Technical Report

Clustered Data ONTAP 8.1 and 8.1.1: An Introduction

Charlotte Brooks, NetApp
January 2013 | TR-3982

SUMMARY

This technical report provides an introduction to the architecture and key customer benefits of clustered Data ONTAP® 8.1 and 8.1.1. Throughout this paper, unless otherwise specified, the term “Data ONTAP” in isolation refers to clustered Data ONTAP. If a reference to 7-Mode or 7G is required, it will be specifically stated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 CLUSTERED DATA ONTAP 8.1: ENTERPRISE NONDISRUPTIVE OPERATIONS

With the release of clustered Data ONTAP 8.1, NetApp brings to market for the first time, enterprise-ready***, unified scale-out storage***. Developed from a solid foundation of proven Data ONTAP technology and innovation, clustered ONTAP is the basis for large virtualized shared storage infrastructures that are architected ***for nondisruptive operations*** over the system lifetime.

This paper provides an overview of clustered Data ONTAP, including its architecture and core capabilities, and positions it firmly in the 21st-century data center. The latest released version, clustered Data ONTAP 8.1.1 adds additional scalability, improved integration of SSD for virtual storage tiering and an option for a big data content repository solution.

1.1 SCALE-OUT

Scale-out is a way to respond to growth in a storage environment. All storage controllers have physical limits to their expandability: 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 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 another controller. One way to do this is to "scale up": that is, each additional controller is a completely independent management entity, which does not provide any shared storage resources. If the original controller is to be completely replaced by the newer and larger controller, data migration is required to transfer the data from the old to the new. This is time-consuming and potentially disruptive and likely requires configuration changes on all the attached host systems.

If the newer controller will coexist with the original controller, there are now two storage controllers to be individually managed, and there are no inbuilt tools to balance or reassign workloads across them. The situation becomes worse as the number of controllers increases. By using scale-up, the operational burden increases consistently as the environment grows, and the end result is a very unbalanced and difficult-to-manage environment. Technology refresh cycles require substantial planning in advance, lengthy outages, and configuration changes, which will introduce risk into the system.

By contrast, using ***scale-out*** means that as the storage environment grows, additional controllers are added seamlessly to the resource pool residing on a shared storage infrastructure. Host and client connections as well as datastores can move seamlessly and nondisruptively anywhere in the resource pool, so that existing workloads can be easily balanced over the available resources, and new workloads can be easily deployed. Technology refreshes (replacing disk shelves, adding or completely replacing storage controllers) are accomplished while the environment remains online and serving data.

While scale-out products have been available for some time, these were typically subject to one or more of the following shortcomings:

- Limited protocol support: NAS only
- Limited hardware support: only supported a particular type of storage controller or a very limited set
- Little or no storage efficiency: thin provisioning, deduplication, compression
- Little or no data replication capabilities

Therefore, while these products are well positioned for certain specialized workloads, they are less flexible, capable, and robust for broad deployment throughout the enterprise.

Clustered Data ONTAP is the first vendor product to offer a complete scale-out solution and offers an adaptable, always-available storage infrastructure 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 overall 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 8.1 also supports a full range of data access protocols. The supported protocols are:

- NFS v3, v4, v4.1, including pNFS
- SMB 1 and 2
- iSCSI
- Fibre Channel
- FCoE

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

SAN DATA SERVICES

With the supported SAN protocols (Fibre Channel, FCoE, and iSCSI), clustered ONTAP provides LUN services: that is, the ability to create and make available LUNs to attached hosts. Since the cluster consists of multiple controllers, there will be multiple logical paths to any individual LUN, and the best practice is to configure at least one path per node in the cluster. 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.

NAS DATA SERVICES

With the supported NAS protocols, CIFS (SMB) and NFS, clustered ONTAP can provide a single namespace: that is, NAS clients can access a very large data container using a single NFS mountpoint or CIFS share. Each client, therefore, needs only to mount a single NFS file system mountpoint or access a single CIFS share, and only the standard NFS and CIFS client code for each operating system is required. 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 or subdirectory, nested off the root of the NFS file system mountpoint or CIFS share. Volumes can be added at any time and will be immediately available to the clients, with no remount required for visibility to the new storage. Furthermore, the clients have no awareness that they are traversing volume boundaries as they move about in the file system, since the underlying structure is completely transparent.

While clustered ONTAP can be architected to provide a single namespace, it also supports the concept of multiple, securely partitioned namespaces to accommodate requirements for multi-tenancy or isolation of particular sets of clients or applications. See section 2.3 for more information.

1.3 NONDISRUPTIVE OPERATIONS

Shared storage infrastructures in today's 24x7 environments provide data 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 no longer an option; storage infrastructures must be always on.

Nondisruptive operations (NDO) in clustered ONTAP are intrinsic to its unique scale-out architecture. NDO is the ability for 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 and network connections anywhere within the storage cluster. The ability to move individual data volumes, known as ***data motion for volumes***, allows data to be redistributed across a cluster at any time and for any reason. Data motion is transparent and nondisruptive to NAS and SAN hosts and enables the storage infrastructure to continue to serve data throughout these changes. Data motion might be performed in order to rebalance capacity usage, to optimize for changing performance requirements, or to isolate one or more controllers or storage components when it becomes necessary to execute maintenance or lifecycle operations.

Hardware and software maintenance operations that can be performed nondisruptively in a clustered Data ONTAP 8.1 environment are listed in Table 1.

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 failed controller or 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

Lifecycle operations that can be performed nondisruptively in a clustered Data ONTAP 8.1 environment are listed in Table 2.

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, Flash Cache, Flash Pool)
Refresh technology	Upgrade storage shelves, storage controllers, back-end switch
Rebalance controller performance and storage utilization	Redistribute data across controllers to improve performance
Rebalance capacity	Redistribute data across controllers to account for future capacity growth
Rebalance disk performance and utilization	Redistribute data across storage tiers within a cluster to optimize disk performance

THE IMMORTAL CLUSTER

Clustered Data ONTAP can effectively deliver an ***immortal cluster***: Over time in most environments, in addition to the typical software updates and configuration changes through the system's lifecycle, the hardware infrastructure will 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 these changes would have been achieved without any outage and at no impact to any of the applications or attached clients and hosts: the cluster entity has persisted intact.

2 CLUSTERED ONTAP ARCHITECTURE

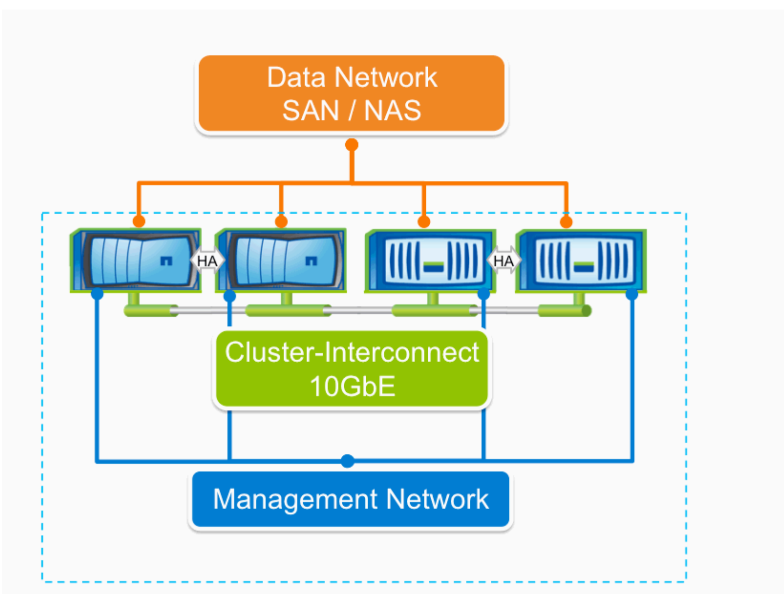
This section describes the architecture of clustered ONTAP, with an emphasis on separation of physical resources and virtualized containers. Virtualization of storage and network physical resources is the basis for scale-out and nondisruptive operations.

2.1 HARDWARE SUPPORT AND BASIC SYSTEM OVERVIEW

As shown in Figure 1, a clustered ONTAP system consists of 2 or more individual NetApp® storage controllers (including V-Series) with attached disks. The basic building block is the HA pair, a term familiar from Data ONTAP 7G or 7-Mode environments. An HA pair consists of 2 identical controllers; each controller actively provides data services and has redundant cabled paths to the other controller's disk storage. If either controller is down for any planned or unplanned reason, its HA partner can take over its storage and maintain access to the data. When the downed system rejoins the cluster, the partner will give back the storage resources.

Note: The term ***cluster*** has been used historically to refer to an HA pair running Data ONTAP 7G or 7-Mode. This usage is discontinued, and HA pair is the only correct term for this. 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 multiple HA pairs are combined together into a cluster to form a shared pool of physical resources available to applications, SAN hosts, and NAS clients. The shared pool appears as a single system image for management purposes. This means there is a single common point of management, whether through GUI or CLI tools, for the entire cluster. While the members of each HA pair must be the same controller type, the cluster can consist of heterogeneous HA pairs. Over time, as the cluster grows and new controllers are released, it is likely to evolve into a combination of several different node types. All cluster capabilities are supported, regardless of the underlying controllers in the cluster.

SCALABILITY

Clustered Data ONTAP allows the inclusion of a wide variety of controller types in the same cluster, protecting protecting the initial hardware investment and giveing 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 data motion capabilities of clustered ONTAP. Flash Cache cards can also be used to provide accelerated read performance for frequently accessed data. In the clustered ONTAP 8.1.1release , Flash Pools are supported, which combine 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. High individual node capacities (for example, 4.3PB maximum storage on high-end controller HA pairs) mean a cluster can scale to a large number of petabytes. This scalability, combined with protocol-neutral storage efficiency can meet the needs of the most demanding workloads.

2.2 NETWORKING IN CLUSTERED ONTAP 8.1

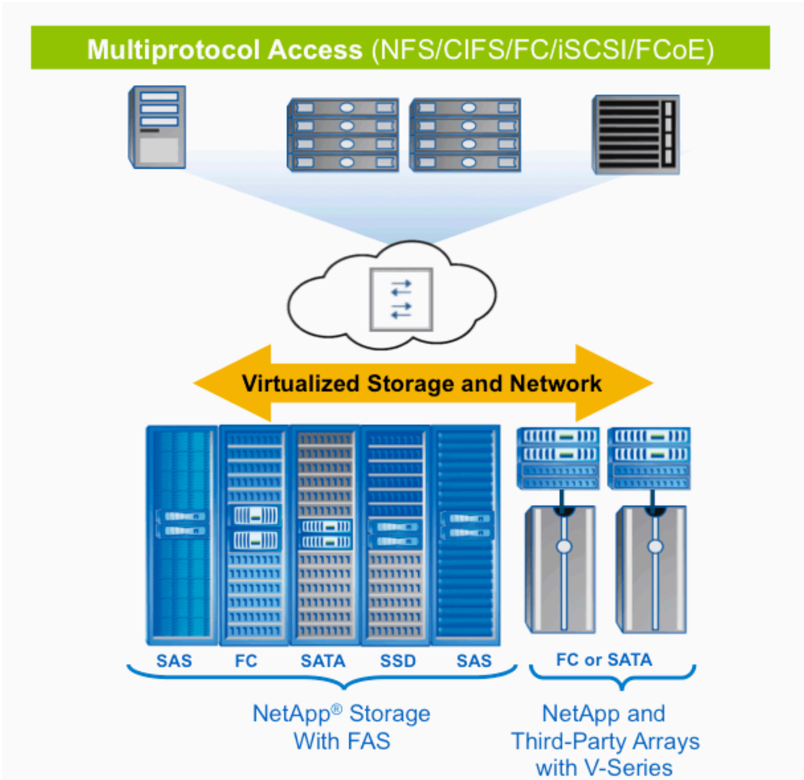
Figure 1 also shows the underlying network architecture of clustered Data ONTAP. Three networks are shown:

- **Cluster interconnect.** A 10Gbps, private, dedicated, redundant, high-throughput network used for communication between the cluster nodes and for data motion within the cluster. The cluster interconnect infrastructure is provided with every clustered ONTAP 8.1 configuration to support this network.
- **Management network.** All management traffic passes over this network. Management network switches can be included in a clustered ONTAP 8.1 configuration, or customer provided switches can be used.

  OnCommand™ System Manager 2.0 and OnCommand Unified Manager 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. Univied Manager provides monitoring and alerts. In addition, a CLI and set of ZAPIs (distributed as an SDK) are available.
- **Data networks.** Provide data access services over Ethernet or Fibre Channel to the SAN hosts and NAS clients. These networks are customer provided according to requirements and could also include connections to other clusters acting as volume replication targets for data protection.

Figure 2 shows a large, heterogeneous cluster, consisting of different controller types, different disk types, and a mix of native NetApp FAS and V-Series controllers. V-Series allows you to front-end third-party storage 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. This will be explained in the next section.

Figure 2) Clustered Data ONTAP.



2.3 STORAGE EFFICIENCY AND DATA PROTECTION

Storage efficiency built into clustered ONTAP provides 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 using "virtual" sizing. They appear to be provisioned at their full capacity, but are actually created much smaller and only use additional space when actually needed. Extra unused storage is shared across all volumes, and the volumes can grow and shrink on demand.
- **Snapshot™ copies.** Automatically scheduled point-in-time copies that write only changed blocks, with no performance penalty. The Snapshot copies consume minimal storage space, since 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.
- **FlexClone® volumes.** Near-zero space, instant “virtual” copies of datasets. The clones are writable, but only changes to the original are stored, so they provide rapid, space-efficient creation of additional data copies ideally suited for test/dev environments.
- **Deduplication.** Removes redundant data blocks in primary and secondary storage with flexible policies on when the deduplication process is run.
- **Compression.** Compresses data blocks. Can be run whether or not deduplication is enabled and can provide additional space savings whether run alone or together with compression.
- **SnapMirror®.** Volumes can be asynchronously replicated either within the cluster or to another cluster.

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 create a physical resource pool, which is virtualized as logical cluster resources to provide data access. Abstracting and virtualizing physical assets into logical resources provide the flexibility and potential multi-tenancy in clustered ONTAP as well as the data motion ability at the heart of nondisruptive operations.

3.1 PHYSICAL CLUSTER COMPONENTS

Storage controllers, while they can be of different types, are by default considered equivalently in the cluster configuration in that they are all presented and managed as ***cluster nodes***.

Individual disks are managed by defining them into ***aggregates***: groups of disks of a particular type that are protected 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 described in Figure 2.2.

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

3.2 LOGICAL CLUSTER COMPONENTS

The primary logical cluster component is the Virtual Storage Server, known as Vserver. Clustered ONTAP supports from one to hundreds of Vservers in a single cluster. Each Vserver is configured for the client and host access protocols it will support – any combination of SAN and NAS. Each Vserver contains at least one volume and at least one logical interface. The administration of each Vserver can also be delegated if desired, so that separate administrators could be responsible for provisioning volumes and other Vserver–specific operations. This is particularly appropriate for multi-tenanted environments or where workload separation is desired.

For NAS clients, the volumes in each Vserver are junctioned together into a namespace for CIFS and NFS access. For SAN hosts, LUNs are defined in the volumes and mapped to hosts as described in section 1.2.

The accessing hosts and clients connect to the Vserver using a ***logical interface*** (or LIF). LIFs present either an IP address (which will be used by NAS clients and iSCSI hosts) or a WWN (for FC and FCoE access). Each LIF has a home port on a NIC or HBA. LIFs are used to virtualize the NIC and HBA ports rather than mapping IP addresses or WWNs directly to the physical ports, because there will almost always be many more LIFs than physical ports in a cluster. Each Vserver 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 temporarily move or migrate to another port on the same or a different controller to preserve availability or to rebalance client performance.

Figure 3 shows a single Vserver that is 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. By default, volumes in a Vserver can be created in any of the defined aggregates and moved at any time from aggregate to aggregate as required. Delegated Vserver administrators can provision volumes in their own Vservers. The delegated Vserver administrator cannot, however, initiate the movement of volumes around across the cluster, since this might affect the entire cluster. For this reason, only a cluster administrator can move volumes.

Note: Optionally, a cluster administrator can restrict the aggregates that may be used to provision volumes in a particular Vserver for provisioning volumes. This allows a Vserver to provide different classes of service: for example, by authorizing the Vserver to use only aggregates consisting of SSD or SATA drives or only aggregates on a particular subset of controllers.

Figure 3) One Vserver in a cluster.

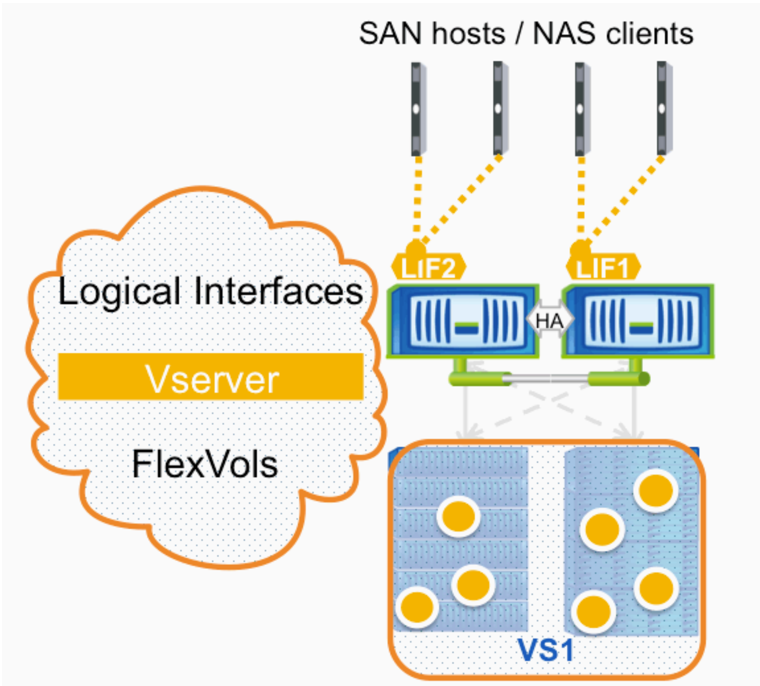
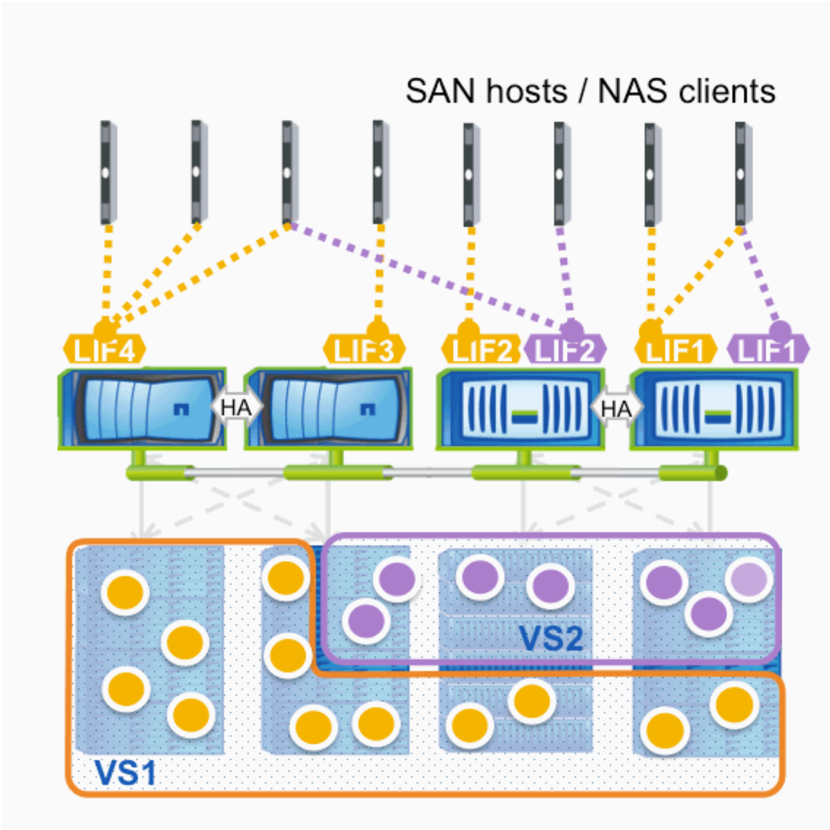


Figure 4 shows a more complex environment. There are 4 nodes in the cluster, and two Vservers providing SAN or NAS data access. Each Vserver consists of different volumes and LIFs, providing secure compartmentalized access. Although the volumes and LIFs in each Vserver share the same physical resources (network ports and storage aggregates), a host or client can only access the data in VS1 through a LIF defined in that Vserver, and similarly for VS2. Administrative controls make sure that a delegated administrator with access to VS1 can only see the logical resources assigned to that Vserver, and a VS2-delegated administrator similarly only sees VS2’s resources.

Figure 4) Multiple Vservers in the same cluster.



By virtualizing physical resources into the Virtual Storage Server (Vserver) construct, clustered ONTAP implements multi-tenancy and scale-out and allows a cluster to host many independent workloads and applications.

INFINITE VOLUME

Clustered Data ONTAP 8.1.1 can optionally be configured to provide a single very large volume (up to 20PB and 2 billion files) with NFSv3 client access, which is ideally suited for enterprise content repositories. For more information on the Infinite Volume, see TR4037 Introduction to NetApp Infinite Volume.

4 SUMMARY

This technical report has provided an overview of clustered Data ONTAP 8.1 and 8.1.1 and shown how it uniquely incorporates industry-leading unified architecture, nondisruptive operations, and storage efficiency in a scale-out architecture.

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.

Go further, faster®

NetApp®

www.netapp.com