NetApp

Technical Report

NetApp Clustered Data ONTAP 8.2

An Introduction

Charlotte Brooks, NetApp

May 2013 | TR-3982

Abstract

This technical report is an introduction to the architecture and key customer benefits of NetApp® clustered Data ONTAP® 8.2.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 NetApp clustered Data ONTAP 8.2: Overview

With the release of NetApp clustered Data ONTAP (clustered ONTAP), NetApp was the first to market with enterprise-ready, *unified scale-out storage*. Developed from a solid foundation of proven Data ONTAP technology and innovation, clustered ONTAP is the basis for virtualized shared storage infrastructures that are architected *for nondisruptive operations* over the lifetime of the system.

This paper is an overview of clustered Data ONTAP, including its architecture and core capabilities, positioning it firmly in today's agile data centers. Clustered Data ONTAP 8.2 increases the scalability, protocol support, and data protection capabilities of previous releases. It also allows the coexistence of multiple Storage Virtual Machines (SVMs) targeted for various use cases, including large NAS content repositories, general purpose file services, and enterprise applications.

1.1 Intelligent Scale-Out Storage Versus Scale-Up

Scale-out storage is the most powerful and flexible way to respond to the inevitable data growth and data management challenges in today's environments. Consider that all storage controllers have physical limits to their expandability—for example, number of CPUs, memory slots, and space for disk shelves—that dictate the maximum capacity and performance of which the controller is capable. If more storage or performance capacity is needed, you might be able to upgrade or add CPUs and memory or install additional disk shelves, but ultimately the controller will be completely populated, with no further expansion possible. At this stage, the only option is to acquire one or more additional controllers. Historically this has been achieved by simple "scale-up," with two options: either replace the old controller with a complete technology refresh, or run the new controller side by side with the original. Both of these options have significant shortcomings and disadvantages.

In a technology refresh, data migration is necessary to copy the data from the old to the new controller and reconfigure the environment there. This is time consuming, planning intensive, often disruptive, and typically requires configuration changes on all of the attached host systems in order to access the new storage resource. Data migrations have a substantial impact on storage administration costs and administrative complexity. If the newer controller will coexist with the original controller, there are now two storage controllers to be individually managed, and there are no built-in tools to balance or reassign workloads across them. Data migrations will also be required. The situation becomes worse as the number of controllers increases.

Using *scale-up* increases the operational burden as the environment grows, and the result is an unbalanced and difficult-to-manage environment. Technology refresh cycles require substantial planning in advance, costly outages, and configuration changes, all of which introduce risk into the system.

With *scale-out*, as the storage environment grows, additional controllers are added seamlessly to the resource pool residing on a shared storage infrastructure. Scale-out, together with built-in storage virtualization, provides nondisruptive movement of host and client connections, as well as the datastores themselves, anywhere in the resource pool. With these capabilities, new workloads can be easily deployed and existing workloads can be easily and nondisruptively balanced over the available resources. Technology refreshes (replacing disk shelves, adding or completely replacing storage controllers) are accomplished while the environment remains online and serving data.

Although scale-out architecture has been available for some time, it is not in itself an automatic panacea for all of an enterprise's storage requirements. Many existing scale-out products are characterized by one or more of the following shortcomings:

- Limited protocol support; for example, NAS only
- Limited hardware support: support for only a particular type or a very limited set of storage controllers
- Upgrades dictate scaling in all dimensions based on the available controller configurations, so that capacity, compute power, and I/O all need to be increased, even if only a subset of these is required
- Little or no storage efficiency, such as thin provisioning, deduplication, compression

- Limited or no data replication capabilities
- Limited flash support

Therefore, although these products may be well positioned for certain specialized workloads, they are not flexible, capable, or robust enough for broad deployment throughout the enterprise.

NetApp clustered Data ONTAP is the first product offering a complete scale-out solution - an intelligent, adaptable, always-available storage infrastructure, utilizing proven storage efficiency for today's highly virtualized environments.

1.2 Multiprotocol Unified Architecture

Multiprotocol unified architecture is the ability to support multiple data access protocols concurrently in the same storage system, over a whole range of different controller and disk storage types. Data ONTAP 7G and 7-Mode have long been capable of this, and now clustered Data ONTAP supports an even wider range of data access protocols. The supported protocols in clustered Data ONTAP 8.2 are:

- NFS v3, v4, and v4.1, including pNFS
- SMB 1, 2, 2.1, and 3, including support for nondisruptive failover in Microsoft® Hyper-V™ enviroments with SMB 3
- iSCSI
- Fibre Channel
- FCoE

Data replication and storage efficiency features are seamlessly supported across all protocols in clustered Data ONTAP.

SAN Data Services

With the supported SAN protocols (Fibre Channel, FCoE, and iSCSI), clustered Data ONTAP provides LUN services; that is, the ability to create and map LUNs to attached hosts. Because the cluster consists of multiple controllers, there are multiple logical paths to any individual LUN, managed by multipath I/O (MPIO). Asymmetric Logical Unit Access (ALUA) is used on the hosts to make sure that the optimized path to a LUN is selected and made active for data transfer. If the optimized path to any LUN changes (for example, because the containing volume is moved), this is automatically recognized and nondsruptively adjusted by clustered ONTAP. If the optimized path becomes unavailable, clustered ONTAP can nondisruptively switch to any other available path.

NAS Data Services

With the supported NAS protocols, SMB (CIFS) and NFS, clustered Data ONTAP offers a namespace that provides access to a very large data container via a single NFS mountpoint or SMB share. NAS clients can mount a single NFS file system mountpoint or access a single SMB share, via their standard NFS or SMB client code – no additional executables need to be installed in order to access the namespace. Internally within clustered ONTAP, the namespace is composed of potentially thousands of volumes junctioned together by the cluster administrator. To the NAS clients, each volume appears as a folder, subdirectory, or junction, attached within the hierarchy of the file system mountpoint or share. Volumes can be added at any time and are immediately available to the clients, with no remount required for visibility to the new storage. The underlying infrastructure of volume boundaries is essentially transparent to clients as they move about in the file system.

As just described, clustered ONTAP offers the ability for NAS clients to mount the entire namespace at the root as a single NFS export or share. However, for many applications and environments, it is preferable to mount or access the namespace below the root so that only a subset view is accessible. This is intrinsically supported in clustered ONTAP; clients can mount the namespace at the root or at any volume junction path, subdirectory, or qtree. In each case, the clients see only the volumes that are mounted below their access point.

Clustered ONTAP also supports Infinite Volume, which is a cost-effective large container that can grow to petabytes of storage and billions of files. See “Infinite Volume” in section 3.2, “Logical Cluster Components.”

Although clustered ONTAP can be architected to provide a single namespace, additional namespaces can be configured to accommodate requirements for multi-tenancy or isolation of particular sets of clients or applications. Clustered Data ONTAP is therefore a platform for one or more Storage Virtual Machines (SVMs). Each SVM can support any or all of the supported client and host protocols. For details, see section 3, “Cluster Virtualization and Multi-Tenancy Concepts.”

1.3 Nondisruptive Operations

Shared storage infrastructures in today’s 24/7 environments provide services to thousands of individual clients or hosts and support many diverse applications and workloads across multiple business units or tenants. In such environments, downtime is not an option; storage infrastructures must be always on.

Nondisruptive operations (NDO) in clustered ONTAP are intrinsic to its innovative scale-out architecture. NDO is the ability of the storage infrastructure to remain up and serving data through the execution of hardware and software maintenance operations, as well as during other IT lifecycle operations. The goal of NDO is to eliminate downtime, whether preventable, planned, or unplanned, and to allow changes to the system to occur at any time.

Clustered ONTAP is highly available by design and can transparently migrate data as well as logical client connections throughout the storage cluster. NetApp DataMotion™ for Volumes is standard and built in to clustered ONTAP. It is the ability to nondisruptively move individual data volumes, allowing data to be redistributed across a cluster at any time and for any reason. DataMotion for Volumes is transparent and nondisruptive to NAS and SAN hosts so that the storage infrastructure continues to serve data throughout these changes. Data migration might be performed to rebalance capacity usage, to optimize for changing performance requirements, or to isolate one or more controllers or storage components to execute maintenance or lifecycle operations.

Table 1 describes hardware and software maintenance operations that can be performed nondisruptively in a clustered Data ONTAP environment.

Table 1) Nondisruptive hardware and software maintenance operations.

Operation	Details
Upgrade software	Upgrade from one version of Data ONTAP to another
Upgrade firmware	System, disk, switch firmware upgrade
Replace a failed controller or a component within a controller	For example, NICs, HBAs, power supplies, and so on
Replace failed storage components	For example, cables, drives, I/O modules, and so on

Table 2 describes lifecycle operations that can be performed nondisruptively in a clustered Data ONTAP environment.

Table 2) Nondisruptive lifecycle operations.

Operation	Details
Scale storage	Add storage (shelves or controllers) to a cluster and redistribute volumes for future growth
Scale hardware	Add hardware to controllers to increase scalability, performance, or capability (HBAs, NICs, NetApp Flash Cache™, NetApp Flash Pool™)
Refresh technology	Upgrade storage shelves, storage controllers, cluster-interconnect switch
Rebalance controller performance and storage utilization	Redistribute data across controllers to improve performance

The Immortal Cluster

NetApp clustered Data ONTAP can effectively deliver the *immortal cluster:* Software updates and configuration changes occur throughout any system's lifecycle. Above and beyond this, in almost all environments, the hardware infrastructure must be added to and replaced, potentially many times. Many years after the system is originally commissioned, the data has outlived the hardware, so that little or none of the original hardware might remain. Through the NDO capabilities, all of these changes can be achieved without outage or impact on the applications or attached clients and hosts—the cluster entity has persisted intact.

2 Clustered Data ONTAP Architecture

This section describes the architecture of clustered Data ONTAP, with an emphasis on the separation of physical resources and virtualized containers. Virtualization of storage and network physical resources is the basis for scale-out and nondisruptive operations.

2.1 Hardware Support and Basic System Overview

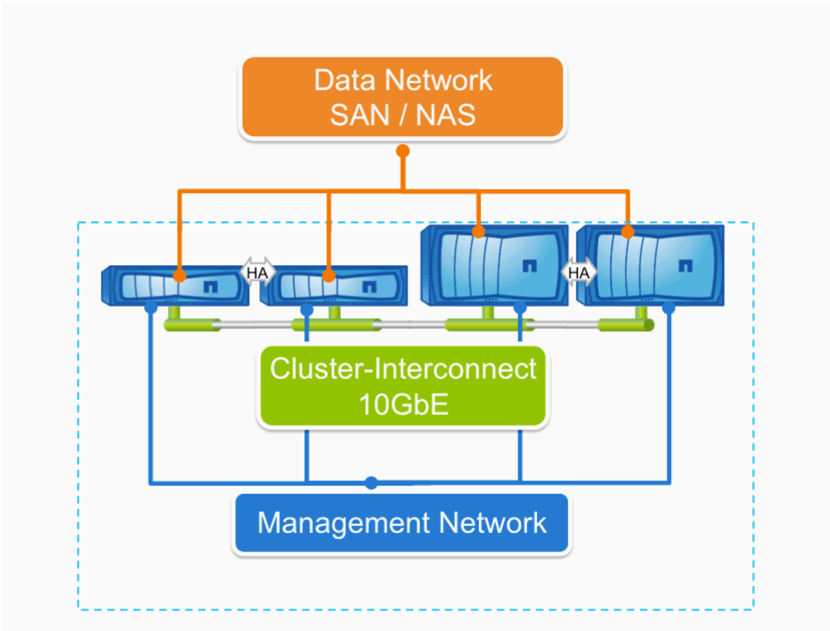
As shown in Figure 1, a clustered ONTAP system consists of NetApp storage controllers (including V-Series) with attached disks. The basic building block is the high-availability (HA) pair, a concept familiar from Data ONTAP 7G and 7-Mode environments. An HA pair consists of two identical *nodes,* or instances of clustered ONTAP. Each node actively provides data services and has redundant cabled paths to the other node's disk storage. If either node is down for any reason, planned or unplanned, its HA partner can take over its storage and maintain access to the data. When the downed system rejoins the cluster, the partner node gives back the storage resources.

The minimum cluster size starts with 2 matching nodes in an HA pair. Using nondisruptive technology refresh, a 2-node, entry-level cluster can evolve to the largest cluster size and most powerful hardware. Clusters with SAN protocols are supported up to 8 nodes with mid and high-end controllers. NAS-only clusters of high-end controllers scale up to 24 nodes and over 69PB of data storage.

Note: Clustered Data ONTAP 8.2 offers the additional option of a *single-node cluster* configuration. This is intended for smaller locations that replicate to a larger data center.

Note: The term *cluster* has been used historically to refer to an HA pair running Data ONTAP 7G or 7-Mode. This usage has been discontinued, and *HA pair* is the only correct term for this configuration. The term *cluster* now refers only to a configuration of one or more HA pairs running clustered Data ONTAP.

Figure 1) Clustered Data ONTAP overview.



One of the key differentiators in a clustered ONTAP environment is that the storage nodes are combined into a cluster to form a shared pool of physical resources that are available to applications, SAN hosts, and NAS clients. The shared pool appears as a single system image for management purposes, providing a single common point of management, through GUI or CLI tools, for the entire cluster.

Scalability

Clustered Data ONTAP allows the inclusion of different controller types in the same cluster, protecting the initial hardware investment and giving the flexibility to adapt resources to meet the business demands of the workloads. Similarly, support for different disk types, including SAS, SATA, and SSD, makes it possible to deploy integrated storage tiering for different data types, together with the transparent DataMotion capabilities of clustered ONTAP. Flash Cache cards can also be used to provide accelerated read performance for frequently accessed data. Starting with clustered ONTAP 8.1.1, Flash Pool intelligent caching is supported, which combines solid state disk (SSD) with traditional hard drives for optimal performance and efficiency using virtual storage tiering. The highly adaptable clustered ONTAP architecture is key to delivering maximum, on-demand flexibility for the shared IT infrastructure, offering flexible options to address needs for performance, price, and capacity.

Clustered ONTAP can scale both vertically and horizontally via the addition of nodes and storage to the cluster. This scalability, combined with protocol-neutral, proven storage efficiency, can meet the needs of the most demanding workloads.

2.2 Clustered Data ONTAP Networking

Figure 1 also shows the underlying network architecture of clustered Data ONTAP. Three networks are shown:

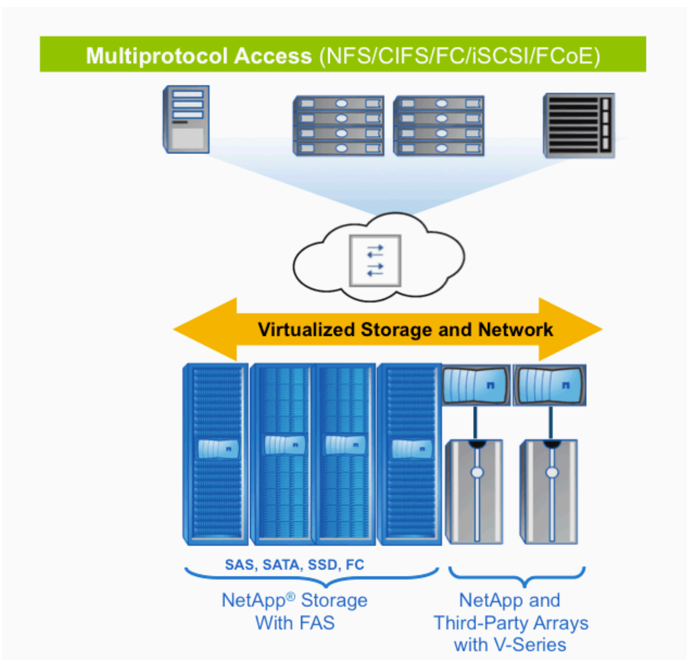
- **Cluster-interconnect.** A private, dedicated, redundant network used for communication between the cluster nodes and for DataMotion data migration within the cluster. The cluster-interconnect infrastructure is provided with every clustered ONTAP configuration to support this network. This infrastructure takes the form of redundant, high-performance, high-throughput 10Gbps enterprise-class switch hardware in clusters of four or more nodes. Clusters of two nodes can be optionally configured without switches, with point-to-point connections used for the cluster-interconnect. This configuration is available for the first time in clustered ONTAP 8.2 and is known as a *switchless cluster.* This entry-level configuration gives all the benefits of clustered ONTAP with a simpler infrastructure. Switchless clusters can be nondisruptively upgraded to include a switched cluster-interconnect when the cluster grows beyond two nodes.

- **Management network.** All management traffic passes over this network. Management network switches can be included as part of a clustered ONTAP configuration, or customer-provided switches can be used.

  NetApp OnCommand® System Manager, OnCommand Unified Manager, and other NetApp applications are available for management, configuration, and monitoring of clustered ONTAP systems. System Manager provides GUI management, including a number of easy-to-use wizards for common tasks. Unified Manager provides monitoring and alerts. A powerful CLI is included and ZAPIS are packaged and distributed in the Manage ONTAP® Software Developer's Kit.
- **Data networks.** Provide data access services over Ethernet or Fibre Channel to the SAN hosts and NAS clients. These networks are provided by the customer according to requirements and could also include connections to other clusters acting as volume replication targets for data protection.

Figure 2 shows a larger cluster with different disk types, and a mix of native NetApp FAS and V-Series controllers. V-Series allows you to use third-party storage as the front end with a NetApp controller, so that it can run clustered ONTAP and participate in a cluster. It also shows the client/host connections and the virtualized storage and network layer. For more information, see section 3, "Cluster Virtualization and Multi-Tenancy Concepts.

Figure 2) Clustered Data ONTAP large cluster.



2.3 Storage Efficiency and Data Protection

Storage efficiency built into clustered Data ONTAP offers substantial space savings, allowing more data to be stored at lower cost. Data protection provides replication services, making sure that valuable data is backed up and recoverable.

- **Thin provisioning.** Volumes are created by using *virtual sizing.* Thin provisioning is the most efficient way to provision storage, because although the clients see the total storage space assigned to them, the storage is not preallocated up front. In other words, when a volume or LUN is created by using thin provisioning, no space on the storage system is used. The space remains unused until data is written to the LUN or the volume, at which time only enough space to store the data is used. Unused storage is shared across all volumes, and the volumes can grow and shrink on demand.
- **NetApp Snapshot™ copies.** Automatically scheduled point-in-time copies that take up no space and incur no performance overhead when created. Over time, Snapshot copies consume minimal storage space, because only changes to the active file system are written. Individual files and directories can be easily recovered from any Snapshot copy, and the entire volume can be restored back to any Snapshot state in seconds.
- **NetApp FlexClone® volumes.** Near-zero space, exact, writable virtual copies of datasets. They offer rapid, space-efficient creation of additional data copies ideally suited for test/dev environments.
- **NetApp FlexCache® volumes.** Provide the ability to cache volumes on other nodes in the cluster, thus balancing the read load to a frequently accessed volume across the cluster. The cache volumes are space efficient, because only the blocks accessed are cached. Data consistency is maintained via read delegation for files
- **Deduplication.** Removes duplicate data blocks in primary and secondary storage, storing only unique blocks, which results in storage space and cost savings. Deduplication runs on a customizable schedule.
- **Compression.** Compresses data blocks by replacing repeating patterns within a subset of a file. Compression is complementary with deduplication; depending on the workload, compression only, deduplication only, or deduplication and compression together may provide the maximum space and cost savings.
- **NetApp SnapMirror®.** Asynchronous replication of volumes, independent of protocol, either within the cluster or to another clustered ONTAP system for data protection and disaster recovery.
- **NetApp SnapVault®**. Volumes can be copied for space-efficient, read-only, disk-to-disk backup either within the cluster or to another clustered ONTAP system.

3 Cluster Virtualization and Multi-Tenancy Concepts

A cluster is composed of physical hardware: storage controllers with attached disk shelves, network interface cards, and optionally Flash Cache cards. Together these components create a physical resource pool, which is virtualized as logical cluster resources to provide data access. Abstracting and virtualizing physical assets into logical resources provides the flexibility and potential multi-tenancy in clustered Data ONTAP as well as the object mobility capabilities that are the heart of nondisruptive operations.

3.1 Physical Cluster Components

Storage controllers, independent of the model, are considered equivalent in the cluster configuration, in that they are all presented and managed as *cluster nodes***.** Clustered Data ONTAP is a symmetrical architecture, with all nodes performing the same data-serving function.

Individual disks are managed by defining them into *aggregates:* groups of disks of a particular type that are protected by using NetApp RAID-DP®, similar to 7G and 7-Mode.

Network interface cards and HBAs provide physical *ports* (Ethernet and Fibre Channel) for connection to the management and data networks shown in Figure 2.

The physical components are visible to cluster administrators but not directly to the applications and hosts that are using the cluster. The physical components provide a pool of shared resources from which the logical cluster resources are constructed. Applications and hosts access data only through Storage Virtual Machines (SVMs) that contain volumes and logical interfaces.

3.2 Logical Cluster Components

The primary logical cluster component is the Storage Virtual Machine; all client and host data access is via an SVM. Clustered ONTAP supports a minimum of one and up to hundreds of SVMs in a single cluster. Each SVM is configured for the client and host access protocols it supports – any combination of SAN and NAS. Each SVM contains at least one volume and at least one logical interface. Administration of each SVM can optionally be delegated, so that separate administrators are responsible for provisioning volumes and other SVM-specific operations. This is particularly appropriate for multi-tenant environments or where workload separation is desired. An SVM delegated administrator would have visibility only to their specific SVM, and have no knowledge of any other hosted SVM.

For NAS clients, the volumes in each SVM are junctioned together into a namespace for CIFS and NFS access. For SAN hosts, LUNs are defined within volumes and mapped to hosts, as described in section 1.2, "Multiprotocol Unified Architecture." A special type of SVM, known as *Infinite Volume,* is described at the end of this section.

The accessing hosts and clients connect to the SVM via a *logical interface* (LIF). LIFs present either an IP address (used by NAS clients and iSCSI hosts) or a WWPN (World Wide Port Name, for FC and FCoE access). Each LIF has a home port on a NIC or an HBA. LIFs are used to virtualize the NIC and HBA ports rather than mapping IP addresses or WWPNs directly to the physical ports, because there are almost always many more LIFs than physical ports in a cluster. Each SVM requires its own dedicated set of LIFs, and up to 128 LIFs can be defined on any cluster node. A LIF defined for NAS access can be temporarily migrated to another port on the same or a different controller to preserve availability, to rebalance client performance, or to evacuate all resources on a controller for hardware lifecycle operations.

Figure 3 shows a single SVM in a two-node cluster providing data services to SAN hosts and NAS clients. Each volume, shown by the orange circles, is provisioned on an aggregate on a cluster node, and the combination of all the volumes constitutes the entire namespace or resource pool for LUNs. Volumes can be moved nondisruptively at any time from any aggregate to any aggregate as required. Delegated SVM administrators can provision volumes only in their own SVMs; these administrators have no visibility to any other SVM, or even awareness that other SVMs exist. The delegated SVM administrator cannot perform volume moves around the cluster, because this operation affects the capacity of aggregates shared by other SVMs. For this reason, only a cluster administrator can move volumes.

If SVM administration has been delegated, the cluster administrator must explicitly specify the aggregates available to the SVM administrator for provisioning volumes. This offers a mechanism for SVMs to provide different classes of service—for example, an SVM could be restricted to using only aggregates with SSD or SATA drives or only aggregates on a particular subset of controllers.

Figure 3) Cluster with a single SVM.

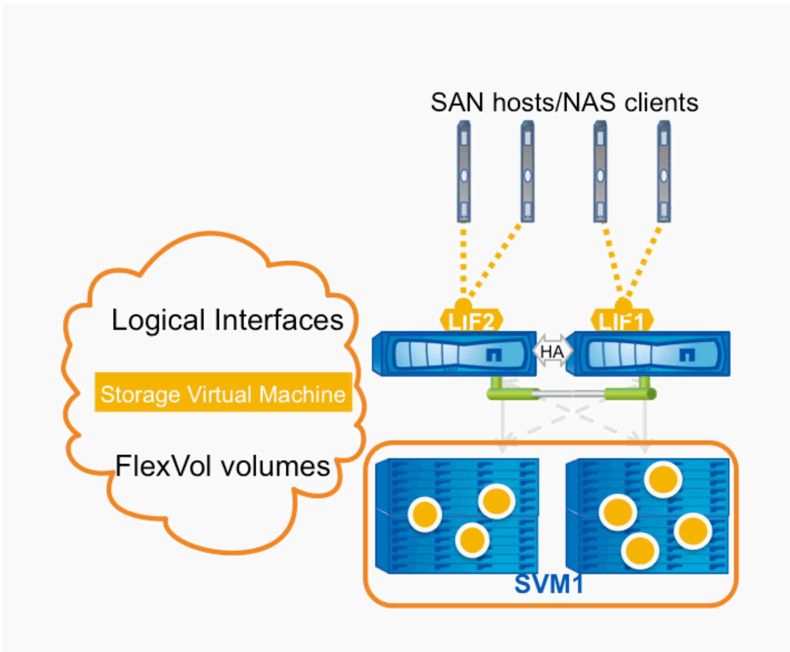
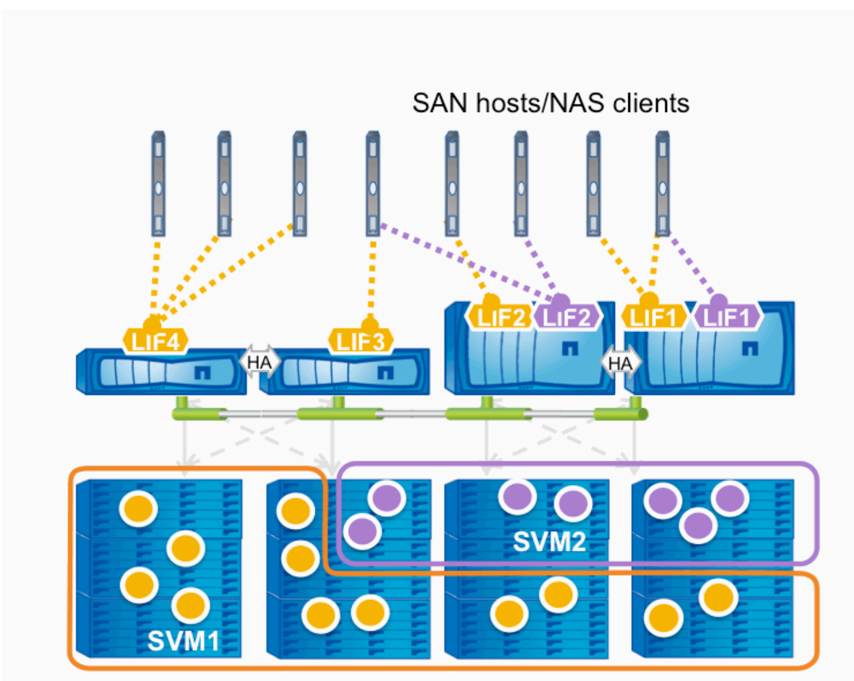


Figure 4 shows a more complex environment. The cluster here consists of four nodes, with two SVMs providing data access. Each SVM consists of different volumes and LIFs, for secure compartmentalized access. Although the volumes and LIFs in each SVM share the same physical resources (network ports and storage aggregates), a host or client can access the data in SVM1 only through a LIF defined in that SVM, and similarly for SVM2. Administrative controls make sure that a delegated administrator with access to SVM1 has visibility only to the logical resources assigned to that SVM, and a SVM2-delegated administrator similarly sees only SVM2's resources.

Figure 4) Cluster with multiple SVMs.



By virtualizing physical resources into the SVM construct, clustered ONTAP implements multi-tenancy and scale-out, allowing the cluster to host isolated independent workloads and applications.

Storage QoS

Clustered Data ONTAP 8.2 provides Storage QoS (quality of service) policies on cluster objects. An entire SVM, or a group of volumes or LUNS within an SVM, can be dynamically assigned to a policy group, which specifies a throughput limit, defined in terms of IOPS or MB/sec. This can be used to reactively or proactively throttle rogue workloads and prevent them from affecting the rest of the workloads. QoS policy groups can also be used by service providers to prevent tenants from affecting each other, as well as to avoid performance degradation of the existing tenants when a new tenant is deployed on the shared infrastructure.

Infinite Volume

Infinite Volume is a type of volume that is contained in a dedicated SVM, which can scale up to 20PB and store up to 2 billion files. In clustered ONTAP 8.2, Infinite Volumes can coexist with standard SVMs and support both NFS and SMB client access. Infinite Volumes are well suited for enterprise NAS content repositories. For more information, see TR4037 Introduction to NetApp Infinite Volume.

4 SUMMARY

This technical report provides an overview of NetApp clustered Data ONTAP 8.2 and shows how it incorporates industry-leading unified architecture, nondisruptive operations, proven storage efficiency, and seamless scalability.

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.

Go further, faster®

NetApp®

www.netapp.com