Technical Report

All Flash FAS SAN-Optimized Configuration

Michael Peppers, NetApp
June 2016 | TR-4480

Abstract

This technical report provides details about the SAN-optimized configuration for NetApp® All Flash FAS clusters.

Data Classification

Public.

Version History

Version	Date	Document Version History
Version 1.0	December 2015	Initial version. Covers NetApp clustered Data ONTAP® 8.3.1 storage operating system.
Version 1.1	June 2016	Updated for ONTAP 9

TABLE OF CONTENTS

Version History **2**

TABLE OF CONTENTS **3**

1 Overview **4**

2 Configuration Details **4**

2.1 Cluster and Node Configuration 4

2.2 Disk Layout and Ownership 4

2.3 FC Target Ports 5

2.4 Cluster and Default Storage Virtual Machine Configuration 6

3 SAN-Optimized Storage Provisioning Workflows **7**

3.1 Provisioning Workflow Overview 7

3.2 Oracle Databases 8

3.3 Microsoft SQL Server Databases 8

3.4 Manual Provisioning 9

4 Other Considerations **9**

4.1 Deploying an All Flash FAS SAN-Optimized Configuration into Production 9

4.2 Storage Failover and Giveback Timing on All Flash FAS Controllers 9

5 Additional Resources **10**

6 Contact Us **10**

LIST OF TABLES

Table 1) Spares required for All Flash FAS root-data partitioning. 4

Table 2) Aggregate disk layouts. 5

Table 3) UTA2 target port configuration details. 5

Table 4) Default LIF assignments. 6

1 Overview

This technical report presents an overview of NetApp All Flash FAS (AFF) storage controllers that run Data ONTAP® 8.3.1 or later, in the SAN-optimized configuration, along with further technical details. It is meant to be a reference for storage administrators and architects.

The purpose of the SAN-optimized configuration is to simplify the deployment of All Flash FAS–based storage clusters. This approach both preconfigures the ONTAP storage operating system optimally for presenting data through Fibre Channel (FC) and offers ease-of-use improvements to the initial storage provisioning workflow.

At first access after installation, an All Flash FAS cluster is immediately ready to provision storage and present it through FC. Section 2 of this technical report details the default configuration options that are initially set for an All Flash FAS SAN-optimized cluster. Section 3 contains more details about the simplified LUN provisioning workflow.

2 Configuration Details

The following section details the preconfigured defaults that are applied to a SAN-optimized All Flash FAS system by NetApp.

2.1 Cluster and Node Configuration

An All Flash FAS SAN-optimized cluster in its initial state has already completed the cluster setup process. The default cluster name is `aff`. The storage controllers that belong to the default cluster are named after the cluster with an appended number in the default fashion. For example, the first node's name is `aff-01`. By default, this node is also the owner of the cluster's management logical interface (LIF); see section 2.4.

For more details about the default configuration of the All Flash FAS SAN-optimized management interface, see the FC SAN Optimized AFF Setup Guide.

2.2 Disk Layout and Ownership

Root-Data Partitioning

All Flash FAS controllers use root-data partitioning to increase storage efficiency by dividing each solid-state drive (SSD) into separate root and data partitioning. This separation effectively spreads the root partition across data drives in the node's first aggregate and eliminates the requirement to reserve SSDs for the exclusive use of a root aggregate. An All Flash FAS cluster in the SAN-optimized configuration arrives partitioned, as specified in Table 1.

All Flash FAS controllers with more than three shelves reserve a non-partitioned SSD as a universal replacement for an SSD that has a failing root or a data partition.

Table 1) Spares required for All Flash FAS root-data partitioning.

Number of Shelves per HA Pair	Spare Root Partitions per Node	Spare Data Partitions per Node	Spare Non-partitioned SSDs per Node
1	1	1	0
2	2	1	0

Number of Shelves per HA Pair	Spare Root Partitions per Node	Spare Data Partitions per Node	Spare Non-partitioned SSDs per Node
≥ 3	0	0	1

For more details, see the section Understanding root-data partitioning in the Data ONTAP 8.3.1 Physical Storage Management Guide.

Disk Ownership

To maximize use of the AFF8000 series storage controller I/O capabilities, an All Flash FAS SAN-optimized system splits shelves of SSDs between storage controllers in an HA pair. The first node owns the SSDs in bays 0 to 11, and the second node owns the SSDs in bays 12 to 23. This approach is a best practice for All Flash FAS systems irrespective of which protocols are used to serve data.

Data Aggregates

An All Flash FAS cluster in the SAN-optimized configuration arrives with a data aggregate that has already been created for both nodes and that spans its available drives, except for spare partitions. The aggregate is sized to appropriately work with the HA pair's shelves, as noted in Table 2.

Table 2) Aggregate disk layouts.

Number of Shelves per HA Pair	Root Aggregate Disk Layout	Data Aggregate Disk Layout
1	9 data, 1 parity	9 data, 2 parity
≥ 2	20 data, 2 parity	21 data, 2 parity

2.3 FC Target Ports

All Flash FAS controllers in the SAN-optimized configuration have their onboard UTA2 (unified target adapter) ports already configured to serve data through the FC protocol. In anticipation of the standard best practice of connecting a storage system to two redundant FC fabrics, the target ports are predesignated for either fabric A or fabric B, as detailed in Table 3.

Table 3) UTA2 target port configuration details.

Target Port	Fabric	Personality	Mode	Speed
0e	A	FC	Target	16Gb (auto)
0f	B			
0g	A			
0h	B			

2.4 Cluster and Default Storage Virtual Machine Configuration

An All Flash FAS SAN-optimized storage cluster is delivered with a preconfigured storage virtual machine (SVM) with appropriate default settings (as detailed in this section) and with the name **AFF_SAN_DEFAULT_SVM**. No volumes or LUNs are configured; volumes and LUNs are configured using the workflows in section 3, SAN-Optimized Storage Provisioning Workflows.

FC Target Logical Interfaces

The default SVM has a logical interface (LIF) that is associated with each of the UTA2 target ports that are configured to `target` mode, as detailed in section 2.3, FC Target Ports. The worldwide port names (WWPNs) of these LIFs are not preconfigured, but instead are sourced from the pre-generated pool of WWPNs that are globally unique for that cluster. The LIFs, listed in Table 4, are named to identify their node, port, and fabric assignments.

Table 4) Default LIF assignments.

Node	Target Port	Fabric	Name
aff-01	0e	A	`aff_node-01_0e_A`
aff-01	0f	B	`aff_node-01_0f_B`
aff-01	0g	A	`aff_node-01_0g_A`
aff-01	0h	B	`aff_node-01_0h_B`
aff-02	0e	A	`aff_node-02_0e_A`
aff-02	0f	B	`aff_node-02_0f_B`
aff-02	0g	A	`aff_node-02_0g_A`
aff-02	0h	B	`aff_node-02_0h_B`

Note: LIFs associated with a node's target ports do not prevent other LIFs from being assigned to the same target port, either within the same SVM or in another SVM within the same cluster.

For more information about the configuration of FC LIFs, see TR-4080: Scalable SAN Best Practices in Clustered Data ONTAP.

Storage Efficiency Policies

A storage efficiency policy in Data ONTAP can indicate that a volume should be deduplicated and/or compressed and can also specify a quality-of-service (QoS) policy and schedule (or no schedule). By default, inline compression is turned on for every All Flash FAS controller and cannot be disabled. This setting serves to increase both storage utilization of SSDs and overall read throughput, because fewer compressed blocks must be read.

All Flash FAS SAN-optimized storage controllers also have the maximum number of concurrent deduplication threads per node set to 1, as a performance- and efficiency-balancing optimization.

Aggregate Full and Nearly Full Thresholds

Aggregates can have a fullness threshold set so that when the total percentage of used space in the aggregate exceeds the threshold, an event is generated. This event can then be forwarded to a SNMP-based monitoring tool.

To strike a balance between optimized performance and storage efficiency, All Flash FAS SAN-optimized storage controllers have their nearly full threshold set to 70% and their full threshold set to 75%. LUNs and volumes that are created by using the workflows in section 3, SAN Optimized Storage Provisioning Workflows, are thinly provisioned. By lowering both thresholds, storage administrators have ample opportunity to take action well before an aggregate is completely filled, despite the smaller storage space that is commonly available when compared with storage controllers that use spinning media.

Management LIF and Administrative User

SVMs appear to hosts that access them to be separate storage arrays, despite occupying the same physical hardware. Therefore, some storage management tools, such as NetApp SnapDrive® software, interface with a storage controller that runs clustered Data ONTAP® at the scope of an individual SVM rather than at the cluster level. This approach requires a LIF that permits management access through an API or a login.

All Flash FAS SAN-optimized clusters are configured to serve data through FC, the FC LIFs configured for `AFF_SAN_DEFAULT_SVM` are unusable for administrative purposes. Therefore, a separate management LIF is configured by default and is assigned to the `e0M` port of the node `aff-01`.

For more information about SVM-scoped management LIFs and administrator accounts, see TR-4080: Scalable SAN Best Practices in Clustered Data ONTAP.

For more information about the default IP address of the All Flash FAS SAN-optimized SVM management LIF, see the FC SAN Optimized AFF Setup Guide.

3 SAN-Optimized Storage Provisioning Workflows

In addition to an initial configuration that is designed to be ready to serve data by using the FC protocol, the All Flash FAS SAN-optimized configuration also simplifies LUN provisioning workflows. To achieve optimal performance, application data must be laid out across multiple nodes, volumes, and LUNs. Rather than using similar but not identical CLI commands, or the OnCommand System Manager LUN provisioning wizard multiple times, an appropriate number of LUNs can be created across the cluster nodes and mapped to host initiators in a single step.

Starting with ONTAP 8.3 an instance of OnCommand System Manager that runs on-box, can be accessed through a cluster management LIF. On SAN-Optimizied AFF systems when OnCommand System Manager is accessed for the first time, it will detect a default **`AFF_SAN_DEFAULT_SVM`** without any LUNs or volumes already provisioned. It will launch a LUN provisioning workflow sheet, allowing storage administrators to select from three common deployment scenarios: An Oracle database, a Microsoft SQL Server database, or a user-specified LUN layout.

3.1 Provisioning Workflow Overview

All the workflows configure LUNs and volumes to be thinly provisioned. The workflows also configure them in such a way that only free space within the LUN (at the host level) and free space at the aggregate level (in Data ONTAP) must be monitored.

Volumes have the following characteristics:

- The volume size is configured to equal aggregate size.
- The space guarantee is set to `none`, and the fractional NetApp Snapshot® reserve space is set to `0`. Only consumed space is deducted from the free space of the containing aggregate.

- Autosize is turned off.

  Volumes do not need to be resized if they are already at their maximum size.
- The Snapshot policy is set to `none`.

  To be internally consistent, LUN Snapshot copies should be managed by an external storage management application such as NetApp SnapManager® or NetApp SnapCenter® software and not by an internal schedule.
- The volume efficiency policy is set to `none`.

  By default, no volume efficiency runs are scheduled. The reason for this approach is twofold: First, deduplication does not necessarily provide a storage efficiency benefit for the database-oriented application provisioning workflows that are included with System Manager. Second, inline compression is always enabled for All Flash FAS controllers.

Additionally, LUNs that are provisioned by using the workflows have their space reserve disabled, so only blocks that are actually allocated are marked as used.

Finally, an igroup or igroups are created by the provisioning workflow by using WWPNs that are provided by a storage administrator, and the LUNs are mapped accordingly. The Oracle provisioning workflow defaults to igroups with the Linux OS type, and the Microsoft SQL Server workflow defaults to igroups with the Windows OS type.

In every case, the provisioning workflows create one volume to contain each LUN.

3.2 Oracle Databases

The Oracle database provisioning workflow uses the following provisioning logic to determine LUN and volume layout.

Inputs:

- Database size (default: 2TB)
- Oracle RAC installation (default: `no`)
- Database instance name (default: `ora1`)

Output:

- LUNs are created for a single database on a single Oracle instance.
- LUNs are presumed to use Oracle ASM or Linux LVM file systems.
- The Oracle LUNs are prefixed with the Oracle instance name.
- The workflow creates eight data LUNs and one log LUN divided equally between the cluster nodes **`aff-01`** and **`aff-02`**. The workflow also creates one LUN for Oracle binaries if the RAC option is not selected. If RAC is selected, a second LUN for binaries is created, along with a LUN for the Oracle Grid Infrastructure binaries.

3.3 Microsoft SQL Server Databases

The SQL Server database provisioning workflow uses the following provisioning logic to determine LUN and volume layout.

Inputs:

- Number of databases to create (default: 2)
- Size of databases to be created (default: 1TB)
- Number of physical cores in each host accessing database LUNs

Output:

- Database names are specified with a **`DB0`*`n`*** prefix.
- Data volumes and LUNs are named **`DB0`*`n`*`_data0`*`n`***.
- The total number of LUNs is a maximum of 8 data LUNs per database, or equal to the number of CPU cores per server, whichever is higher. The data LUNs are divided equally between the cluster nodes **`aff-01`** and **`aff-02`**.

3.4 Manual Provisioning

Manual provisioning allows a storage administrator to provision an arbitrary number of LUNs of an arbitrary size and to automatically map them to a new igroup that is specified by entering a host WWPN. Each LUN receives its own volume, and the volumes are distributed among the aggregates of the cluster nodes **`aff-01`** and **`aff-02`**.

4 Other Considerations

4.1 Deploying an All Flash FAS SAN-Optimized Configuration into Production

The All Flash FAS SAN-optimized configuration is designed to simplify the deployment of FC-based storage and to minimize the number of steps that are necessary to serve data during a demonstration or a proof of concept. Therefore, not every setting that might be necessary in a production environment is configured. The FC SAN Optimized AFF Setup Guide has a checklist of configuration options and/or additional setup that should be considered before moving an All Flash FAS SAN-optimized storage cluster to a production deployment.

4.2 Storage Failover and Giveback Timing on All Flash FAS Controllers

Customers who deploy a mission-critical/business-critical storage workload and implement NetApp prescriptive AFF SAN configurations for such a scenario can expect planned storage failovers to complete in less than 15 seconds. They can expect unplanned storage failovers to complete in less than 30 seconds. As a result of performance issues associated with specific workloads and identified special storage operations, up to 5% of failover or giveback events could exceed these failover times.

5 Additional Resources

- TR-4080: Scalable SAN Best Practices in Clustered Data ONTAP
  www.netapp.com/us/media/tr-4080.pdf
- FC SAN Optimized AFF Setup Guide
  https://library.netapp.com/ecm/ecm_download_file/ECMP12462582
- Data ONTAP 8.3.1 Physical Storage Management Guide
  https://library.netapp.com/ecmdocs/ECMP12458210/html/index.html

6 Contact Us

Let us know how we can improve this technical report.

Contact us at docfeedback@netapp.com.

Include TECHNICAL REPORT 4480 in the subject line.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.