utilize root-data HDD partitioning and SSD FlashPool partitioning on one system simultaneously?

Yes, on an entry system root-data partitioning will be pre-enabled by Data ONTAP 8.3 during system initialization. If you have SSDs on your system, you can create storage pool and take advantage of Flash Pool SSD partitioning.

4 Flash Pool SSD Partitioning

What is Flash Pool SSD Partitioning?

In this use case each SSD is divided into four equal partitions. SSDs are provisioned as a cache in a flash pool aggregate. This use case is very different from root-data partitioning. Root-data partitioning allows root aggregates to be stored on smaller root-partitions, which helps in reducing root aggregate space utilization and allows for increased usable capacity. Flash Pool SSD Partitioning on the other hand is

used just for caching and it enables you to share each SSD with up to four flash pool aggregates. Sharing SSDs as a caching tier in your flash pool aggregate reduces parity and spare capacity overhead because you can now allocate a single SSD for parity/spare and share it among multiple flash pool aggregates.

Without partitioning, the capacity and performance of an SSD is only dedicated to a single aggregate. With partitioning, you can share SSDs, which are capable of generating a lot of IOPs and throughput among multiple flash pool aggregate. Therefore, it improves SSD storage utilization and helps you achieve higher performance.

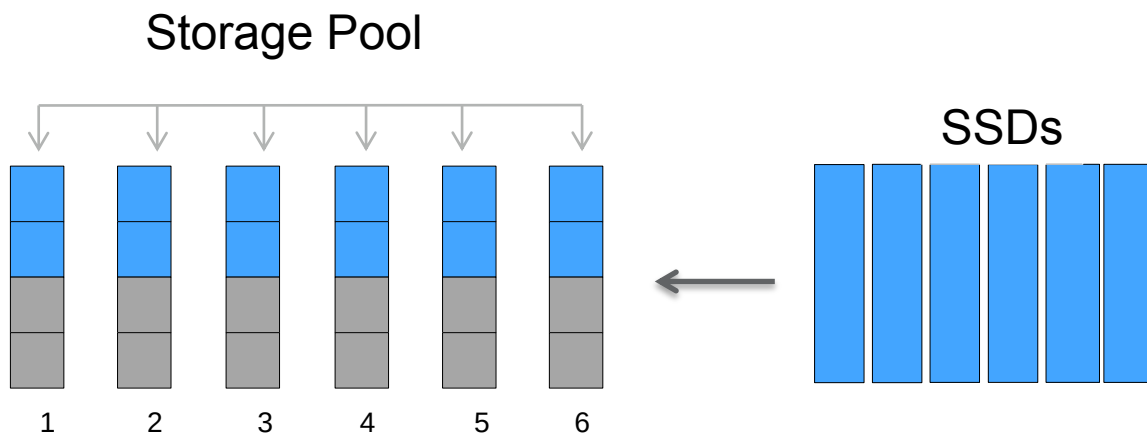
Unlike root-data partitioning, SSDs in this use case are not partitioned automatically because SSDs here are used just for data provisioning (caching). Thus, it is designed to provide more control over to the users. Therefore, users have to manually enable partitioning by creating an SSD storage pool.

What is an SSD Storage Pool?

Storage Pool is a new concept which is introduced as part of the flash pool SSD partitioning functionality. It increases cache-allocation flexibility by allowing you to provision cache based on capacity rather than the number of drives. It is a new capacity-based provisioning model, which allows our system admins to provision cache based on capacity, instead of traditional drive-based provisioning. This is NetApp's vision in going forward with high capacity SSD drives in order to increase SSD storage utilization. NetApp's intelligent I/O scheduling module allows SSDs to share among both the nodes within an HA pair and with up to four flash pool aggregates.

Furthermore, storage pool can be created using *storage pool create command*, SSDs are partitioned when SSD storage pool is created. Following is an example of an SSD storage pool with 6 SSD drives:

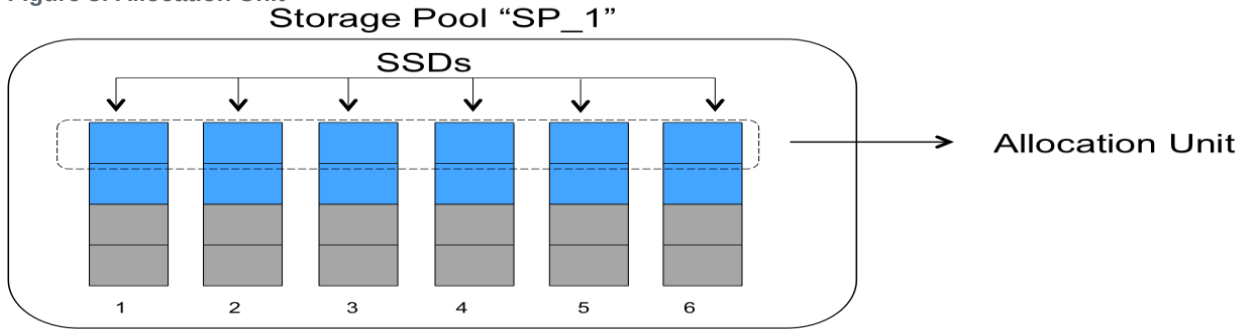
Figure 7: Storage Pool



What is an Allocation Unit (AU)?

Storage pool is divided into four Allocation units. Each allocation unit contains one partition from each SSD in an SSD storage pool i.e. one allocation unit represents 25% of the cache from the storage pool. Allocation unit is also a minimum unit of allocation (MUA) that can be added to a flash pool aggregate as a single raid group. Therefore, you can also think of an AU as a single raid group.

Figure 8: Allocation Unit



Is there a concept of ownership in an SSD Storage Pool?

When you create an SSD storage pool. By default, two allocations are assigned to each node. There is an option to change the ownership of allocation units (AUs) by running *storage pool reassign* command.

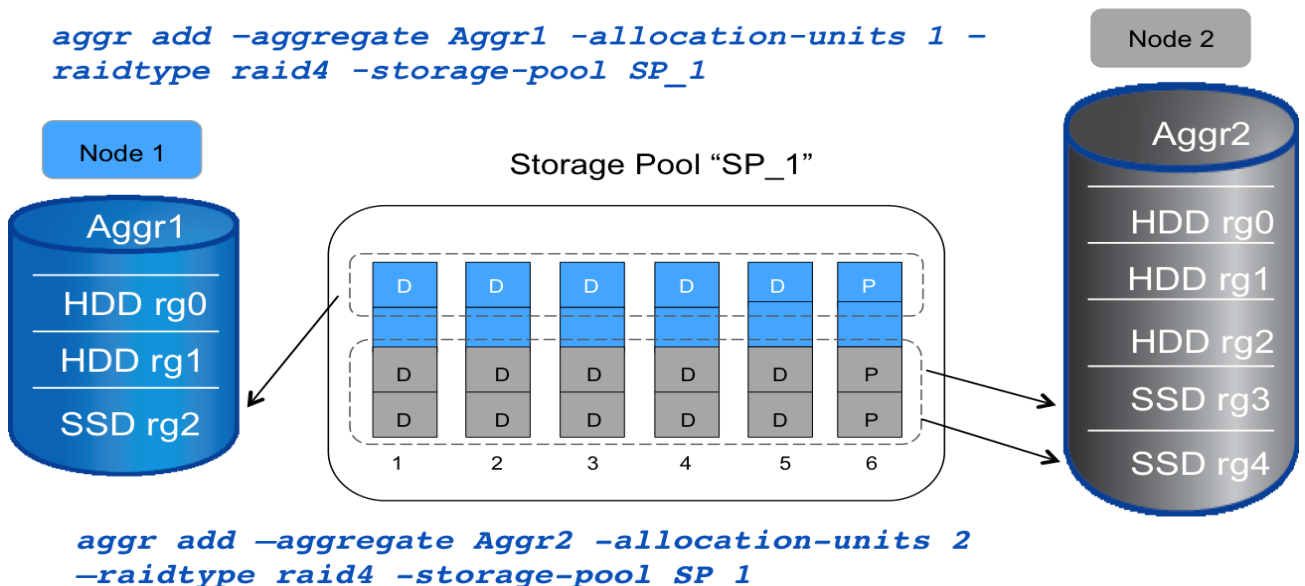
Can I use system manager to provision cache using Flash Pool SSD Partitioning?

Yes, system manager is fully supported with Flash Pool SSD partitioning. It allows you to create SSD storage pool and assign allocation units to your flash pool aggregates. System manager's user friendly GUI allows you to easily manage your cache without having to remember CLIs. It also recommends best practices for managing spare drives.

How do I provision SSD cache in a Flash Pool aggregate?

As mentioned earlier, in order to provision cache in a flash pool aggregate, create an SSD storage pool and then assign allocation units based on your needs. For example the following diagram shows how to provision a flash pool aggregate using partitioned SSDs.

Figure 9: Flash SSD Partitioning using storage pool



5 Root-data partitioning for SSDs (AFF)

What is root-data partitioning for SSDs?

This use case is very similar to the first use case i.e. the root-data partitioning for HDDs. Same concept is applied to the AFF product line to increase usable as well as effective capacity from our AFF systems by significantly reducing the root aggregate space allocation.

Root-data partitions are created during system-initialization and ONTAP divides each SSD into two partitions:

1. **Root Partition:** Root partition is a smaller partition and it is used to create root aggregate.
2. **Data Partition:** Data partition is a larger partition and it is used to create data aggregate.

Root-data partitioning for SSDs efficiently manages root aggregate allocation which opens up more space to provision data aggregates. This enables reduced \$/GB across the board on our entry, mid and high-end AFF systems.

Sharing SSDs into root and data partitioning lets you manage spare capacity efficiently. Users can—assign partitions for spare-capacity instead of dedicating the whole drive. This technique reduces parity and spare capacity overhead. Therefore, with all the benefits of advanced drive partitioning combined with the world class Flash performance. Running on top of world's number one storage operating system, this makes our AFF product line extremely powerful.

How many SSDs are partitioned with root-data partitioning in an AFF system?

Maximum of 48 SSDs are partitioned per HA pair and atleast 8 SSDs are required to have root-data partitioning on an HA pair. This minimum requirement is driven by the size of the root aggregate needed by DATA ONTAP. Therefore, at least 8 SSDs are required to enable partitioning on an AFF system.

6 Supported Configurations

What platforms are supported with Advanced Drive Partitioning?

Entry = Root-data partitioning for HDDs (use case #1)

Flash Pool = Flash pool SSD partitioning (use case #2)

AFF = Root-data partitioning for SSDs (use case #3)

Data ONTAP®	Clustered Data ONTAP 8.3	✓
Platforms	FAS25xx – Entry, Flash Pool, and AFF	✓
	FAS22xx – Entry, Flash Pool, and AFF	✓
	FAS80xx – Flash Pool, and AFF	✓
	FAS62xx – Flash Pool, and AFF	✓
	FAS32xx – Flash Pool, and AFF	✓

What drives are being supported with Advanced Drive Partitioning?

Drive Support		
Drive Category	Drive Type	Capacity
Capacity	SATA/NL-SAS – 7.2K RPM	2TB, 3TB, 4TB, 4TBNSE & 6TB
Performance	SAS – 10K RPM	600GB, 900GB, 900GB NSE, & 1.2TB
Ultra Performance	SSD	200GB, 400GB, 800GB, 1.6TB, & 800GB NSE
Comments	- Root aggregate is configured by cDOT automatically - User to provision Flash Pool SSD Partitions - User to provision data aggregates	

Go further, faster®



NetApp® Advanced Drive

www.netapp.com

NetApp provides no representations or warranties regarding the accuracy, reliability, or serviceability of any information or recommendations provided in this publication, or with respect to any results that may be obtained by the use of the information or observance of any recommendations provided herein. The information in this document is distributed AS IS, and the use of this information or the implementation of any recommendations or techniques herein is a customer's responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. This document and the information contained herein may be used solely in connection with the NetApp products discussed in this document.