Technical Report

# DataMotion for Volumes
## NetApp clustered Data ONTAP 8.2 and 8.3

Jay Bounds, NetApp

March 2015 | TR-4075

## Abstract

This technical report describes the operation of and best practices for the clustered Data ONTAP® `volume move` command known as DataMotion™ for Volumes.

DataMotion for Volumes is a basic,core component of nondisruptive operations within clustered Data ONTAP.

# Version History

| Version | Date | Document Version History |
|---------|------|--------------------------|
| Version 1.0 | April 2012 | Original publication |
| Version 2.0 | March 2014 | Update for clustered Data ONTAP 8.2 and Data ONTAP 8.3.x. Merged with TR-3075 |
| Version 3.0 | March 2015 | Updated for clustered Data ONTAP 8.3.x |

## TABLE OF CONTENTS

**LIST OF TABLES**

**LIST OF FIGURES**

# 1   Introduction

Today's business environments require 24/7 data availability. An enterprise's storage must deliver the base building block for IT infrastructures, providing data storage for all business applications and objectives. Therefore, constant data availability begins with architecting storage systems that facilitate nondisruptive operations (NDOs). Nondisruptive operations provide functionality in three principal areas: hardware resiliency, hardware and software lifecycle operations, and hardware and software maintenance operations. For the purpose of this paper we focus on NetApp® DataMotion for Volumes best practices and usage optimization as a solution for hardware and software lifecycle operations.

There are additional NetApp DataMotion capabilities in previously available versions of clustered Data ONTAP, including DataMotion for vFiler® and NetApp DataMotion for Volumes for Data ONTAP 7-Mode. These data mobility capabilities target specific use cases in 7-Mode and are based on different data replication implementations. This paper does not cover additional information for either of the 7-Mode data mobility solutions. For more information on these, see TR-3814, NetApp DataMotion for vFiler, or TR-3873, NetApp DataMotion for Volumes.

DataMotion for Volumes is the official product name for the movement of volumes in clustered Data ONTAP. For the purpose of this paper and for clarity where required, volume move and vol move are used interchangeably with the term clustered Data ONTAP DataMotion for Volumes.

Basic knowledge of the architecture and terminology of clustered Data ONTAP is assumed throughout this TR. Refer to TR-3982: NetApp Clustered Data ONTAP 8.3 and 8.2.x: An Introduction for more information.

## 1.1   Overview

This paper was written with two main goals: to explain how DataMotion works to nondisruptively move data across the physical storage resources within a cluster and to present common use cases and best practices for DataMotion for Volumes. The best practices presented are intended to be applied during common nondisruptive operations in order to optimize the impact and duration of a volume move job. Although each scenario and customer environment is unique, the common principles set forth in the paper are meant to enhance the user's understanding of the volume move feature and how it interacts within the NetApp ecosystem and to apply it to any given use case in which volume move will be performed.

The high-level topics in this paper are:

* Overview of DataMotion for Volumes
* Operation and phases of a volume move
* Common use cases for DataMotion for Volumes
* DataMotion for Volumes usage, caveats, and limitations
* Best practices during the different phases of volume move

## 1.2   DataMotion for Volumes Overview

NetApp DataMotion for Volumes is a native clustered Data ONTAP feature for data mobility and a key functional capability for nondisruptive operations of storage systems. As a core clustered Data ONTAP component, it is automatically enabled for all clusters; no additional license is required.

DataMotion for Volumes provides the capability to move a volume within a storage virtual machine (SVM, formerly known as a Vserver) from one aggregate, referred to as the source, to another aggregate, referred to as the destination. DataMotion for Volumes is nondisruptive for all supported protocols such as NFS, CIFS, FC, FCoE, and iSCSI. Regardless of the location of the volume before and after the move, access to the data remains the same for client/host applications. Put another way, vol move changes the

physical location of the volume (its aggregate) while leaving the logical location (position in the namespace, or paths to LUNs) unchanged. If you need to move a volume logically to another SVM so that it will be accessed through a different LIF and network path, SnapMirror® technology can be used to copy the data to the new SVM. This requires the client to remount via the new path and, therefore, incur a disruption.

Volume move utilizes native asynchronous replication technology to copy Snapshot™ copies containing the data content of the volume from the source aggregate to a new volume created on the destination aggregate. The transfer of data from a volume on the source aggregate to a volume on the destination aggregate is done over the cluster interconnect, except when source and destination aggregates reside on the same node. Figure 1 shows that the Snapshot copy is transferred from volume B to volume Bt over the cluster interconnect; the source and destination aggregates reside on different nodes. Once the copy is complete, volume Bt is renamed to the original volume B, preserving location transparency to the hosts and clients.

**Figure 1) DataMotion for Volumes replication process.**
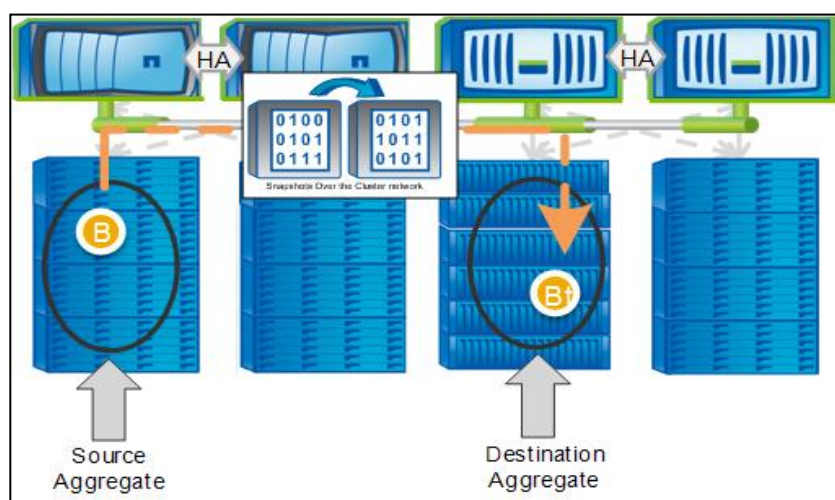


Before we get into the details of DataMotion for Volumes, it is important to understand which volume types are supported. Table 1 shows the currently supported volume types.

**Table 1) Volume type support for DataMotion for Volumes.**

| Volume Type | Supported |
|---|---|
| FlexVol® volumes | Yes – data volumes only, including SVM root volume. Node root volumes cannot be moved. |
| Data protection (DP) flexible volumes | Yes; see section 5.2 for more information |
| Load-sharing mirror volumes | No |
| FlexCache® volumes | No |
| Volumes in 32-bit and 64-bit aggregates | See Table 6 for more information on aggregate type compatibility |
| Infinite Volume | Available only at the constituent level with the assistance of NetApp Support; more information on Infinite Volume is on the Field Portal |
| Traditional volumes | No, not supported in clustered Data ONTAP |

# 2   How Volume Move Works

A volume move can be performed by a cluster administrator in any of the following ways. Volume move requires cluster administration privilege. More information on how to perform a vol move is in section 0.

- A single command, `volume move start`, commonly abbreviated as vol move.
- OnCommand® System Manager can initiate a volume move from either the Volume or Aggregate display panels.
- OnCommand Workflow Automation (WFA) canned workflow.
- Programmatically via ZAPI. See the NetApp Developer Community and the NetApp Manageability SDK Technical Overview for information and documentation on the API.

No matter how it is initiated, a volume move operation consists of a number of phases. Once the volume move is initiated, the progression of stages is automatic, but it is important to understand each phase and the underlying activities. There are five phases to the volume move process, detailed in the next section.

## 2.1   Phases of Volume Move

The different phases of the volume move job are shown in Table 2.

**Table 2) Phases of DataMotion for Volumes.**

| Volume Move Phase | Phase Description |
|---|---|
| 1. Validation Phase | In this phase, several checks occur to verify that a volume move is possible between the volume on the source aggregate and the selected destination aggregate. If any of the checks fails, the vol move job does not proceed and an appropriate error message is issued. |
| | The validation phase checks for sufficient capacity on the target aggregate so that it can contain the volume. It also checks that mutually exclusive operations are not in progress when the vol move is initiated (see section 5.2 for more information). Other compatibilities are also checked, for example, aggregate compatibility, since vol move from a 64-bit aggregate source to a 32-bit aggregate destination is not permitted in Data ONTAP 8.2 and 32-bit aggregates are no longer supported in Data ONTAP 8.3 and higher. See Table 6) Aggregate compatibility for DataMotion for Volumes. |
| 2. Setup Phase | In this phase, a new volume on the destination aggregate is created. |

| Volume Move Phase | Phase Description |
|---|---|
| 3. Iterative Phase | In this phase, data is replicated between the source volume and the destination volume by replicating groups of Snapshot copies over the cluster interconnect.<br><br>**Cutover Decision Point:** After each iteration of replication, the volume move job checks to see if the amount of changed data between the source and destination is small enough so that a final replication can be completed in the time defined for the cutover phase. If this decision point is successful, the volume move job moves to the cutover phase; if not, the iterative phase continues. Best practices for the iterative phase are discussed in section 6. |
| 4. Cutover Phase | In this phase, all I/O access is suspended, and I/O requests are blocked to the source volume. Further I/O requests are queued. The final replication transfer is completed and the cluster volume locator database is updated with the new volume information. Once the volume data is updated, the queued I/O resumes on the volume at the new location. The cutover must be able to complete before or at the maximum defined "cutover period" that is within an acceptable window of time for client/host application. Best practices for predictable cutover are discussed in section 7.1. If the cutover cannot complete within the defined cutover window, I/O resumes on the source and the cutover is retried after a short delay. |
| 5. Complete Phase | In this phase, if the cutover has completed successfully, the original volume on the source aggregate is removed after appropriate cleanup operations. The new volume on the destination aggregate then serves all client and host data. The new volume goes through a period of cleanup during which physical block location information is updated, referred to as the deswizzle portion of the complete phase. See section 8 for more information. |

## 2.2   DataMotion for Volumes as a Nondisruptive Operation

DataMotion for Volumes has strict guidance for the cutover phase and how the transition from the volume on the source aggregate to the volume on the destination is completed within a defined period to maintain transparency and nondisruptiveness to client/host applications. During the cutover period, data requests are queued at the node level until the volume cutover is complete. If the cutover cannot complete in the specified cutover time, then the cutover phase terminates. A failed cutover reverts data access to the volume on the source aggregate. Any pending requests that were queued during the cutover attempt are completed and the iterative phase of volume move resumes until the next cutover can be attempted, up to the number of cutover attempts specified.

The cutover period is specified with the `vol move -cutover-window` option. The cutover window can be defined between 30 and 300 seconds. The sensitivity of the cutover window depends on the protocols accessing the data. Some protocols and applications do not tolerate a cutover greater than 45 seconds while others can withstand a cutover of 120 seconds or more. The default value for the cutover window is 45 seconds.

Optionally, the `vol move` command can be issued such that it does not automatically attempt cutover, but remains in the iterative phase until the administrator explicitly triggers the cutover. Section 4.4 explains the cutover options in more detail.

Details regarding `vol move` command usage and best practices for specific environments are discussed in section 5.3.

Another important aspect in keeping vol move nondisruptive to normal cluster operation is automatic throttling of the replication on the initial full data transfer. This throttling automatically adjusts itself based on host/client traffic and other cluster workloads in order to minimize the impact of the vol move. More information on this throttling is in section 6.

# 3   DataMotion for Volumes and Nondisruptive Operations Use Cases

Several different customer use cases benefit from using DataMotion for Volumes: cluster expansion, hardware lifecycle management, and the alignment of business priorities with data volumes. This section discusses each of these use cases as well as the nondisruptive solutions available in clustered Data ONTAP. For each use case, the workflow is based on the nondisruptive volume move functionality. Although the actual end goal varies, the process for volume move is generally the same.

1.  Identify the volume(s) to be moved. The volume being moved is chosen based on the desired end goal. For example, a customer might select a volume in an aggregate that is about to run out of space and move it to an aggregate with more space to accommodate future volume growth.

2.  Identify a destination aggregate to move the volume to. Choosing a destination aggregate is influenced by the initial objective for moving the volume. For example, if the objective is moving a volume to accommodate future growth, then the destination aggregate needs to have enough space for projected future growth. If the objective is to alter the performance characteristics of the volume, an aggregate of the required disk type (for example, SSD, Flash Pool™ intelligent caching, SATA, or SAS) should be chosen. The command `vol move target-aggregate` provides a filtered list of suitable target aggregates, based on size and aggregate type. OnCommand System Manager and WFA also provide information on available aggregates.

3.  Initiate the vol move using either the command line interface (CLI) or one of the GUI methods available with OnCommand System Manager or Workflow Automation. A CLI vol move gives the user more flexibility in defining options associated with the vol move compared to the GUI in OnCommand System Manager. The GUI uses a wizard specifically purposed for volume moves. The wizard has predefined default values for the vol move. For more information on CLI and OnCommand System Manager usage, refer to section 0.

4.  Monitor the vol move progress. Detailed status is available with the `volume move show` command. See section  4.3 for more on monitoring the progress of a vol move.

The remainder of this section describes the major use cases for DataMotion for Volumes. Within each scenario, there can be a number of variations to match a particular requirement; therefore, only an overview of each use case is given. This overview is not meant to replace or replicate the resources in the product documentation, which should be consulted for detailed processes.

## 3.1   Cluster Expansion

There is no foreseeable end to the growth in data storage. An on-demand scalable storage solution to meet growth needs is essential to eliminate sizing headaches and meet rapidly changing resource needs. DataMotion for Volumes allows customers to scale a storage system as needed and relocate data volumes to new hardware to balance provisioned resource pools on demand.

This user scenario can be as simple as a customer adding additional drives to an existing system or as expansive as adding completely new systems to a cluster. Although the scale of each of these scenarios differs, the underlying usage of DataMotion for Volumes is the same.

NetApp storage solutions have always scaled on demand within a storage system. A storage administrator has the ability to nondisruptively add shelves or a loop/stack to increase the amount of storage on a system. Clustered Data ONTAP takes this scaling capability further with its extensible scale-out cluster infrastructure. In clustered Data ONTAP, one or more HA pairs operate and communicate as a single system, and a storage administrator can grow the cluster when required by adding nondisruptively a new HA pair with storage shelves to the cluster. With DataMotion for Volumes, data can be moved from the existing cluster nodes to the newly deployed storage or nodes without disrupting the client/host applications. The following diagrams show three scenarios: adding new shelves to an existing system within a cluster, adding a new loop/stack to an existing system within the cluster, and adding a new system (controllers and storage) to a cluster.

For cluster expansion, NetApp recommends managing the volume moves during times when foreground I/O is lower and jobs that use background resources and the cluster interconnect are minimal. In addition, apply the best practices in section 6 and following sections.

Figure 2 shows three scenarios for nondisruptive storage expansion with vol move. These are:

- Scenario 1: Add shelves to existing controllers and disk stack. If the new storage is to be added to an existing aggregate, vol move is not applicable.
    a. Add shelves to an existing loop.
    b. Define aggregate(s) on the new storage.
    c. Move volumes from existing aggregates to the newly added aggregate(s) as desired to rebalance space and performance.
- Scenario 2: Add a new loop/stack of shelves to existing controllers and disk stack. If the new storage is to be added to an existing aggregate, vol move is not applicable.
    a. Install disk adapter if required.
    b. Install new storage on a new loop/stack.
    c. Define aggregate(s) on the new storage.
    d. Move volumes from existing aggregates to the newly added aggregate(s) as desired to rebalance space and performance.
- Scenario 3: Add a new HA pair with storage to the cluster, increasing the cluster size and capacity.
    a. Join new nodes to the cluster as an HA pair, taking care to remain in conformance with the platform mixing rules.
    b. Define aggregate(s) on the new storage.
    c. Move volumes from existing aggregates to the newly added aggregate(s) as desired, to rebalance the space and performance.

**Figure 2) Nondisruptive storage expansion**
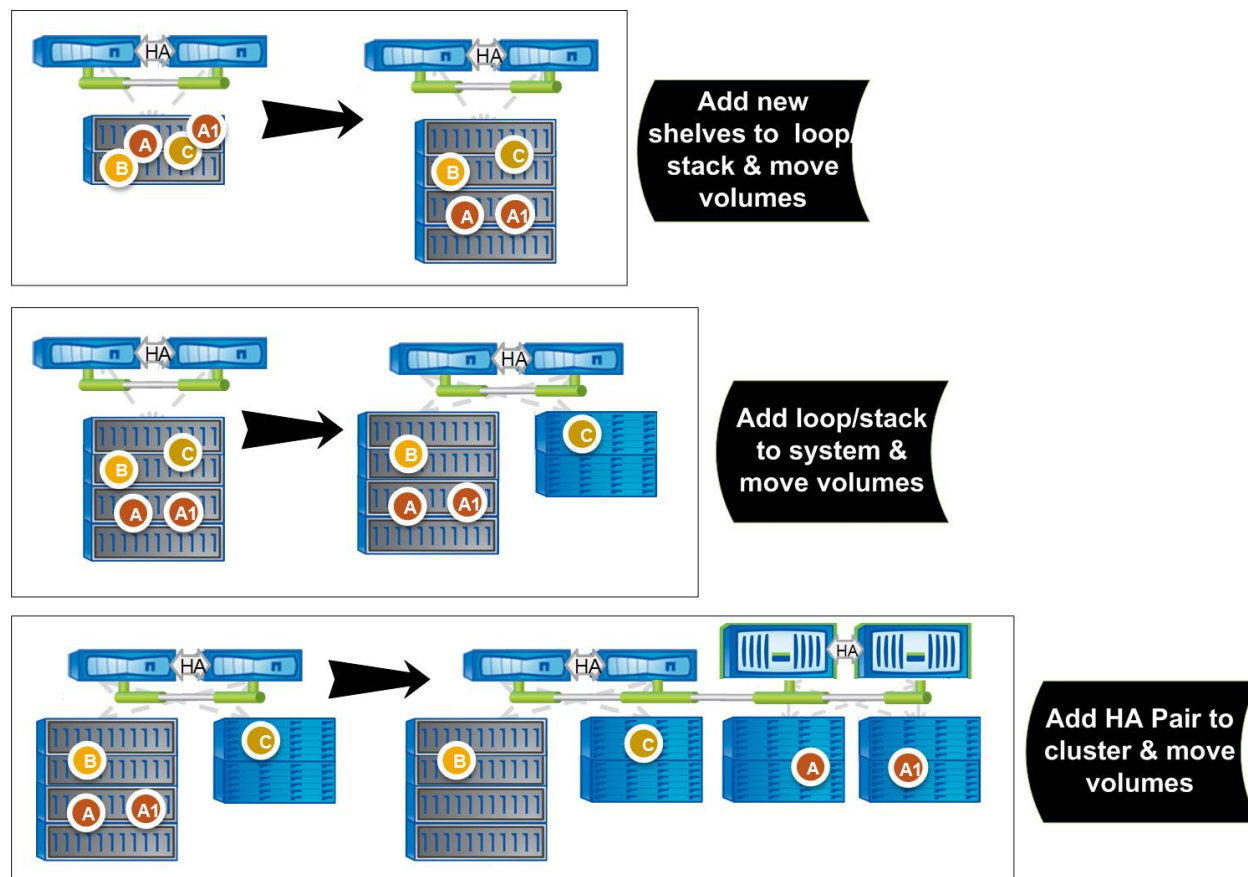


## 3.2 Hardware Lifecycle Management

The lifetime of hardware is generally surpassed by the lifetime of the data being stored on the hardware. What this means is that the data residing on hardware today will, at some point, need to be relocated when the hardware reaches a point of retirement and needs to be decommissioned. DataMotion for Volumes enables datasets to be migrated from hardware that has reached the extent of its lifespan to newly deployed hardware within the cluster.

Many components of a storage system will go through hardware lifecycle management. The most common are shelves and controllers.

### Upgrading Shelves

Upgrading shelves using DataMotion for Volumes requires that the cluster nodes have the capacity to add shelves before removing the old shelves, since the shelves being upgraded must be completely evacuated before removal. The following diagram shows a two-node cluster with two DS14Mk4 shelves. Two DS4243 shelves are added as a new stack to the system and the volumes that resided on the DS14Mk4 shelf are nondisruptively moved to the DS4243 shelves using DataMotion for Volumes. The volumes are seamlessly moved to shelves with new capacity drives on SAS technology without disrupting data availability to the client/host applications.

In Data ONTAP 8.2 and higher, supported storage shelves can be nondisruptively removed from a controller, provided that all data on the shelves are completely evacuated, using DataMotion for Volumes.

The shelves being removed must contain only unowned disks prior to removal. See the High Availability Guide for more information on how to nondisruptively remove storage shelves.

Figure 3 shows the scenario for upgrading disk shelves.

a. Install disk adapter if required.

b. Install new storage on a new loop/stack.

c. Define aggregate(s) on the new storage.

d. Move volumes from existing aggregates to the newly added aggregate(s) to evacuate the storage being refreshed.

e. After all volumes have been evacuated from the older storage, delete the aggregates and nondisruptively remove the shelves (requires at least clustered Data ONTAP 8.2 or higher version)

**Figure 3) Nondisruptive technology refresh for shelves.**
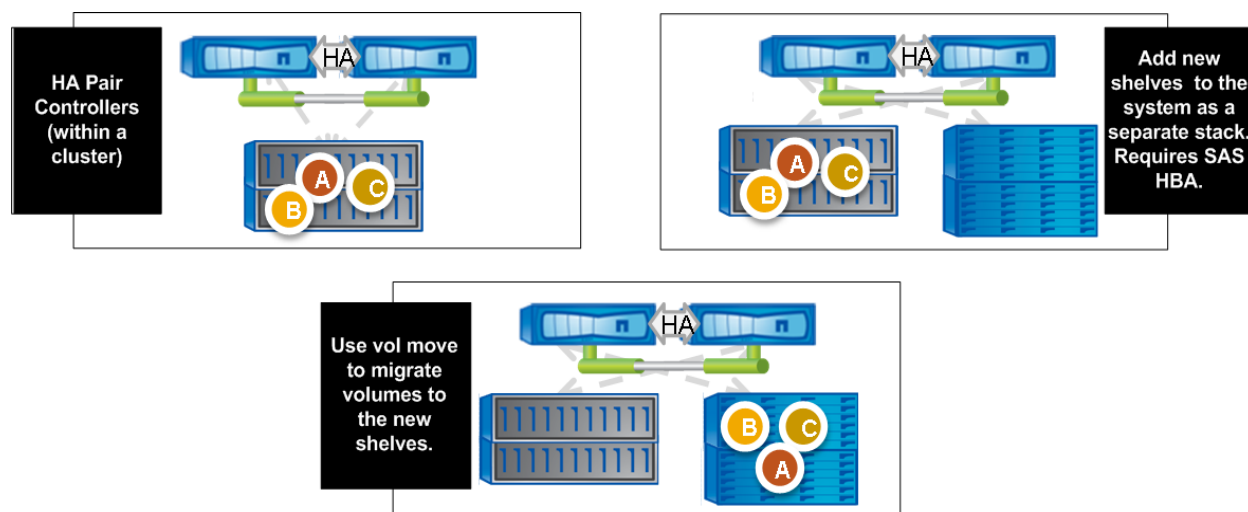


## Upgrading Storage Controllers

There are different possibilities to upgrade controllers. In all instances, the composition and size of the cluster must remain in conformance with the cluster platform mixing rules at all times.

- In-place controller upgrade with ARL: In systems with clustered Data ONTAP 8.2 or higher, aggregate relocation (ARL) is often the preferred solution compared to DataMotion for Volumes to nondisruptively upgrade controllers, because it does not require physical data movement. This option is available for controllers that do not use internal drives for either root or data aggregates and applies where just the controller will be upgraded, not the storage. For more information on aggregate relocation for controller upgrades, see TR-4146: Aggregate Relocate Overview and Best Practices for Clustered Data ONTAP.

- In-place controller upgrade without ARL: In clusters of four or more nodes, it is possible to upgrade the controllers in place if there is sufficient available capacity in the cluster to evacuate an HA pair. In summary, the process is:

a. Identify an HA pair for controller upgrade.

b. Evacuate all data using DataMotion for Volumes and migrate LIFs to other nodes in the cluster.

c. The HA pair then does not serve any data. Shut down the HA pair to perform the upgrade on both nodes.

      d.   Move the volumes back and send the LIFs back to their home ports.

      e.   Repeat the process to upgrade other HA pairs as required.

- Upgrade of controllers and shelves together: In this use case, the new controllers are joined as new cluster nodes. This scenario requires significantly less planning and fewer and simpler steps compared to an in-place controller upgrade with ARL; however, it requires that the storage be refreshed with the controllers. OnCommand Workflow Automation 2.1 includes a workflow for this use case that can further simplify the process.

      a.   Identify an HA pair for controller upgrade.

      b.   Add new nodes and storage to the cluster.

      c.   Evacuate all data using DataMotion for Volumes, delete all aggregates, and delete or rehome all LIFs from the HA pair being upgraded.

      d.   Unjoin the evacuated HA pair from the cluster.

For technology refreshes using volume move, NetApp recommends managing the volume moves during times when foreground I/O is lower and jobs that use background resources and the cluster interconnect are minimal. In addition, apply the best practices in section 6 and following.

## 3.3 Matching Business Priorities for Data Volumes

Data storage volumes are a dynamic resource for workloads. The required priority of any given volume is a function of the workload's service-level objective (SLO) requirements and the capacity growth expectations. The ability to move volumes around to meet business priorities is essential to maintain performance expectations of the dataset. DataMotion for Volumes provides the transparent movement of volumes to alternate resources within the cluster to better accommodate the resource needs of the workload on the volume.

There are a number of scenarios for vol moves of this type, and these may also be combined:

- Between smaller and larger-capacity drives, for example, SAS to SATA
- Between storage tiers with different performance characteristics, for example, standard to Flash Pool
- Between controllers with different bandwidth and performance speeds, for example, FAS3200 to FAS6200

As an example, in the figure below, volume B moves from a midrange controller with 7200 RPM SATA drives to a high-end controller with 15000 RPM SAS drives. Volume Bt represents the new location of the volume after the move.

**Figure 4) Nondisruptive volume move to meet required SLOs.**
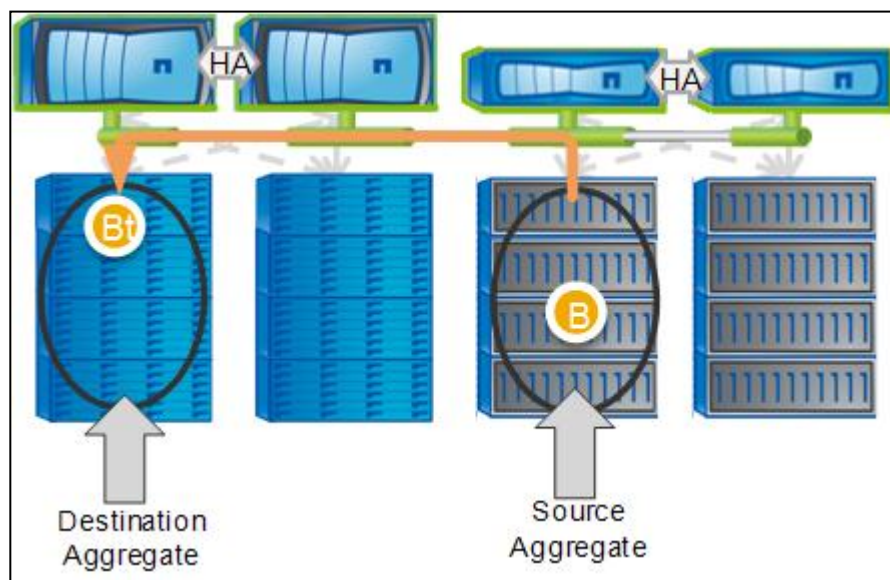


For capacity management, alerts should be configured for aggregates. OnCommand Unified Manager (OCUM) provides monitoring and alerting for storage capacity. For more information on configuring OCUM, see OnCommand Unified Manager Online Help. If aggregates approach their capacity, volumes should be proactively moved to lower utilized aggregates. It is important to plan accordingly by scaling out the cluster with additional new hardware prior to having to move a volume. Space available on existing aggregates might be limited.

For performance management, the following general guidelines apply:

- Maintain an average CPU utilization of controllers at 50%.
- Analyze usage trends over longer periods to highlight repetitive spikes. Anticipating periods of high traffic will provide triggers for a proactive volume move. For example, if a cluster is servicing a finance organization and the last week of each quarter sees a spike in traffic on files for quarter-end documents, then that volume can be preemptively moved each quarter to service the increased level of traffic on those volumes. After the spike in activity, the volumes could be moved again to hardware resources that better fit the workload requirements for the rest of the month. OnCommand Performance Manager can help identify performance trends and spikes.
- Move lightly utilized volumes in preference to more heavily utilized volumes if possible, especially when the system is under heavy load
- If a heavily utilized volume will be moved to a node which has more headroom, schedule the move over a period of lower overall system activity if possible.

# 4   Syntax and Usage

This section discusses execution of DataMotion for Volumes via OnCommand System Manager and the CLI. The CLI provides an increased ability to manage the different options of the volume move job while System Manager provides a simple volume move job wizard that uses default options. The following sections provide an overview of the main capabilities of each interface. OnCommand Workflow Automation also includes a DataMotion for Volumes workflow.

## 4.1   OnCommand System Manager Usage

Vol move can be initiated in OnCommand System Manager from either the Volumes or Aggregates display.

System Manager uses the default vol move options for cutover, so these cannot be customized when using System Manager, unlike in the CLI. Where the default options are sufficient, the System Manager interface provides a simple wizard method to initiate a volume move.

### Volume Move from the Aggregates Panel

The following figure shows how to initiate a volume move from the Aggregates display. Click Volume Move when an aggregate is highlighted to start the volume move wizard.

**Figure 5) System Manager initiate vol move from aggregates display.**
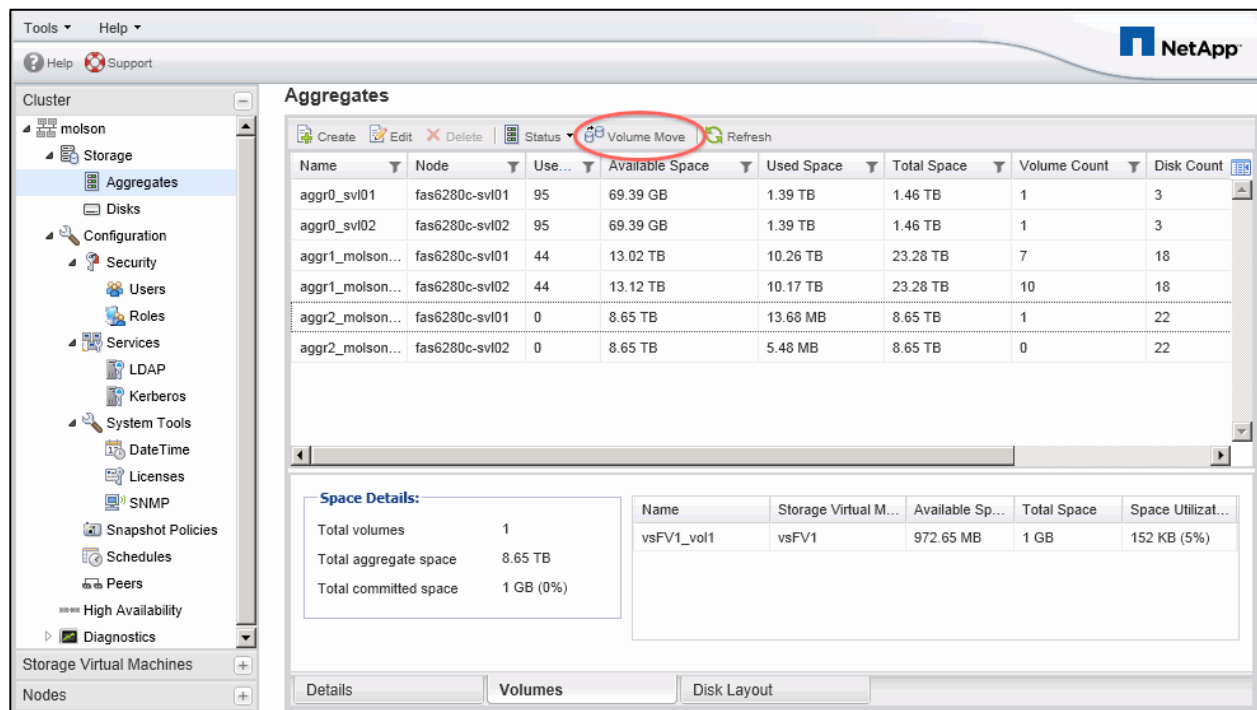


The wizard prompts with a list of volumes in the aggregate. After selecting which volume should be moved, the wizard prompts to select the destination aggregate. Only aggregates with sufficient space are displayed. Node root aggregates are not displayed since vol move is not possible to an aggregate of this type. The before and after effect on space utilization for the destination aggregate is shown.

**Figure 6) System Manager select vol move destination aggregate.**



A summary of the operation and its impact on utilized space is presented for confirmation before proceeding.

**Figure 7) System Manager vol move summary.**

## Volume Move from the Volumes Panel

Vol move can also be initiated by selecting the volume to be moved directly from the Volumes display, as shown in the figure.

**Figure 8) System Manager initiate vol move from volumes display.**
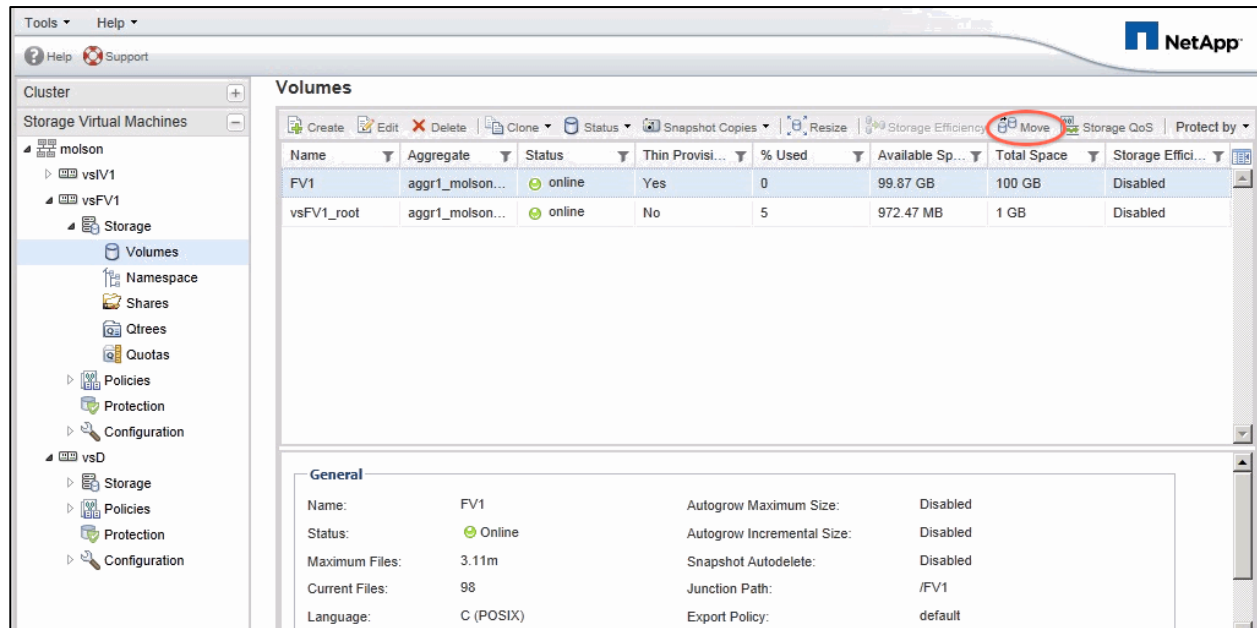


System Manager will display the eligible destination aggregates (data aggregates only), sorted by available space for selection. The effect on space utilization in the source and destination aggregate is also shown. After you click Move on this screen, the move proceeds after one final confirmation.

**Figure 9) System Manager select vol move destination aggregate.**



## 4.2 CLI Usage

Use the `vol move start` command at the cluster shell. The available options for the command may be left at their default values or modified within the acceptable ranges of values. The table below lists the options for the `vol move` command, which are all available at admin privilege. There are no advanced privilege options for the command. The command requires clusterwide credentials; it is not available to SVM administrators since the command can have clusterwide effect and is not confined within the resources of the SVM.

**Table 3) DataMotion for Volumes CLI options.**

| Value | Options | Usage |
|---|---|---|
| -vserver | Vserver_name | The SVM that owns the volume. Only volumes in data SVMs can be moved. Moving node root volumes is not possible. |
| -volume | Volume_name | The volume to be moved. |
| -destination-aggregate | Aggregate_name | The aggregate to move the volume to. A node root aggregate cannot be entered as the destination aggregate, only a data aggregate. |

| Value | Options | Usage |
|---|---|---|
| -cutover-window | Numeric_value | Number of seconds in which the cutover must complete. Cutover can be between 30 and 300 seconds and the default is 45 seconds. The cutover window specifies the *maximum* amount of time allowed; if the cutover can complete in less time than the window, then it will, allowing user access to resume earlier. |
| -cutover-attempts | Numeric_value | Number of times to retry a cutover. If this value is exceeded without a successful cutover, the action taken depends on the value of cutover action. Cutover attempts can be from 1–25. The default is 3. |
| -cutover-action | retry_on_failure\|abort_on_failure\|defer_on_failure\|force\|wait\| | The action for cutover.The default values are: *retry_on_failure* for ONTAP 8.3 and *defer_on_failure* for ONTAP 8.2. Details on possible cutover actions are in section 4.4. |
| -perform-validation-only | true\|false | Perform the validation checks only for the specified move without actually starting the move. The default is false. |
| -foreground | true\|false | Runs volume move as a foreground process. The default is false (operation runs in the background). |

When a vol move is initiated, the output includes a sample `vol move show` command that can be used to monitor progress, as described in section 4.3.

```
Sylvester::> vol move  start -vse vs1 -volume vol1 -destination-aggregate aggr1_Sylvester_02

[Job 7455] Job is queued: Move "vol1" in Vserver "vs1" to aggregate "aggr1_Sylvester_02". Use the
"volume move show -vserver vs1 -volume vol1" command to view the status of this operation.
```

## Additional Vol Move Commands

The following table lists additional commands associated with vol move. These commands all require cluster admin privilege at the admin privilege level. For more information, refer to the "Clustered Data ONTAP Commands: Manual Page Reference," available as part of the clustered Data ONTAP Product Documentation.

**Table 4) DataMotion for Volumes additional commands.**

| Command | Description and Usage |
|---|---|
| Vol move abort | Cleanly terminate a specified running vol move process. The volume in the target aggregate, if already created, is deleted, and user access continues to the volume in the source aggregate. |
| Vol move show | Show status of a running vol move job. See section 4.3 for more information. |
| Vol move target-aggr show | List compatible destination aggregates for a specified volume using specified criteria. |
| Vol move trigger-cutover | Initiate cutover of a deferred vol move job; that is, one that was started specifying –cutover-action wait. |

## 4.3 Monitoring Progress—`vol move show` Command

Once the volume move has been initiated by any method, the operation proceeds automatically through the phases described in section 2.1, "Phases of Volume Move." The phases are executed in the same manner up until the cutover phase. Executing vol move processes can be monitored using the `vol move show command`. This command is a replacement for monitoring via the job number, because it includes substantially more statistics and progress information.

This command shows detailed information on vol move jobs. The syntax and options are described further in the "Clustered Data ONTAP Commands: Manual Page Reference," available as part of the clustered Data ONTAP Product Documentation.

When used without arguments, as in the first command example, a brief summary of all jobs is displayed, including the state, current phase, percent complete, and estimated time to complete. For more detailed information on a job, specify the particular move of interest by SVM and volume, as shown in the second command example. Some of the fields are permanently displayed throughout the operation, including static parameters such as cutover action and cutover attempts. Others display only once a particular phase has been entered, such as Time User Triggered Cutover. It can be particularly helpful to monitor The Percentage Complete, Replication Throughput, and Estimated Remaining Duration.

```
Sylvester::> vol move show
Vserver    Volume     State    Move Phase Percent-Complete Time-To-Complete
---------  ---------- -------- ---------- ---------------- ----------------
vs1        vol1       healthy  replicating
                                          38%              Fri Jan 17 08:40:59 2014

Sylvester::> vol move show -vserver vs1 -vol vol1

                      Vserver Name: vs1
                       Volume Name: vol1
            Actual Completion Time: -
        Specified Action For Cutover: defer_on_failure
          Specified Cutover Attempts: 3
        Specified Cutover Time Window: 45
         Time User Triggered Cutover: -
Time Move Job Last Entered Cutover: -
             Destination Aggregate: aggr1_Sylvester_01
                   Detailed Status: Transferring data: 5.75GB sent.
        Estimated Time of Completion: Fri Jan 17 08:41:01 2014
                     Managing Node: Sylvester-02
              Percentage Complete: 82%
                       Move Phase: replicating
          Prior Issues Encountered: -                    <<< requires advanced mode
      Estimated Remaining Duration: 00:00:07.000
           Replication Throughput: 157.6MB/s
                   Duration of Move: 00:00:42.000
                 Source Aggregate: aggr1_Sylvester_02
               Start Time of Move: Fri Jan 17 08:40:14 2014
                       Move State: healthy
```

## 4.4 Controlling the Cutover Phase

The cutover phase is the critical phase during vol move to assure continuity of data access. The behavior at the point of cutover is controlled by the `-cutover-action` parameter. There are different options to allow the cutover to be as conservative or aggressive as necessary. The `vol move start` CLI and WFA workflows allow `-cutover-action` to be specified, while System Manager always uses the default setting – *retry_on_failure* in ONTAP 8.3 and *defer_on_failure* in ONTAP 8.2

The possible options for `-cutover-action` are described in the following table.

**Table 5) DataMotion for Volumes -cutover-action options.**

| -cutover-action Value | Vol Move Cutover Behavior |
|---|---|
| retry_on_failure (default value – 8.3) | When a *retry_on_failure* action is specified, the job retries to cutover indefinitely and it never enters a hard-deferred state. After exhausting cutover attempts, the move job waits one hour before trying to cutover again. The user can issue a volume move trigger-cutover command at any time to halt or restart the cutover process.<br><br>This option is set as default starting in clustered Data ONTAP 8.3 and requires advanced privileges to modify. |
| defer_on_failure | A cutover is retried until it is either successful or reaches the specified number of `-cutover-attempts`. If the cutover fails on the last retry, then the cutover is deferred and continues in the iterative phase, keeping source and destination in sync, until the administrator manually triggers a further cutover. A cutover is triggered using the `vol move trigger-cutover` command. When the cutover is triggered it is retried the next time a cutover decision point is reached and continues to retry up to the number of cutover attempts. |
| abort_on_failure | A cutover is retried until it reaches the specified number of `-cutover-attempts`; if the final cutover fails on the last retry, then the move operation is aborted and the temporary destination volume is removed from the destination aggregate. |
| wait | After the baseline transfer, vol move remains in the iterative phase, periodically replicating the source to target volume. Cutover is not attempted until the administrator manually triggers it using the `vol move trigger-cutover` command. Then the cutover is attempted the next time a cutover decision point is reached.<br><br>This option is particularly useful for very active volumes for which cutover within the required interval may be difficult to achieve. The bulk of the replication can proceed concurrently with user access and the cluster administrator can wait for a known less busy time to trigger the cutover. The cutover can even be deferred to a planned maintenance window. |
| Force | A cutover is retried until it reaches the specified number of `-cutover-attempts`; if the final cutover fails on the last retry, then the cutover is forced. A forced cutover can cause disruption and does not abide by the specified -cutover-window.<br><br>Use this option with caution and only if the possible disruption is acceptable to end users and applications. |

# 5  Restrictions and Compatibility Considerations

DataMotion for Volumes is a nondisruptive data mobility feature that is supported for all commonly supported client/host applications; however, there are some compatibility considerations around other cluster operations that impact DataMotion for Volumes. The following section covers the finer details of vol move interactions within the ecosystem of Data ONTAP software and NetApp storage systems.

## 5.1 Cutover-Phase Guarantee

The cutover guarantee enables the final replication of and access to the new location of the volume to complete in an amount of time that does not interrupt access of the client/host to the data stored on the volume. The cutover guarantee is the key to DataMotion for Volumes being nondisruptive.

During the cutover phase of the volume move, client/host requests are queued until the cutover completes. The following are best practices for the cutover phase.

- The default cutover duration is 45 seconds for volume move. A cutover can be 30 to 300 seconds. The cutover value for a volume move should not exceed the expected timeout value requirements of the client or host applications for that dataset. Consult the documentation or best practices guides for your host/client OS or applications for recommended timeout settings.

- Execute a volume move during a period of lower system utilization. See sections 6 and 7 for more details.

- Prior to Data ONTAP 8.3, restrict the number of concurrent vol move jobs initiated to no more than eight references to a source or target aggregate on a particular node. In Data ONTAP 8.3, vol move performs this function automatically to limit the number of concurrent jobs executed. See section 7 for more information.

- Use volume move options as discussed in section 4.4 to control the cutover behavior.

    - Do not use the force value for `cutover-action` unless absolutely necessary. A forced cutover may be disruptive, and appropriate precautions should be taken depending on the customer environment.

## 5.2 Interactions with Other Cluster Operations

Several operations are discussed in this section that, like volume move, update the volume location or other volume information in the cluster's internal shared metadata. In order to assure the integrity of the cluster metadata, these operations may be mutually exclusive, or one type of operation may preempt the other during certain vol move phases. An operation that is mutually exclusive with DataMotion for Volumes is defined as a job or an operation that cannot take place at the same time as a volume move. Mutual exclusivity is verified during the validation portion of the setup phase. Improvements have been made to the interoperability of vol move with other cluster operations in various releases of Data ONTAP, and these are highlighted where appropriate.

### SnapMirror

There are considerations for interactions between volume move for volumes in a SnapMirror relationship. In all versions of Data ONTAP, a source volume of a SnapMirror relationship can be moved while the volume is being mirrored. SnapMirror services encounter a brief pause during the cutover phase of the volume move job.

Prior to Data ONTAP 8.3, a destination volume of a SnapMirror relationship is mutually exclusive with a volume move, meaning that a volume move and a SnapMirror operation, like a SnapMirror update, cannot be executed concurrently. For example, if a destination volume of a SnapMirror relationship is locked by a SnapMirror update, then a vol move cannot be executed. Similarly, if a destination volume of a SnapMirror relationship is locked by the volume move operations, then a SnapMirror update is deferred until the lock placed by the volume move is released. If a SnapMirror operation is a priority, a volume move can be aborted to allow the SnapMirror operation to complete.

Data ONTAP 8.3 eases this restriction to allow SnapMirror/SnapVault® technology and the vol move of a destination volume to execute in parallel. In the iterative phase, SnapMirror/SnapVault updates and vol move run concurrently. In evaluating whether a cutover is possible in the cutover phase, priority between

the cutover and SnapMirror or SnapVault updates is administered on a first-come/first-served basis. Either operation will block until the operation occurring first completes.

SnapMirror resync and break and SnapMirror restore using snaprestore are mutually exclusive with volume move with all releases of Data ONTAP.

## NDMP Tape Operations

Prior to Data ONTAP 8.3, DataMotion for Volumes is a mutually exclusive operation with NDMP tape operations. A volume move job must complete before a tape operation can begin. A tape backup that is initiated or scheduled while a volume move is in progress will fail. If you run a version of Data ONTAP earlier than 8.3, NetApp recommends that you coordinate vol moves with the tape backup schedule so as not to induce conflicts.

In Data ONTAP 8.3, a vol move operation and an NDMP tape backup/restore operation can execute in parallel. Vol move operations are blocked in favor of tape operations until the vol move baseline copy is completed. Once the baseline vol move copy is complete, similar to SnapMirror, ndmp and vol move operations are prioritized on a first-come/first-served basis.

### NDMP Backup Applications and Vol Move

Prior to Data ONTAP 8.3, NDMP backup is node-scoped. If a volume is moved, the backup configuration must be updated to specify the new node hosting the volume. In most cases, a full baseline of the volume is required on the next backup operation after a vol move.

Once a cluster has been fully upgraded to clustered Data ONTAP 8.3 or higher, storage virtual machine (SVM)–aware NDMP backups are available. This helps in optimizing NDMP backup performance by choosing efficient data transfer paths, and it is fully compatible with the integrated nondisruptive operations and volume mobility capabilities of clustered Data ONTAP. In particular, SVM-aware NDMP backup does not require a full volume baseline backup after a vol move. To support this feature, the CAB (Cluster Aware Backup) extension has to be implemented by the backup vendor application. Refer to the Interoperability Matrix (IMT) for the vendor backup application support list. If the vendor application does not support the extension, the node-scoped backup behavior for NDMP prior to Data ONTAP 8.3 will be in effect.

The following table summarizes the behavior of NDMP backup in conjunction with vol move, depending on whether node-scoped or SVM NDMP mode is enabled. More information on NDMP backup with clustered Data ONTAP is in the "Data Protection Tape Backup and Recovery Guide" in the Product Documentation.

**Figure 10) NDMP backup and vol move.**

| NDMP Backup Type | Vol Move and Backup Behavior |
|---|---|
| SVM-aware NDMP mode when CAB extension is supported by the backup application | Vol move does not require a new baseline backup. Incremental tape backup and restore operations will continue. |
| SVM-aware NDMP mode when CAB extension is NOT supported by the backup application<br>Or<br>Node-scoped NDMP mode | Vol move requires a new baseline backup unless the LIF configured in the backup policy is also migrated to the node that hosts the destination aggregate. |

## Storage Efficiency

A volume that is moved with deduplication and compression enabled retains the storage efficiency attributes and savings after the move. During the move there is a window of time in which storage efficiency savings are deferred until after the volume move completes; inline compression continues without interruption. Once a volume move is initiated, Snapshot transfers begin copying data to the new volume location on the destination volume. After the baseline transfer is completed for the vol move, a transfer containing the deduplication metadata is transferred to the new volume and postprocess compression and deduplication are stopped on the volume; however, the data blocks continue to be fingerprinted for deduplication or compression. Once the volume move completes, deduplication and compression resume on the data blocks that were previously deferred. If deduplication or compression is in progress when the transfer of fingerprint database is initiated by the vol move job, postprocess compression and deduplication are aborted and the volume move job continues. Once the volume move job completes, deduplication and compression processes are automatically restarted at the next scheduled time or can be manually restarted on the volume, now located on the destination aggregate. It continues from the last checkpoint.

**Note:** Exercise caution if you leave a move of a storage efficiency-enabled volume in the cutover deferred state for an extended period of time. The change log may fill, which can result in loss of storage efficiency.

## Volume Modification Operations

Prior to Data ONTAP 8.3, many commands that modify a volume were mutually exclusive with a move of that volume. These include volume delete, offline, and modify (most attributes). In Data ONTAP 8.3, most volume modify operations can execute in parallel with a volume move up to the cutover phase. Volume offline, restrict, destroy, and volume snapshot restore remain mutually exclusive with a volume move.

## Flash Pool

Flash Pool is a component of the NetApp Virtual Storage Tier. Flash Pool combines solid-state disks (SSDs) and traditional hard disk drives (HDDs) in a single Data ONTAP aggregate. The NetApp storage system takes advantage of the latency and throughput benefits of SSD while maintaining the mass storage capacity of HDD. DataMotion for Volumes is fully supported for Flash Pool—a volume can be moved freely to and from Flash Pool aggregates as well as between a Flash Pool aggregate and a homogenous HDD aggregate in either direction. However, the administrator must pay attention to any potential performance impact if a volume is moved from a Flash Pool aggregate to regular HDD. The `vol move` command itself will not alert that a volume is being moved out of a Flash Pool aggregate.

If a volume is moved between two Flash Pool aggregates, the heat map is not transferred with the volume; hence reads may take longer while the cache is repopulated.

More information on Flash Pool is on the [Field Portal](Field Portal).

## Flash Cache

Cached blocks in Flash Cache are not moved with the volume. Hence, the Flash Cache cache on the target node needs to be rewarmed with frequently read data, and reads may take longer while this is in process.

## FlexClone

A volume with child FlexClone® volumes can be moved without restriction, and the child clones remain on the original aggregate. However, the original parent volume, instead of being deleted at the end of the volume move, remains on the source aggregate in restricted mode to provide the shared blocks for the remaining FlexClone volumes. This restricted parent volume is not, and cannot be, junctioned into the SVM and is not otherwise available for client or host access. When all the child clones are deleted or split,

the restricted parent volume can be deleted. Any such volumes are automatically deleted by the internal *volume reaper* job, which runs periodically.

If a child clone volume is moved, it is automatically split from the parent and inflated so that it consumes the full amount of required space and no longer accesses any shared blocks.

In the case of chained clones, consider the example of volume A cloned to volume B, which is then cloned to volume C. If volume A is moved, the behavior is as described above for the parent volume. If volume B is moved, it is inflated on the target aggregate. It cannot be split from A in this case, since it has a child clone C. If volume C is moved, it is split from its parent and is inflated on the target. The original volume C on the source aggregate is deleted.

## HA Pair Controller Storage Failover

Prior to Data ONTAP 8.3, an unplanned storage failover event can result in termination of a running volume move operation. If a node fails over before or during the iterative phase of the vol move, the job aborts and cleans up, and access to the source volume resumes on the partner node. Although this process does not disrupt client/host applications, the vol move job needs to be manually reinitiated and will start from the beginning of the validation phase. If a node fails during the cutover phase, there are two possible outcomes. There is a checkpoint ("point of no return") in the cutover phase where data for the updated volume location is committed. A volume move cannot be reversed once this checkpoint has been reached. Therefore, when the vol move job fails over on a node failover, if the vol move has not yet reached the point-of-no-return checkpoint, it will roll back and be cleaned up, just as if it failed over in the iterative phase. If the checkpoint has passed, it will roll forward and continue to completion on the node that took over.

In Data ONTAP 8.3, multiple checkpoints are automatically taken during the vol move iterative phase. Hence, a vol move automatically restarts after an unplanned storage failover on the source or destination node—resuming the replication close to the point where it was interrupted by the failover. This provides much better restartability and resilience for vol move jobs.

Prior to Data ONTAP 8.3, planned failover or giveback operations cannot occur while a volume move is in progress. In Data ONTAP 8.3, planned failover or giveback is only mutually exclusive when a vol move is in the cutover phase. If the vol move is in the iterative phase, the vol move data copy is interrupted to allow the failover or giveback to execute. It then resumes after the failover or giveback has completed.

## Upgrade and Revert

Clusters in the process of being upgraded to a newer Data ONTAP release have constraints for moving volumes between nodes running varying versions of Data ONTAP. At Data ONTAP 8.1 and higher, DataMotion for Volumes uses physical replication SnapMirror technology to replicate data. Similar rules to those for SnapMirror, therefore, apply to volume movement between aggregates on nodes running different clustered Data ONTAP versions. In particular, the destination aggregate of a volume move must be owned by a node running the same or higher major Data ONTAP version than the node that owns the source volume. A volume cannot be moved from a higher major version node to a lower major version node. This scenario would occur when a cluster runs in mixed-version mode; for example, during an upgrade of Data ONTAP. NetApp recommends running mixed-version clusters only for as long as absolutely necessary; for example, for the amount of time it takes to complete a rolling or batch Data ONTAP upgrade of the entire cluster. The guidance is to complete the upgrade within a week.

For example, during a major upgrade to Data ONTAP 8.3, some nodes in the cluster will still run Data ONTAP 8.x while others will have completed the upgrade to Data ONTAP 8.3. A volume on a Data ONTAP 8.x node can be moved to any other node, since Data ONTAP 8.x is the lower major version of Data ONTAP. A volume on a Data ONTAP 8.3 node can be moved only to another Data ONTAP 8.3. A volume move job will fail if the source aggregate resides on a node with a higher version of Data ONTAP than that of the destination aggregate's node. NetApp recommends that customers avoid running volume

move jobs during major version upgrades or when a cluster is otherwise in a mixed-version state. See the Nondisruptive Operations FAQ for more information on running in a mixed-version state.

The restriction does not apply to minor Data ONTAP upgrades; Vol move works between nodes running different minor releases.

**Note:** In clustered Data ONTAP 8.1 and later, DataMotion for Volumes uses physical replication to move volumes. Previously, in 8.0, logical replication was used. Physical replication has the ability to transfer sets of Snapshot copies simultaneously as opposed to logical-based replication in which a single Snapshot copy is transferred at a time. The change from logical to physical replication for vol move has implications for clusters upgrading from 8.0 to 8.1. In clusters running in mixed-version, in which some nodes are at 8.0 and some are at 8.1, volumes cannot be moved between nodes running different Data ONTAP versions in any direction. For example, volumes cannot be moved from a source aggregate owned by a node running Data ONTAP 8.0 to a destination aggregate owned by a node running Data ONTAP 8.1 or vice versa.

## Aggregate Types

Node root volumes cannot be moved. Volumes cannot be moved to a node root aggregate. This is prevented in both System Manager and the CLI, as noted in section 0. The following table summarizes the aggregate compatibility rules for vol move. Moving data from a 32-bit aggregate to a 64-bit aggregate automatically converts the data in the volume to the 64-bit format; however, Snapshot copy data, being read-only, remains in the 32-bit format until it is explicitly converted or expires.

Note the following restrictions and recommendations in vol move between aggregates of different types.

1. In Data ONTAP 8.3., a volume cannot be moved from a 64-bit aggregate to a 32-bit aggregate. In Data ONTAP 8.3.x, 64-bit aggregates are the default aggregate type, and NetApp highly recommends creating and using only 64-bit aggregates in all releases of Data ONTAP. In prior releases, although it is supported, it is preferable to move volumes of any type only to 64-bit aggregates.

2. 32-bit data is not supported in a cluster running Data ONTAP 8.3 or higher. All 32-bit volumes must be upgraded to 64-bit, and all 32-bit Snapshot copies must be removed before upgrading any cluster node to Data ONTAP 8.3. Since a 32-bit volume cannot therefore exist in Data ONTAP 8.3, vol move of such data cannot ever occur. See TR-3978: In-Place Expansion of 32-Bit Aggregates to 64-Bit Overview and Best Practices for more information on the mandatory detection and conversion of 32-bit data before upgrading to Data ONTAP 8.3.

   **Note:** Discussion of ONTAP capabilities is provided in order to facilitate planning for 64-bit aggregates prior to the Data ONTAP 8.3 release. This is not a guarantee of future functionality and is subject to change without notice.

**Table 6) Aggregate compatibility for DataMotion for Volumes.**

| Source Aggregate | 32-Bit Target Aggregate | | 64-Bit Target Aggregate | |
|---|---|---|---|---|
| | Before Data ONTAP 8.2 and Data ONTAP 8.3 | Data ONTAP 8.2 and Data ONTAP 8.3.x | Data ONTAP 8.2 and Data ONTAP 8.3.x | Data ONTAP 8.3.x |
| 32-bit | Yes, but not recommended. Best practice is to move to 64-bit. | Yes, but not recommended. Best practice is to move to 64-bit. | Yes | **No** All 32-bit volumes must be upgraded to 64-bit, and all 32-bit Snapshot copies must be detected and deleted before upgrading any node to Data ONTAP 8.3. |

| Source | 32-Bit Target Aggregate | | 64-Bit Target Aggregate | |
|---|---|---|---|---|
| 64-bit | Yes, but not recommended. Best practice is to move to 64-bit. | **No** | Yes | Yes |

## 5.3    Protocol-Specific Behavior

### SAN Environments

DataMotion for Volumes supports the movement of volumes containing LUNs. The movement is at the volume level and all LUNs in the volume are automatically moved along with the containing volume. The use of MPIO and ALUA within Data ONTAP and on the hosts enables the paths to the LUNs to be automatically updated to preserve the most optimal path for application access, provided that best practices for configuring SAN LIFs are followed.

The main consideration when moving a SAN volume is the host-side timeout value. The cutover window defined for a volume move with a SAN volume should not exceed the expected timeout value on the host side. NetApp recommends that the cutover window not exceed 120 seconds for volumes in SAN solutions. When a SAN volume is moved, ALUA is used for optimized access to the volume. Each node must have a data LIF for optimized access to the volume.

### SMB Environments

DataMotion for Volumes is nondisruptive for SMB volumes so long as the source node of the aggregate on which the volume is located does not encounter a failure. During the volume move, the cutover phase must be defined to be within an acceptable range for SMB clients. For all supported SMB versions, NetApp recommends a cutover window of 30 seconds to maintain sessions and eliminate disruption. When the cutover attempt is made, the SMB I/O will be queued at the controller level until the cutover is complete, up to a maximum of 30 seconds. The CIFS application waiting for acknowledgment of the I/O should have a timeout window equal to or greater than 30 seconds. Once the cutover completes and the I/Os are routed to the new volume location, the application will see acknowledgments of the pending I/Os without disrupting the application.

If a LIF migrate will occur in conjunction with the vol move, it is important to note that LIF migrate is disruptive for SMB 1.0 environments. With SMB 2.0 and later releases, LIF migrate is nondisruptive.

### NFS Environments

Clustered Data ONTAP 8.1 introduced NFSv4.x referrals. A referral directs a client to another LIF in the SVM. NFSv4 clients use referrals to access data over the referred LIF path, regardless of the LIF specified in the mount statement. Referrals are issued when there is a LIF in the SVM that resides on the cluster node where the data volume resides. In other words, if a cluster node receives an NFSv4 request for a nonlocal volume, it refers the client to the data LIF corresponding to the direct path for that volume. This allows clients faster access to the data via a direct path and avoids indirect traffic on the cluster interconnect.

If the volume is moved to another node, there is no new referral, and the client continues to access the volume by the previously referred to LIF. The client traffic is therefore indirect at this point, even as the vol move proceeds nondisruptively to the client, as for all protocols. If it is necessary to continue direct access to the volume, the NFSv4 client needs to unmount and remount the file system to receive a referral to the LIF providing the direct path to the volume.

When pNFS is used in clustered Data ONTAP, the client automatically is given the direct LIF path by the metadata server in clustered Data ONTAP. If the volume moves to another node, the new direct path is returned to the client, so direct access is assured provided best practices for configuring pNFS have been followed with a data LIF configured on each node per pNFS-enabled SVM.

For more information on NFSv4 referrals and pNFS, see TR-4067: Clustered Data ONTAP NFS Implementation Guide.

# 6  Best Practices for Iterative Data Transfer Phase

The iterative phase is broken down into two parts: the initial or baseline transfer and the incremental transfer. The initial transfer replicates the full contents of the volume from the source to the newly created destination volume, current at the time when the transfer is started. During this portion of the transfer, expected throughput is streamlined since data can be replicated sequentially in large blocks. The second part of the iterative phase, the incremental transfer, starts when the initial transfer completes. The incremental transfer replicates only the changes in a dataset since the previous transfer, rather than the entire chunk. The incremental transfers must cycle through the dataset to flag changes and then replicate them. Smaller blocks of data are replicated and throughput is lower for more rapidly changing datasets.

Perhaps of greater significance during transfer operations is the throughput throttling of vol move. In Data ONTAP 8.3, automatic throttling occurs during the initial baseline phase to assure that the workload imposed by the vol move does not adversely affect end-user latency and throughput. The throttling automatically adjusts itself as client or host traffic increases and decreases, effectively deprioritizing vol move in favor of end-user applications. Throttling may also take effect while certain scanners, such as those for storage efficiency, are running.

Vol move is resilient when transient errors occur; these can happen during the iterative phase, such as when rebooting the node or during the restart of many internal processes. If a data transfer is interrupted or cannot complete, vol move enters a soft-defer state and automatically retries the operation periodically. Vol move can remain in this automatic retry state indefinitely or until interrupted by user.

Based on these capabilities, vol move is essentially a "low-touch" administrative activity. Its primary objective is always nondisruptiveness and elimination of client impact, which are therefore prioritized over end-to-end performance.

## 6.1  Factors and Best Practices for Optimal Performance for Iterative Phase

The iterative phase is the longest part of the volume move job. Since vol move uses the same underlying engine as SnapMirror, theoretical maximum performance on a minimally loaded cluster is similar to that observed during an intracluster SnapMirror replication. However, clusters are rarely "lightly loaded," and, because of the automatic throttling, vol move throughput and associated total execution time may vary widely and are difficult to predict. A general rule of thumb is to plan for maximum throughput on the order of hundreds of MB/sec, rather than expecting speeds approaching that of the cluster interconnect, with the caveat that throttling will reduce throughput to below 100MB/s if and as required. It is possible for some vol moves to take days or even weeks if the volume is close to the maximum size. The Estimated Remaining Duration and Estimated Time of Completion fields of the `vol move show` command (see section 4.3) are useful indicators of duration.

The following sections discuss other factors that impact vol move performance beyond application-related traffic, as well as best-practice recommendations to optimize the iterative phase.

### Cluster Network Bandwidth

A volume move job within a cluster will transfer Snapshot copies of data from the volume on the source aggregate to a volume on the destination aggregate. Except where the source and destination aggregate

reside on the same cluster node, the vol move traffic traverses the cluster interconnect. Although vol move uses the same resources as SnapMirror and indirect data access, in general, the cluster network is architected so that it should not be a bottleneck to such traffic. As more jobs are initiated within the cluster, the available cluster network bandwidth may decrease, slowing the vol move data transfer rate.

The vol move governor in Data ONTAP 8.3 (see "Concurrent Volume Move Jobs," below) tries to maximize throughput by limiting the number of concurrent vol move jobs. This should allow existing move jobs to finish faster, rather than spreading the throughput among more jobs so that they all proceed more slowly and finish in a longer period of time. Vol move throttling (discussed in section 6 Best Practices for Iterative Data Transfer Phase) also slows throughput as needed.

The `dashboard performance show` command is one of the simplest methods to monitor the effect of vol move on the cluster interconnect; for example, on a cluster running a release earlier than 8.2.1 (before the vol move governor is available), or where there is substantial concurrent SnapMirror traffic.

## CPU Usage

A volume move job shares background resources with other jobs that were initiated on the controller. Job requests for background processes are prioritized behind foreground I/O such as application-level traffic. For a controller that is running with a CPU utilization greater than 50%, a vol move job might be bottlenecked by available CPU resources to optimally complete a volume move job and service all other background requests. NetApp recommends that a volume move job be executed during a period when the controller CPU is at less than 50%. Because of the automatic vol move throttling in effect, vol moves initiated on more heavily loaded systems experience extremely low throughput rates, leading to potentially longer execution times. As stated previously, this is a deliberate design choice that is implemented to minimize client impact and prevent service disruption.

## Rate of Change of Data on Source Volume

The amount of data that changes on the volume being moved has an impact on the overall time for the vol move to complete. The incremental transfer employs an algorithm that cycles through the blocks, selects those that have changed, and flags them for replication. As data continues to change, more cycles to check for changes are required, therefore increasing the total time of subsequent replications after the initial transfer and the volume move job as a whole. However, throttling only occurs during the initial transfer; incremental transfers are not subject to throttling.

## Concurrent Volume Move Jobs

Volume move jobs can run concurrently on the same node or different nodes in a cluster. In general, all jobs serviced by the same controller share the available background resources, which are also shared with other traffic such as SnapMirror. Vol move replication is automatically and transparently throttled during the initial transfer so as not to increase client latency. Therefore, kicking off many volume move jobs in parallel on a single node may not result in the vol moves completing any faster than if multiple batches of volume move jobs were started sequentially.

For clusters on releases lower than Data ONTAP 8.3, the general guidance is to initiate no more than 8 vol move jobs on a given node, or 25 vol moves throughout the cluster, for a reasonable completion time. This recommendation is based on typical throughput and performance of a cluster; initiating more moves than this will not achieve higher throughput.

In Data ONTAP 8.3, an automatic and transparent governor process queues and regulates vol move jobs so that an administrator can initiate as many individual jobs as required. Counters are kept of the number of references to the aggregates owned by each node in all the current vol move jobs, whether as a source or target aggregate. If 8 or fewer aggregate references for a node are involved in the initiated vol move jobs, all jobs will execute uninterrupted. If additional references are made to aggregates on that node in subsequent vol move jobs, these will wait in the setup phase after creating the target volume until one of the executing jobs completes.

As slots open up, waiting jobs are randomly selected to automatically proceed to the iterative phase without administrator intervention. In this way, the workload on each node is automatically queued, and the administrator can start as many vol move jobs as are required for the particular use case. For example, suppose 12 vol moves are started concurrently from SAS to SATA aggregates on node A. Only 4 of these jobs (since each vol move job references 2 aggregates on the same node, making 8 references in total) will execute normally; the other 8 will wait in setup. As individual vol moves complete, the counter on the node will drop below 8, allowing the waiting jobs to proceed. Similarly, if 10 vol move jobs are initiated from node A to node B, 8 jobs will proceed to the iterative phase and the remaining 2 will wait in setup until the counters drop below 8 for each node.

Therefore, in Data ONTAP 8.3, there is no need for the administrator to keep track of how many vol moves are executed in parallel—the cluster automatically regulates the workload for optimal execution time and use of required resources.

# 7  Best Practices for Cutover Phase

During the cutover phase, all client I/O is queued at the controller level, possibly yielding a degree of increased latency for a very short period of time. The maximum period of time when increased latency can occur is defined by the cutover window as specified with the volume move job. Once the iterative phase has completed, the volume move will take no more than the defined cutover-window setting to complete before I/O resumes. If subsequent cutover attempts are necessary, each of these individually will similarly take no more to complete than the value specified in the cutover window.

Clearly, appropriate setting of the cutover window is key to the vol move completing without disrupting end users. The default of 45s (or 30s in SMB environments) will be appropriate in many scenarios; however, NetApp recommends customizing the window if required for particular application environments. Consult your application-specific documentation for guidance on the tolerance for timeout.

During the cutover phase, throughput and performance are not generally of concern. When the cutover phase is initiated, the amount of data to be replicated will be relatively small compared to the available throughput. The throttling used in the initial transfer is not in effect; indeed, the vol move job is prioritized at cutover so that it gets all available resources to complete the cutover in the very short amount of time allotted. During the iterative phase, volume moves share background resources with other jobs and are prioritized accordingly. Once the volume move job hits the cutover checkpoint, the volume move job automatically takes priority over other jobs aside from foreground requests from the user applications.

If multiple jobs are ready to enter cutover at the same time, automatic controls manage them so that a given node cuts over only once as either a source or a destination. For example, a cutover from node A to node B could execute concurrently with a cutover from node B to node A; however, subsequent cutovers involving either node, such as node A to node C or node D to node B, automatically defer until the previous cutover is complete.

## 7.1  Factors and Best Practices for Optimal Performance for Cutover Phase

Cutover is triggered when the data to be replicated meets certain criteria that allow the final replication to complete in the defined cutover window. In addition to application recommendations discussed above, CPU usage and volume workload also affect the success of a cutover.

### CPU Usage

When the cutover attempt is initiated, the amount of resources that are available is based on how busy the source and destination node CPUs are at that particular checkpoint when cutover is initiated. If the CPU were to experience a burst of foreground I/O after the cutover is initiated, it might affect the success of the cutover. If the cutover attempt cannot be completed in the cutover window, the cutover is aborted and ultimately retried once the foreground I/O slows and the CPU bandwidth returns to an acceptable rate

for the cutover to be retried. NetApp recommends assigning values for the `–cutover-window` that are consistent with host-/client-side settings.

### Type of Workload on Volume

For a volume with a write-intensive workload, the cutover duration may be insufficient to complete the vol move. If so, the cutover attempt ceases and I/O is resumed on the source volume until a subsequent cutover retry is initiated. Initiation of a cutover signifies that a check has been completed and that the amount of data that has changed is within a range that can be replicated in the defined cutover window. However, with a write-intensive workload, more blocks of data will change, rendering the calculation inaccurate and potentially inhibiting cutover. Minimizing write-intensive workloads during a vol move job (or as the cutover checkpoint approaches) optimizes the cutover for success. For these workloads, specifying a value of wait for the `cutover-action` parameter of the `vol move` command gives manual control and allows the cutover to be deferred until a time of lower activity.

### Cutover Summary

The cutover phase, as controlled by the cutover-window setting, enables the final replication and access to the new location of the volume to complete in an amount of time that does not disrupt client/host access to the volume's data. This cutover guarantee is the key to DataMotion for Volumes being nondisruptive.

During the cutover phase of the volume move, the client/host requests are queued until the cutover completes. For volume move to remain nondisruptive during the cutover while I/O is being queued, it is important to use the following best practices.

- The default cutover duration is 45 seconds for volume move. A cutover can be 30 to 300 seconds. The cutover value for a volume move should not exceed the expected timeout value required by the client or host applications for that dataset.

- Execute a volume move during a period of lower system utilization.

- When using Data ONTAP prior to 8.3, do not exceed more than 25 concurrent volume moves within a cluster.

- Use volume move options as discussed in section 4.4 to specify settings for the number of times to retry the cutover or when to execute a cutover.

- Do not force a cutover unless absolutely necessary. A forced cutover may be disruptive; therefore, appropriate precautions should be taken in the customer environment (for example, notifying users).

## 7.2 Expected Throughput

Although cutover replication includes certain metadata such as lock state information, the total amount of data to be replicated at this phase is generally small and is not restricted by the available bandwidth for data throughput. NetApp recommends following the earlier-stated best practices for optimal performance and nondisruptiveness during the cutover phase of the volume move procedure.

# 8 Best Practices for Postcutover Phase

Moving a volume involves physically replicating data to a new physical location within the cluster. Although the logical block assignments of the data are consistent after a volume move, the references for the corresponding physical location of the data must be updated. This period is referred to as the deswizzle period, which occurs after the cutover phase. The completion of the cutover phase means that all pending I/O is flushed out to the new volume. A scanner for obtaining the physical location of each block kicks in and begins to map it to the virtual block data (which has remained unchanged). For each

read, an extra I/O is incurred to update the information until all blocks are updated. As each block is updated, the performance penalty is reduced. Depending on other workloads in the cluster, read access performance to the moved volume may be impacted until the scanner finishes.

As mentioned in section 5, cached data in a Flash Pool or Flash Cache are not transferred during a move; hence, read access is slower while caches rewarm.

### 8.1   Factors and Best Practices for Optimal Performance for Deswizzling

During the deswizzle process, there are two ways that performance can be affected: in the overhead associated with running the deswizzle scanner and the additional I/O associated with each read request.

The deswizzle scanner cycles through the blocks of data; each cycle requires a certain amount of bandwidth that may affect system throughput. While the deswizzle scanner is running, I/O requests are still sent to the volume. For writes, the request is stored in NVRAM, and no impact is seen. For read requests, if the physical block location was not updated, an extra I/O is required for that read request to obtain the physical location of the block and update the metadata information for that block. Intensive read workloads may divert resourced bandwidth required by the deswizzle scanner. Therefore, read-intensive workloads can prolong the amount of time it takes for a volume to exit the deswizzle phase and for normal performance expectations to return.

## Summary

This paper provides an in-depth discussion of the operation of DataMotion for Volumes and information on how to utilize it as a cornerstone of nondisruptive operations for clustered Data ONTAP.

## References

The following references were used in this TR:

- Cluster Platform Mixing Rules
  https://fieldportal.netapp.com/Core/DownloadDoc.aspx?documentID=93126&contentID=125939
- Storage Platform Mixing Rules

  https://library.netapp.com/ecm/ecm_get_file/ECMP1644424
- TR-4067: Clustered Data ONTAP NFS Implementation Guide
  www.netapp.com/us/system/pdf-reader.aspx?m=tr-4067.pdf
- Nondisruptive Operations FAQ

  https://fieldportal.netapp.com/Core/DownloadDoc.aspx?documentID=76323&contentID=87206