



# VTL with Deduplication: Best Practices Guide

## Contents

### 3 Introduction

- Audience
- Terminology

### 5 Optimizing Performance

- Optimize the backup stream
- Select the proper virtual tape library configuration
- Assign the robot control responsibilities to the master backup server
- Do not share virtual drives between multiple servers
- Always use VTL compression
- Do not run VTL Cache near capacity
- Run full backups less frequently

### 8 Optimizing Deduplication

- Manage deduplication policy
- Avoid randomizing the data stream
- Audit the tape deduplication ratios
- Put no more than 500 tapes in a deduplication policy
- Run the VTL Console from a server within the same data center as the VTLs
- Expect the overall deduplication ratio to take two or three weeks to reach 10:1

### 11 Optimizing Replication

- Dedicate one or more alternate NICs for replication
- Expect to seed the replication over time
- Ensure that there are enough virtual devices set up between the VTLs and SIRs for post-processing deduplication

### 12 Resource Management

- VTL Cache
- SIR Repository
- SIR RAM
- SIR Index
- Growing your environment
- Automatic Reclamation
- Manage Scratch Pools
- Configuration for Cost-Effective High Availability
- Cluster with Deduplication

### 15 Physical Tape Integration

- Dedicate the FC HBA Initiator ports to the physical resources
- Run exports from the VTL Cache, prior to dedupe
- Manage the scratch pool media
- Reduce the virtual tape size to allow for compression differences
- Set the advanced ATC reclamation parameters

Copyright © 2018 FalconStor Software. All Rights Reserved.

FalconStor Software and FalconStor are registered trademarks of FalconStor Software, Inc. in the United States and other countries.

Windows is a registered trademark of Microsoft Corporation.

All other brand and product names are trademarks or registered trademarks of their respective owners.

FalconStor Software reserves the right to make changes in the information contained in this publication without prior notice. The reader should in all cases consult FalconStor to determine whether any such changes have been made.

PIDVTLWPBPG112018



# VTL with Deduplication: Best Practices Guide

## Introduction

The Virtual Tape Library (VTL) User Guide provides general instructions on how to set up and use the features of the application. It is important to review the User Guide and understand the basics of the VTL application. This best practice guide recommends how to apply VTL in common backup environments. The goal is to utilize the resource in such a way that it takes minimal effort to manage. Environments change over time and data grows, so capacity management is key with any storage resource.

This document primarily focuses on the special capability that the deduplication process adds to VTL. The primary purpose of VTL is to emulate physical tape devices and allow fast write to raw disk. Most people understand the impact of using compression on the data stream and are aware that some data compresses better than others. The same applies to deduplication. If you understand how the process works, you can better utilize the resource.

The VTL version that this guide is based on is 9.00 build 10520-02 released in September, 2018. Backup-application-specific best practices are maintained in separate documents.

### AUDIENCE

This document is intended for new VTL users who have a good understanding of backup applications and a basic understanding of virtual tape libraries. It is assumed that the system has already been architected and deployed in the environment.

### TERMINOLOGY

The VTL environment consists of several components that are normally transparent during normal backup operation. In order to discuss the reasoning behind the recommendations in this document, you need to understand some of the components and technologies involved in the solution. The terminology used in this document is briefly defined below.

<b>VTL-S</b>	Single-appliance solution in which VTL and deduplication processing are performed on a single server.
<b>VTL VTL Cluster Node</b>	The appliance that functions as a VTL server in a cluster; it is associated with a separate Single Instance Repository (SIR) cluster for deduplication.
<b>SIR SIR Cluster Node SIR Cluster</b>	One or more SIR nodes can be clustered with one or more VTL servers to provide a scalable deduplication engine.
<b>VTL Cache</b>	Storage assigned to the VTL. Backup data is written to the VTL cache pending deduplication.
<b>Repository SIR Repository</b>	Storage assigned to the SIR cluster. The repository is dedicated to storing unique data identified during deduplication.
<b>Virtual Tape Virtual Index Tape (VIT)</b>	There are three primary states of a virtual tape in VTL. A 'Virtual Tape' contains backed-up data. After data on the tape has been deduplicated, the tape is referred to as a 'VIT'. It now contains pointers to the data on SIR storage, freeing space for more data.
<b>Folder Lun Metadata Lun</b>	Storage on the SIR nodes that is dedicated to maintaining metadata.
<b>Hash SHA-1</b>	Each block of data has a 20-Byte 'hash value' calculated by the SIR deduplication process to determine and track if the block already has been stored in the SIR repository.
<b>Compression</b>	Data compression utilizes smaller patterns to represent larger patterns within a stream of data, reducing the amount of space required to store the original data.
<b>Deduplication</b>	FalconStor's data deduplication utilizes a SHA-1 hashing algorithm to identify matching blocks of data, insuring that you store them only once and reducing storage requirements. Deduplication works across an entire repository, not simply across a data stream or buffer, thereby getting better reduction than compression alone.
<b>Overall Repository Dedupe Ratio ("redundancy elimination ratio")</b>	The measure of the total amount of data stored compared to the amount of repository storage space actually used. This can be found in the 'Deduplication Results' section of the 'Deduplication Statistics' tab.
<b>Daily Tape Dedupe Ratio</b>	The measure of the total amount of data on a tape compared with the amount of data written to the repository for that specific tape. This can be determined by looking at the 'Tape History' tab or by resetting the 'Deduplication Statistics' at the bottom of the 'Deduplication Statistics' tab.
<b>SIR Reclamation</b>	The process that runs automatically according to parameters, or manually, to delete unneeded data from the SIR index disk and SIR data disk.

# Optimizing Performance

This section describes factors that affect the throughput performance and the utilization performance of FalconStor VTL. Your goal should always be to get as much out of the resource as possible without causing secondary problems. Some recommendations simply address usability concerns in order to prevent extra effort.

## OPTIMIZE THE BACKUP STREAM

This document does not provide details on how to determine bottlenecks that impact how fast data can be sent to VTL. Remember that the backup application can back up the data only as fast as the source storage and server can provide it. Each backup application should have Optimizing Performance documentation.

### *Large block buffering*

FalconStor VTL performs best with large (256K) blocks. Backup applications may default to a 64K buffer depending on the tape drive that you are emulating. Most backup applications allow you to set the block size. SATA storage works best with large block sequential read/write streams.

Consider using backup software's multiplexing to limit concurrent streams

FalconStor VTL lets you emulate 1024 tape drives per VTL, but sending data to them all at once will likely kill the overall performance. SATA storage is optimized for sequential I/O. If you hit the same SATA raid group with too many streams (each active virtual tape drive is a stream), the performance will drop for all the streams.

Your storage configuration for the VTL Cache will dictate how many virtual tape drives you can write to (or read from) without seeing diminishing returns. This recommended maximum should be documented by the solution architect.

Multiplexing can reduce deduplication effectiveness by up to 5%.

### *Consider using NetBackup Accelerator for VMWare virtual machine backups*

FalconStor VTL supports Veritas™ NetBackup™ OpenStorage Technology (OST) and its Accelerator and Optimized Synthesis Backup options. With Accelerator, full backups can be performed at the cost of incremental backups. With the combination of FalconStor VTL and NetBackup Accelerator, white space and data discarded by applications running on the VMWare virtual machine will not be backed up and the penalty incurred by VMWare's Change Block Technology (CBT) can be avoided. The backup time (and the backup window) for VMWare virtual machine backup can be reduced significantly.

## SELECT THE PROPER VIRTUAL TAPE LIBRARY CONFIGURATION

Some backup applications may specify the library configuration they formally support when using a VTL. This requirement is primarily a decision to simplify support and not a reflection on true interoperability. Check with your backup software vendor to make sure they will support you when you call.

### *Create a large number of slots in the virtual library*

Both the number of drives and the number of slots can be set to higher values than the actual physical tape library would support. You can have up to 60,000 in a single library. As long as your backup application is not licensed by the slot, create twice as many slots as you think you will ever use. Increasing the number later requires a restart of the VTL processes.

***Emulate tape drives that you already support***

The type of tape drive you emulate should not affect the actual performance, but if you emulate the same physical drive type as you were using before implementing VTL, you will not need to load different drivers. Size does not matter because you can make the virtual tape size any size you want, unless doing direct export to physical tape.

It is possible that some backup applications will try to set default buffer settings for specific drive types. If in doubt, emulate a newer drive type that is more likely to have larger buffer settings.

***Test the tape driver by stopping and starting the VTL processes***

In rare situations, a tape driver has been able to hang a (Windows 64-bit cluster) server. Test the tape driver with the emulated tape drives by stopping and restarting VTL, causing the devices to go away and return. This is a problem with the driver and not VTL.

***Enable Capacity-On-Demand (COD)***

When utilizing deduplication, you virtualize the capacity and tapes reduce to a smaller VIT automatically. This allows you to create more virtual tapes than you have physical storage for. The VTL User Guide describes the algorithm that is used to expand the virtual tapes on VTL storage.

The primary reason not to use COD is when you are not deduplicating and you require that the backup application be able to report the actual media usage. If you have specific storage that should be used for specific tapes, create a separate library without COD and create static tapes from the specific storage.

***Note on tapes inheriting virtual library properties***

In earlier versions of VTL, you could create some virtual tapes that did not use COD and put them in a library that was configured with COD. When the virtual tape is rewritten, it automatically inherits the tape properties of the library it is in. This is convenient if you want to change your default tape size, but it eliminates the ability to mix tape libraries.

***Set the virtual tape initial size to 1G***

FalconStor VTL does not assign storage to the virtual tape until it is mounted for the first time.

Even if you have expired and scratched/erased the tape, you still utilize the initial size on disk. 2000 x 1G = 2 TB of space. The only way to recover this data is to actually delete and recreate the tape (see Manage Scratch Pools).

***Set the virtual tape incremental tape size to 10% of the full tape size***

The smallest incremental size you can set is 1/64th the full tape size. The larger the incremental size, the larger the potential unused space that may be assigned to a specific tape. Once the virtual tape is deduplicated, the resulting VIT is reduced to a minimum of 1 GB. Making the incremental size too small adds unnecessary overhead (although very little) to the VTL as it assigns the next segment.

**ASSIGN THE ROBOT CONTROL RESPONSIBILITIES TO THE MASTER BACKUP SERVER**

FalconStor VTL can talk to ACSLS-controlled physical libraries. It can NOT emulate an ACSLS controlled library. Therefore there will always be one server to manage the robotic control for the virtual library. This is a single point of failure, so why not assign it to another single point of failure... the master backup server (the server managing the catalog/database for the backup application).

FalconStor VTL supports a Hosted Backup option that allows most backup applications to be loaded on the VTL server. This is not a recommended solution in most environments because it adds complexity to VTL configuration, including the need to certify the backup application on the VTL operating system.

### **DO NOT SHARE VIRTUAL DRIVES BETWEEN MULTIPLE SERVERS**

Many backup applications provide features that allow a single physical device to be presented and used by multiple servers. This is done with the intention to increase utilization of a single device. Unfortunately, this causes complications when attempting to troubleshoot stability in the environment. FalconStor VTL can emulate 1024 virtual tape drives at no additional cost.

Although VTL supports drive sharing, it is recommended that you dedicate the virtual drives to a specific server and use scheduling to insure that you are not trying to run too many drives at one time.

### **ALWAYS USE VTL COMPRESSION**

Compression reduces the amount of VTL Cache storage utilized during backups.

Software compression is best for the smaller VTL configurations where the reduced throughput actually helps avoid over-running the deduplication processing and filling up the VTL Cache.

*Note: The unique blocks are always compressed in the SIR Repository.*

### **DO NOT RUN VTL CACHE NEAR CAPACITY**

Just as with your home or office PC, if you start running short on disk space, the system will start to slow down. The best time to add storage to the VTL Cache is before it is maxed out. The best performance occurs when you have many LUNs and RAID groups available to share the load. If you fill up four RAID groups and add another two RAID groups, you will write only to the two new RAID groups.

If you have the time, there are processes for moving data between LUNs (raid groups) to balance the system. But if the system is busy trying to keep up because its performance is poor, you may find yourself in a "Catch-22" position.

### **RUN FULL BACKUPS LESS FREQUENTLY**

The improved recovery time from VTL eliminates most of the justification for taking full backups every week or every day. If you stagger your full backups every other weekend, you essentially double your throughput capacity. Running incremental backups reduces the amount of data that is actually transferred, thereby reducing the amount of time required to run the backups and deduplication.

Don't be surprised if your deduplication ratios drop when you do fewer full backups. In all practicality, the same amount of storage will be used. It is only a question of how much data you have to process to get there.

# Optimizing Deduplication

## MANAGE DEDUPLICATION POLICY

To maximize deduplication performance and minimize storage needs, FalconStor offers three deduplication methods:

- Inline deduplication – for tape and NAS deduplication
- Turbo deduplication – for tape deduplication
- Post-processing deduplication – for tape deduplication

The following table compares the different deduplication methods:

	Inline Deduplication	Turbo Deduplication	Post-Processing
<b>Tape or NAS Deduplication</b>	<ul style="list-style-type: none"> <li>• Available for tape deduplication.</li> <li>• Used for NAS deduplication</li> </ul>	<ul style="list-style-type: none"> <li>• Available for tape deduplication</li> </ul>	<ul style="list-style-type: none"> <li>• Available for tape deduplication</li> </ul>
<b>Storage and its Bandwidth Requirements for Backup Landing Area and SIR Repository</b>	<ul style="list-style-type: none"> <li>• Requires the least amount of landing storage.</li> <li>• As the deduplication ratio increases over time, overall backup/ deduplication performance will increase because fewer writes will need to be made to the repository.</li> <li>• Requires the least storage bandwidth with reasonable deduplication ratio</li> </ul>	<ul style="list-style-type: none"> <li>• Requires appropriate capacity for your daily backups plus additional space for VITs (about 2% of the size of the pre-deduplication capacity).</li> <li>• Pre-processing can significantly improve deduplication performance as the backup data to tape won't be read again in the later deduplication stage. Mixed disk write and read requests can be reduced.</li> <li>• Require high Backup Landing storage bandwidth and spindles</li> </ul>	<ul style="list-style-type: none"> <li>• Requires appropriate capacity for your daily backups plus additional space for VITs (about 2% of the size of the pre-deduplication capacity).</li> <li>• Deduplication and backup ingest can create disk contention as the former performs write and the latter performs read.</li> <li>• Require high Backup Landing storage bandwidth and spindles</li> </ul>
<b>CPU Processing Power</b>	Requires the most CPU computation power (12 cores or more recommended).	The more CPUs in the VTL server, the less impact to backup performance.	<ul style="list-style-type: none"> <li>• CPU requirements can be somewhat relaxed (8 cores for example).</li> <li>• Can utilize external SIR cluster nodes for deduplication.</li> </ul>
<b>Can be used with Tape Caching</b>	No. Inline deduplication and Tape Caching are mutually exclusive. To reduce disk contention, data written to cached tapes will not be deduplicated in real time and will follow the schedule of the tape caching policy.	Yes	Yes

## AVOID RANDOMIZING THE DATA STREAM

### *Never use client-side compression*

Client-side compression randomizes data from one backup to the next, causing low deduplication ratio performance.

Most backup applications allow you to enable client-side compression to reduce the bandwidth required to send it to the backup server. Small changes in the data result in a complete change in the next backup's data stream, eliminating the commonality. FalconStor VTL knows how to uncompress the VTL compression before deduplicating the data, but it does not accommodate client-side compression.

Many databases also have the ability to compress their data before sending it to the backup server; this also causes low deduplication ratio performance.

### *Never use encryption*

Encryption randomizes the data from one backup to the next, causing low deduplication ratio performance.

### *Do not bother deduplicating data that will not dedupe well*

One benefit of post-processing deduplication is that you do not have to deduplicate everything. If the data is already compressed and won't dedupe, you are not going to save any space, so don't waste time trying. Increase the VTL Cache storage and keep it there. If you don't have time or bandwidth to replicate it, you should consider a CDP solution with replication.

### *Remember that most TSM environments use client-side compression*

Even if you turn off client-side compression when you start to use VTL, consider that any data that is migrated from the old primary storage pool into the new virtual tape library is likely to have been compressed at the client side. This leads to lower overall deduplication ratios. It is best to force new full backups of the environment (without client-side compression) and let the majority of the old pool expire.

## AUDIT THE TAPE DEDUPLICATION RATIOS

During the initial rollout, and periodically through the year, you should audit the Daily deduplication ratios of your backup tapes. Any tape with a deduplication ratio of less than 3:1 should be investigated as follows:

1. Identify what type of backup data is written to the tape.
2. Confirm that the data stream is not being modified before reaching VTL by doing two consecutive backups of the same data, deduplicating the tape after each backup. The second backup should have a very high deduplication ratio. If not, check for compression and encryption on the source (see Always use VTL compression and Never use encryption).
3. If the data continues to get poor deduplication performance, contact FalconStor Technical Support to determine if a special parser enhancement may be required. An example would be a database agent that puts a variable-length header on the backup streams. Or, a backup application may make a change that might require an updated parser.
4. If you have the VTL Cache space, don't dedupe the poorly performing tapes until the issue is resolved.

### *Deduplication ratios include compression*

All deduplication vendors report deduplication ratios as the total of the reduction due to dedupe plus the reduction due to compression. This is helpful to understand when evaluating the poorly deduping tapes.

An 8:1 deduplication ratio can be 2:1 compression and 4:1 deduplication, or 4:1 compression and 2:1 deduplication. The key thing to remember is that any tape reporting anything near a 1:1 deduplication ratio is most likely compressed data.

This issue applies to most vendors' deduplication.

### **PUT NO MORE THAN 500 TAPES IN A DEDUPLICATION POLICY**

VTL Console performance slows down when managing large policies. Each virtual tape has a lot of metadata that must be loaded into the console, which causes frustrating delays. Network performance between VTL and the console can have a serious impact on performance.

### **RUN THE VTL CONSOLE FROM A SERVER WITHIN THE SAME DATA CENTER AS THE VTLS**

Use a remote console interface to a server that has the VTL Console and resides in the same data center (and preferably network) as the VTL and SIR servers. This will insure the best performance should the amount of metadata increase or if you have a poor remote network connection.

### **EXPECT THE OVERALL DEDUPLICATION RATIO TO TAKE TWO OR THREE WEEKS TO REACH 10:1**

Assuming that you back up the entire environment once in 7 days, it will take a week or two of retention to see the overall dedupe ratio climb to near 10:1. A single copy of the environment will have some commonality and compression, but rarely more than 4:1. The second and third copies are where you see the greatest benefits.

Some environments that do full backups every day can see over 30:1 dedupe ratios, but this is a sign of an inefficient backup environment.

## Optimizing Replication

FalconStor VTL replicates data after it has been deduplicated by replicating the VIT tapes (metadata) between VTLs. The remote SIR component resolves the VIT and pulls the blocks that do not already exist in the remote site. When completed, the new VIT is copied into the vault and is ready for use.

### **DEDICATE ONE OR MORE ALTERNATE NICs FOR REPLICATION**

Keep replication traffic off your primary production network. Each VTL and SIR server can be configured with additional NICs on a private subnet so that you can route the traffic separately.

#### *Network latency and bandwidth matter*

There is no magic in moving data across a WAN. It is important to dedicate the write bandwidth and quality of WAN to move the necessary data.

### **EXPECT TO SEED THE REPLICATION OVER TIME**

You will not immediately benefit from the deduplication during replication until the remote repository is populated. The easiest solution is to gradually add systems to the replication policies so you do not exceed your available bandwidth and get behind.

### **ENSURE THAT THERE ARE ENOUGH VIRTUAL DEVICES SET UP BETWEEN THE VTLS AND SIRs FOR POST-PROCESSING DEDUPLICATION**

FalconStor will tune the environment to obtain the best possible replication performance. When post-processing deduplication is occurring, each tape being processed utilizes two virtual tape devices that are presented to the SIR server. When the SIR resolves the replicated VITs, it takes two virtual tape devices. When you are replicating to a VTL environment that is also doing backups, you may not have the proper number of virtual devices, which can stifle the replication performance.

## Resource Management

In order to get the most out of the Falconstor VTL with deduplication, you should understand the resources involved. With straight VTL, you only had to manage the VTL Cache space. Now, in order to reduce overall storage and to be able to replicate, you need to manage the SIR as a unique optimized storage resource. You should understand how both RAM and storage affect SIR capacity.

If you do not want to get deep into this subject, just recognize that you should monitor the following resources:

1. VTL Cache - Backups go here
2. SIR Repository - Unique data is stored here
3. SIR RAM (memory) - Hash lookup index runs here
4. SIR Index Space - Stores the Hash Index

Built-in reports can be configured to periodically send emails so that you do not have to log into the consoles (see the Email Alerts feature).

### VTL CACHE

The VTL Cache is needed to store non-deduped data and the metadata (VITs). If you run out of VTL Cache, your backups will stop.

### SIR REPOSITORY

The SIR Repository stores the unique blocks in a compressed format. As the repository fills, you need to periodically run SIR Reclamation to clean out the data that can be deleted. If the repository fills, all deduplication will stop/fail. The VTL Cache will thus start to fill up and eventually you will not be able to perform backups.

### SIR RAM

In order to dedupe at sufficient speeds, the index table containing all the hash values is kept in memory. Each block has a SHA-1 hash value of 20 bytes calculated, and then checked against the previously encountered hash values (in memory). So every block stored in the repository will have a 20 byte hash (plus 5-7 bytes) that needs to be stored in memory.

When FalconStor states how much memory is required to support a certain amount of repository, they are estimating the average compression rate of the blocks stored. If your blocks compress more, you will use less repository space. But you will use the same amount of memory for the same number of blocks.

At present, there is no obvious way to know your average compression on the repository blocks, but this variable can cause either the RAM or the storage to fill up first. The rough sizing estimate is that every 2 GB of RAM supports 1 TB of SIR Repository.

### SIR INDEX

The SIR Index resource is the storage required to store the hash index information that is running in RAM. This space is also used during SIR Reclamation and is cleaned up as a separate Index Reclamation process. In most cases, this storage is over-sized enough not to be an issue.

## GROWING YOUR ENVIRONMENT

### VTL Cache

You can add LUNs to VTL at any time, but pay attention to the issues mentioned early in the section and do not run VTL Cache near capacity. FalconStor capacity licenses may apply.

**Folder and Index Luns**

Additional Folder and Index storage are normally added as 4% of any additional SIR repository storage.

**SIR Repository**

It is important to add storage to the SIR repository in the same number of LUNs per the hash columns that were originally assigned to the repository. The details are beyond the scope of the document, but consult FalconStor for the minimum and recommended incremental storage configurations.

Data is purged from the SIR Repository by a background process called SIR Reclamation (unofficially known as ‘garbage collection’). This process consists of two phases: one phase clears up the repository data space, while the second cleans up the index space. They can be triggered manually, scripted to run at a certain time, or left to trigger based on capacity. Greater detail can be found in the VTL User Guide.

**AUTOMATIC RECLAMATION**

Make sure that automatic reclamation is enabled, which will trigger reclamation based on utilization watermarks. If you look at the Deduplication Repository tab (below), you should see the markers. You can trigger reclamation manually using the VTL Console.

**DEDUPLICATION REPOSITORY USAGE**

**Index Cache Capacity**



80% threshold

■ Used before reclamation	0%
■ Used since reclamation	0%
■ Free	100%

**Folder Disk Capacity**



80% threshold

■ Used before reclamation	0%
■ Used since reclamation	0%
■ Free	100%
Folder disk capacity	84GB
Folder disk used	2 MB

**Index Disk Capacity**



50% threshold

■ Retained by pruning	0%
■ Used since pruning	0%
■ Free	100%
Index disk capacity	16GB
Index disk used	32KB
Index pruning:	Idle

**Data Disk Capacity**



50% threshold

■ Used before reclamation	0%
■ Used since reclamation	0%
■ Free	100%
Data disk capacity	200GB
Data disk used	2 MB
Space reclamation:	Idle

**MANAGE SCRATCH POOLS**

Tapes that have been expired by the backup application, but not re-used, are still using storage in the VTL and SIR repository. When running near capacity or when wanting to get the

most space back during SIR Reclamation, you should trigger the backup software to overwrite the start of the scratch tapes. This will signal VTL that the data is no longer needed.

The process for doing this will vary depending on the backup application. Some have options to reinitialize the scratch tapes, while others can be told to re-label the scratch pool using a script. Insure that the backup application process is actually rewriting from the start of the virtual tape.

### *CommVault Critical Process Change*

When Commvault ages tapes, it moves the tapes to the Scratch Media Pool (does not load into a tape drive). When Commvault is ready to backup to a tape, it selects a tape from the Scratch Media Pool. Commvault starts writing data at the end of the On Media Label (Starts at an unknown mark within the tape) to the end of the tape. The OML is never overwritten or rewritten even if the tape is to be recycled. This practice will not allow Reclamation to free up any SIR storage.

There is a simple solution to this problem. The directions below are assuming that the backup administrator has experience in using Commvault and are not step by step directions.

1. Edit the Commvault storage policy going to the VTL, set all the media that is recycled to be marked to be erased also.
2. Right-click on each virtual tape library system within Commvault and select **Erase Spare Media** schedule the process of erasing those media flagged to be erased. Schedule the media erasing to occur between the time most backups complete and the time DeDuplication occurs.

*Note: Erasing all tapes at once does not put an exorbitant load on the system. The amount of time it takes will depend on how many tapes the customer is erasing (less than a minute per tape).*

3. Step 2 automates the process of erasing tapes that are to be recycled moving forward. However, it does not address the current scratch pools in the environment. You'll need to manually erase each of these tapes (one at a time).

Overall goal of the directions above is to have Commvault select a tape that has been erased by Commvault instead of overwriting a tape.

### *RAM and SIR Clusters*

The RAM and cluster server configurations must remain identical in the SIR Cluster. Some customers will deploy more SIR cluster nodes, but only half populate them with RAM. It is then much easier to add RAM and storage to the servers than add additional servers to the cluster.

## **CONFIGURATION FOR COST-EFFECTIVE HIGH AVAILABILITY CLUSTER WITH DEDUPLICATION**

A high availability cluster with deduplication can be formed with a pair of active-active HA nodes running VTL plus an external SIR cluster. Using external SIR cluster for deduplication can offload VTL nodes from performing post-processing deduplication but incurs additional cost for SIR heads nonetheless.

For cost sensitive deployments, an alternative to the above way is to form an active-active VTL-S HA pair. No external SIR cluster is required and hardware cost be reduced.

## Physical Tape Integration

FalconStor VTL has the ability to directly interface with a physical tape library, including ACSLS controlled libraries. When deploying VTL in a standard configuration, the backup servers are utilized to copy the data from VTL through the server to the physical tape devices. Several advanced features, such as Automatic Tape Caching, allow VTL to send data directly to a physical tape drive without involving the backup application.

This document does not attempt to recommend when to architect one of the advanced configurations. If you intend to utilize these advanced features, there are a few best practices that you should follow.

### DEDICATE THE FC HBA INITIATOR PORTS TO THE PHYSICAL RESOURCES

There are several potential problems that can occur, beyond simple bandwidth contention, when you share an HBA Initiator port with the physical tape library resources and other storage devices.

### RUN EXPORTS FROM THE VTL CACHE, PRIOR TO DEDUPE

If you are cutting physical tape copies at the local site, schedule deduplication to run AFTER the exports or migration of data occurs. (Automatic Tape Caching policies do this by design.) Exports of deduped tapes put a load on the SIR Repository storage that may slow down other deduplication.

### MANAGE THE SCRATCH POOL MEDIA

When the backup application re-uses an expired tape in the scratch pool, it causes VTL to temporarily mount the physical tape until it realizes that the tape is being overwritten. This can interrupt other export jobs that are using the physical tape drive.

Re-initialize the scratch pool during a period when you have available physical drives, causing the disk cache to be re-established. The next time the backup application requests the tape, the physical tape will not be mounted.

### REDUCE THE VIRTUAL TAPE SIZE TO ALLOW FOR COMPRESSION DIFFERENCES

Set the virtual tape max size to 15% less than the native capacity of the physical tape you are exporting to. This will waste a little tape, but will ensure that you never run out of physical tape due to differences in the compression engines.

### SET THE ADVANCED ATC RECLAMATION PARAMETERS

Automatic Tape Caching (ATC) reclamation policies can be modified by parameters that are not in the VTL Console. These include the ability to reclaim only full tapes or to set a percentage. Do not combine ATC Reclamation with SIR Reclamation. Consult FalconStor Technical Support for proper configuration, especially when utilizing the watermark triggers.

Information in this document is provided "As Is" without warranty of any kind, and is subject to change without notice by FalconStor, which assumes no responsibility for any errors or claims herein. Copyright 2018 FalconStor Software. All rights reserved. FalconStor Software, FalconStor and RecoverTrac are trademarks or registered trademarks of FalconStor Software, Inc. in the United States and other countries. All other company and product names contained herein are or may be trademarks of the respective holder. PIDVTLWPBPG112018

## FALCONSTOR®

**Corporate Headquarters**  
701 Brazos Street, Suite 400  
Austin, TX 78701  
Tel: +1.631.777.5188  
salesinfo@falconstor.com

**Europe Headquarters**  
Rosa-Bavarese-Strasse 3  
80639 München, Germany  
Tel: +49 (0) 89.41615321.10  
salesemea@falconstor.com

**Asia Headquarters**  
Room 1901, PICC Office Tower  
No. 2 Jian Guo Men Wai Avenue  
Chaoyang District Beijing 100020 China  
Tel: +86.10.6530.9505  
salesasia@falconstor.com