White Paper

EMC Symmetrix VMAX Virtual Provisioning Space Reclamation and Application Considerations

Applied Technology

Abstract

This white paper discusses EMC® Symmetrix® VMAX™ Virtual Provisioning™ space reclamation functionality in the context of server applications, detailing the scenarios in which applications and data migration tools write zeros that can subsequently be reclaimed from virtually provisioned devices.

February 2011

Part Number h6730.3

Table of Contents

Executive summary **4**
Introduction **4**
- Audience 4
- Terminology 4

Zero space reclamation **5**
Executing Virtual Provisioning reclamation **5**
- Solutions Enabler 5
- Symmetrix Management Console 7

Application considerations **10**
- VMware ESX Server 10
- VMware Virtual Infrastructure 3 10
- VMware vSphere 4 11
  - Migrating from Virtual Infrastructure 3.x to vSphere 4.0 11
  - Considerations for space reclamation in VMware environments 12
- Oracle 12
  - REDO log files 12
  - Datafiles 12
  - ASM considerations 12
  - Reclaiming space from deleted Oracle objects 13
  - Summary of Oracle considerations 13
- Microsoft SQL Server 13
  - Transaction log files 13
  - Database files 13
- Microsoft Exchange Server 15
  - Microsoft Exchange log files 15
  - Microsoft Exchange database files 16
- Microsoft Hyper-V 16

Migration and replication **17**
- File system considerations 17
- Open Migrator/LM 18
- EMC Open Replicator 18
- TimeFinder/Clone thick to thin replication 19
- Symmetrix Remote Data Facility (SRDF) 19
- PowerPath Migration Enabler 20
- Veritas Storage Foundation SmartMove 20

Space reclamation from previously used devices **21**
- Online reclamation of previously used space 21

Conclusion **22**
References **22**

Executive summary

Organizations continually search for ways to both simplify storage management processes and improve storage capacity utilization. When provisioning storage, administrators must consider and otherwise estimate an application's future capacity requirements in addition to current needs. Estimating storage requirements is a difficult process and can lead to miscalculations in the allocation of storage. Administrators frequently choose to overallocate in order to avoid the disruptions and management overhead of exhausting available capacity. Overallocation results in unused storage space, which carries inefficiencies and unnecessary costs.

EMC® Virtual Provisioning™ helps to address these challenges by providing a mechanism of presenting large "thin" LUNs to a host while consuming physical storage from a shared pool only as needed. Symmetrix Virtual Provisioning can improve storage capacity utilization and simplify storage management by presenting an application with sufficient future capacity, reducing the need to frequently provision new storage, while also avoiding costly allocated but unused storage. Virtual Provisioning also simplifies data layout, with automated wide striping that enables organizations to achieve equivalent performance to standard provisioning, with less planning and labor required.

One of the goals when deploying Virtual Provisioning is to ensure the applications and migration tools used to store or to move data from standard or fully provisioned environments to thin environments do not cause unneeded storage allocations. In some cases an application may write a contiguous series of zeros to represent available or initialized but unused space for a volume or file in a file system. When a standard environment is overallocated and a datafile or volume has a high amount of unused space, an application that creates or copies this unused space to a thin device in the form of zeros can cause inefficient space utilization.

To help address this specific concern, Symmetrix® Virtual Provisioning now enables Symmetrix VMAX™ users to automatically reclaim "chunks" containing all zeros. This is most beneficial after migrating from standard volumes to thin volumes but can also prove beneficial in some application scenarios, in order to reduce capacity requirements and TCO.

Introduction

Many applications have the potential to write zeros to free space as a part of standard initialization, allocation, or migration processes. Depending on the way zeros are written, there is the potential to reclaim the storage space allocated due to these processes. This white paper discusses some of the most common applications and environment scenarios in which zeros will be written to storage devices. In addition, this paper will outline when it is possible to reclaim this all-zero space and how to reclaim storage using EMC Solutions Enabler or the Symmetrix Management Console.

Audience

This white paper is intended for server and application administrators, storage architects, customers, and EMC field personnel who want to understand application considerations for Virtual Provisioning space reclamation on the Symmetrix VMAX.

Terminology

Term	**Description**
Device	A logical unit of storage defined within a Symmetrix array.
Thin Device	A host accessible device that has no storage directly associated with it.
Data Device	An internal device that provides storage capacity to be used by thin devices.
Thin Device Extent	The minimum quantum of storage that must be mapped at a time to a thin

	device. Extent also is known as a “chunk.”
Data Device Extent	The minimum quantum of storage that is allocated at a time when dedicating storage from a thin pool for use with a specific thin device
Thin Pool	A collection of data devices that provide storage capacity for thin devices.
Bind	The process by which one or more thin devices are associated to a thin pool.
Unbind	The process by which a thin device is disassociated from a given thin pool. When unbound, all previous extent allocations from the data devices are erased and returned for reuse.
Pre-allocating or Pre-provisioning	User specified operation performed against a thin device for the purposes of reducing the operational impact of allocating extents, or for guaranteeing a specified amount of storage for a thin device in a thin pool

Zero space reclamation

Solutions Enabler 7.1, in conjunction with the Enginuity™ 5874 Q4 service release, was the first Symmetrix software release to provide the ability to free up, also referred to as “de-allocate,” storage ranges found to contain all zeros. Space reclamation was designed to be used primarily following a replication or migration activity from a regular device to a thin device, in which software tools such as Open Replicator and Open Migrator (for certain environments) copy all-zero, unused space to the target thin volume. Enginuity 5875 further extends the Virtual Provisioning space reclamation capabilities of the VMAX platform to include on-the-fly zero detection during Open Replicator and Symmetrix Remote Data Facility (SRDF®) migration scenarios.

Zero space reclamation is an extension of the existing Virtual Provisioning space de-allocation mechanism. Previous versions of Enginuity and Solutions Enabler allowed the reclamation of allocated but unused thin device space from a thin pool. Administrators now have the ability to reclaim all-zero space, including both host-unwritten extents/chunks and chunks that contain all zeros due to file system or database formatting methods. The space reclamation process is nondisruptive and can be executed with the targeted thin device ready and read/write to operating systems and applications.

Starting the space reclamation process will spawn a back-end disk director (DA) task that will examine the allocated thin device extents on specified thin devices. A thin device extent is 768 KB (or 12 tracks) in size and is the default unit of storage at which data is allocated for a thin device. For each allocated extent, all 12 tracks will be brought into Symmetrix cache and examined to see if they contain all zero data. If the entire extent contains all zero data, the extent will be de-allocated and added back into the pool, making it available for a new extent allocation operation. Should any portion of the extent contain non-zero data, none of the extent will be reclaimed.

Currently there are some restrictions on space reclamation with Symmetrix-based local and remote replication. Space reclamation is not supported on actively replicating SRDF volumes. The link must be suspended prior to running the reclamation operation. Additionally, reclamation will not be performed on tracks that participate in a local replication session, including TimeFinder®/Clone and TimeFinder/Snap.

Executing Virtual Provisioning reclamation

Solutions Enabler

Virtual Provisioning space reclamation can be executed using the Solutions Enabler version 7.1 or later “symconfigure” command line interface (CLI). Within symconfigure the “free tdev” syntax should be used with the thin device being the target for reclamation. As a part of the free tdev syntax, a range of cylinders representative of the thin device needs to be specified, based on a beginning and end cylinder value, or size in cylinders, megabytes, or gigabytes. The most common scenario will be to target an entire thin device for reclamation, which can be specified by a starting cylinder of 0 and an ending cylinder of "last_cyl."

The free tdev command offers two types of reclamation options. The first type, called "unwritten," was available with previous versions of Solutions Enabler and Enginuity. Reclaiming unwritten space is equivalent to reclaiming pre-allocated but unused storage within a thin pool. Pre-allocating storage is a user-specified operation generally performed to guarantee space for a thin device in a pool, or to reduce the operational impact of allocating extents. Pre-allocation can be done when binding a thin device to a pool, or after a thin device is bound with the "allocate tdev" symconfigure command.

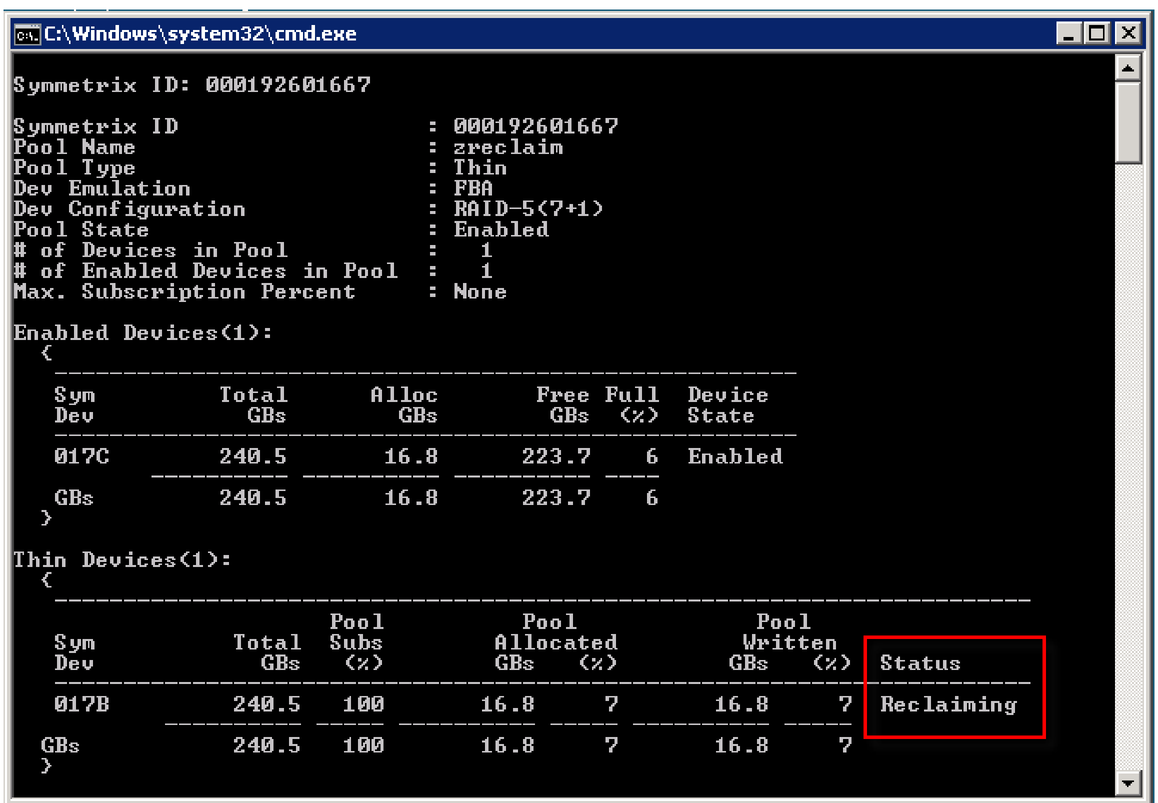
The second and new form of reclamation is called "reclaim." The new reclaim operation will perform the previously mentioned reclamation of pre-allocated but unused (unwritten) space, as well as perform the reclamation of space where a thin device extent contains all zero data.

The example below provides the appropriate syntax for starting the reclamation process against an entire thin device, 17B, using the symconfigure command.

symconfigure -cmd "free tdev 17b start_cyl=0 end_cyl=last_cyl type=reclaim;" commit -sid 67

The symconfigure command will kick off a background process to perform the specified reclamation. Once initiated the status of the thin device will change from a "bound" state to a "reclaiming" state. One method to view the status of the reclamation process and validate the reclaiming state is to look at the detail of the thin pool where the thin device resides. This can be accomplished by running the symcfg show command against a thin pool with the "-detail" option. The syntax of the command is below, with the resulting output shown in Figure 1.

symcfg show -thin -pool zreclaim -detail -sid 67 –gb



```
C:\Windows\system32\cmd.exe

Symmetrix ID: 000192601667

Symmetrix ID                    : 000192601667
Pool Name                       : zreclaim
Pool Type                       : Thin
Dev Emulation                   : FBA
Dev Configuration               : RAID-5(7+1)
Pool State                      : Enabled
# of Devices in Pool            :    1
# of Enabled Devices in Pool    :    1
Max. Subscription Percent       : None

Enabled Devices(1):
  {
   ----------------------------------------------------------
   Sym         Total      Alloc        Free Full  Device
   Dev           GBs        GBs         GBs  (%)  State
   ----------------------------------------------------------
   017C        240.5       16.8       223.7    6  Enabled
          ---------- ---------- ---------- ----
   GBs         240.5       16.8       223.7    6
  }

Thin Devices(1):
  {
   ---------------------------------------------------------------------------
                          Pool         Pool              Pool
   Sym          Total     Subs       Allocated          Written
   Dev            GBs      (%)      GBs     (%)      GBs     (%)  Status
   ---------------------------------------------------------------------------
   017B         240.5     100      16.8      7      16.8      7   Reclaiming
           ---------- ----- ---------- ----- ---------- -----
  GBs           240.5     100      16.8      7      16.8      7
  }
```

Figure 1. symcfg show command to display the reclaiming thin device state

The reclamation process can also be queried by using the "symcfg verify" command. The verify command has been enhanced to include a "reclaiming" parameter for the purposes of verifying whether a thin device has entered a reclaiming state. The verify command can be useful when scripting operations, which will check to ensure a given thin device has begun reclaiming. An example for running the symcfg verify command is next:

symcfg verify -tdev -reclaiming -dev 17b -sid 67

Once a thin device has completed the reclaim background task it will change status from reclaiming back to bound. The same symcfg show and symcfg verify commands as previously mentioned can also be used to query and otherwise verify that a thin device has completed reclamation and returned to a bound state.

Symmetrix Management Console

The Symmetrix Management Console (SMC), with version 7.1 or later, can also be used to easily initiate and monitor the Virtual Provisioning space reclamation process. One method of utilizing SMC to perform this operation is to browse the tree view and highlight the thin device (TDEV) folder located under the Devices folder. In the Properties pane, a given thin device can be highlighted and targeted with a right-click of the mouse to display the control menu. Within the control menu, select **Device Configuration> Allocate/Free Thin Device** as depicted in Figure 2.

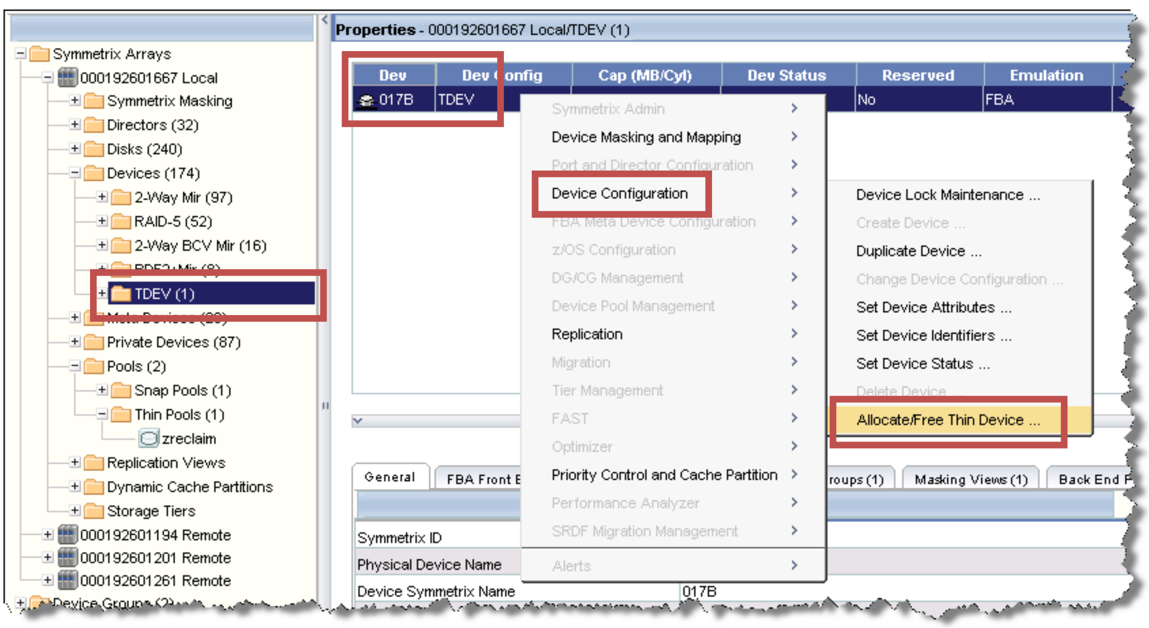


Figure 2. In SMC, selecting the Allocate/Free Thin Device option

Once Allocate/Free Thin Device is selected, a dialog box, as shown in Figure 3, will open to allow the selection of specific options. For space reclamation, the first option selected should be the **Free** radio button. From there one can select the Unwritten or Reclaim options as discussed previously. In order to reclaim zeroed space, the Reclaim radio button should be selected. Additionally the option to target the full thin device or a range of the thin device based on cylinders is available. Once the appropriate selections are specified, click the **Add to Config Session List** button to proceed.

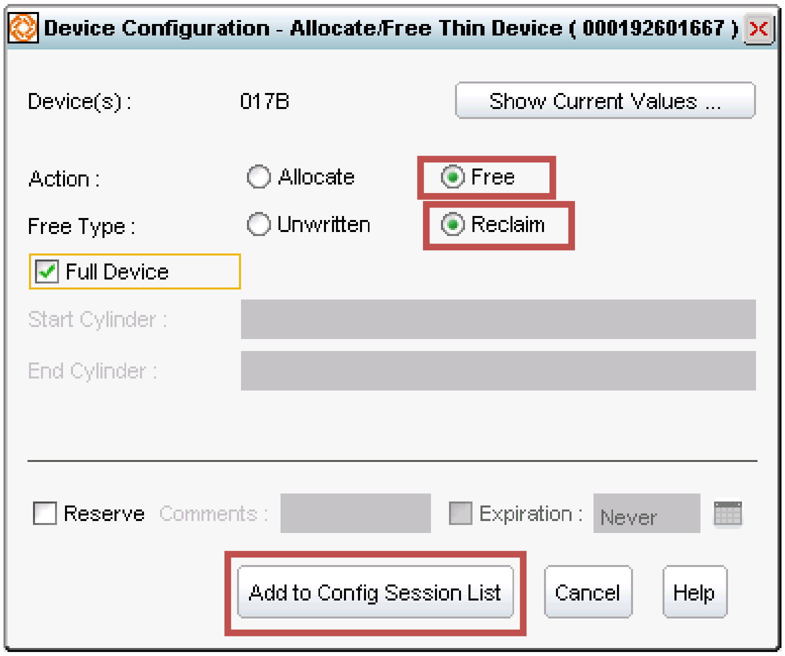


Figure 3. Allocate/Free Thin Device configuration change dialog box

The act of clicking **Add to Config Session List** will create a configuration change task that needs to be executed in order to initiate the reclamation process. SMC should automatically navigate to the configuration session task window upon the creation of the task. To start the task, click the **Commit All** button as shown in Figure 4.

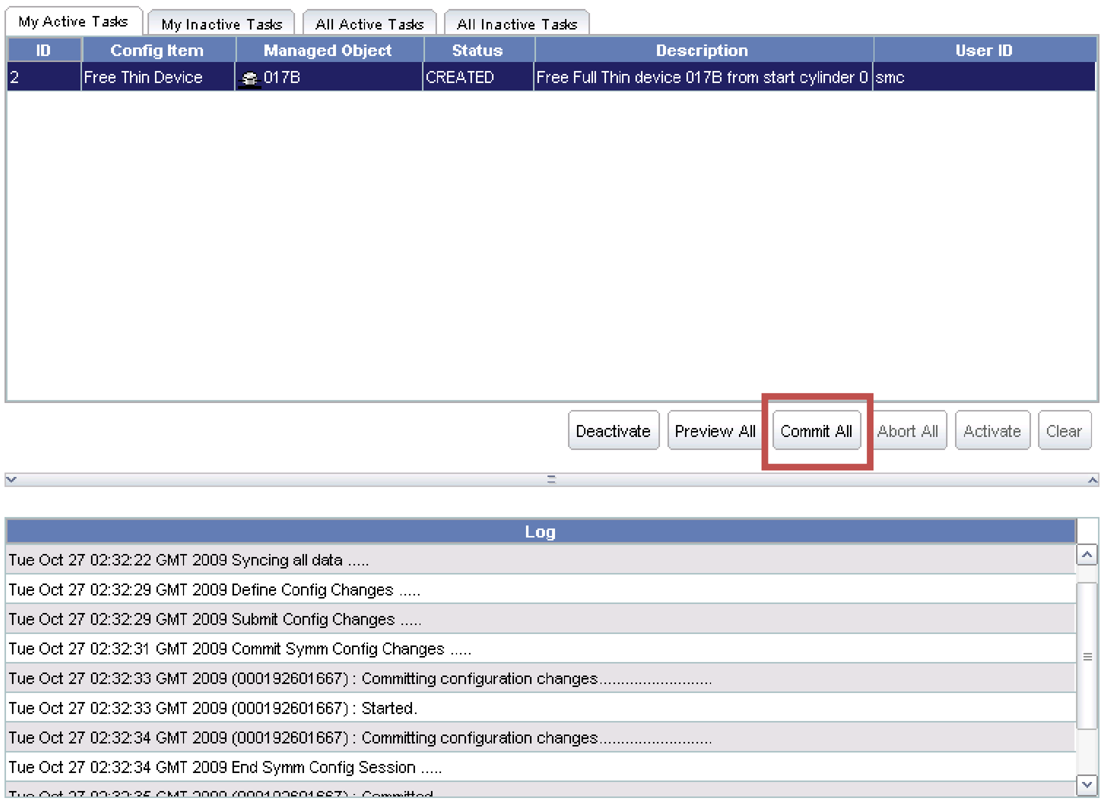


Figure 4. Space reclamation configuration change task

After the configuration change is committed, the reclamation process will operate as a background task in the Symmetrix. To verify that reclamation has started for a given thin device, navigate to the Thin Pools folder in the tree view and select the thin pool where the thin device targeted for reclamation resides. Once selected, click the **Bound Thin Device** tab in the Properties pane to view the status of the thin device. The thin device should report a status of reclaiming, like in Figure 5. Be sure to refresh the view in SMC to get the most up-to-date status for the selected thin pool.

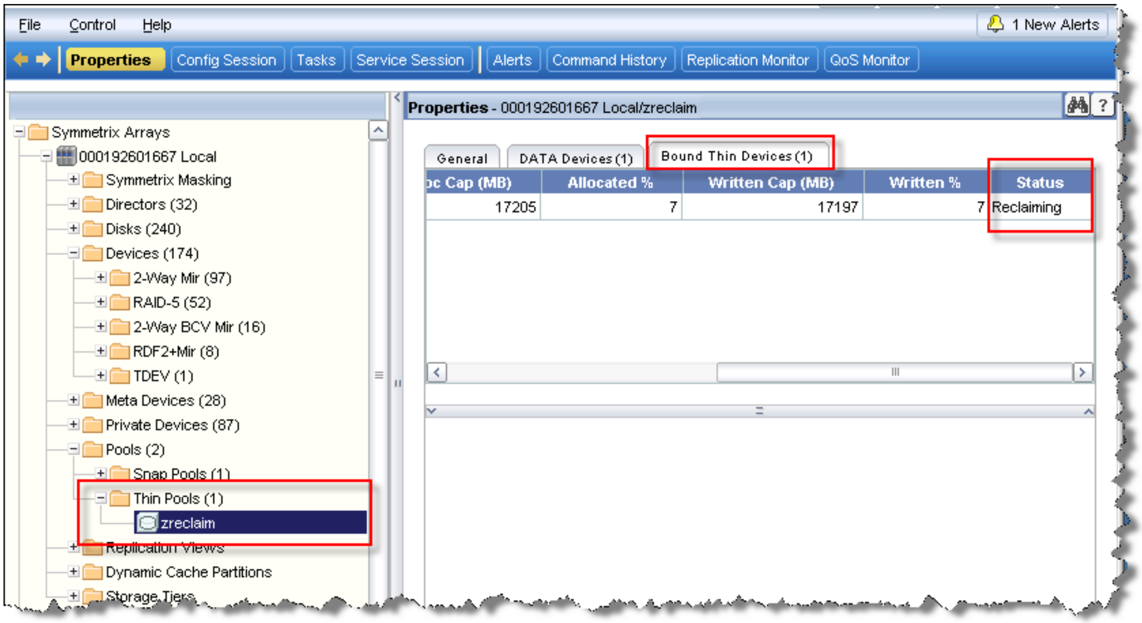


Figure 5. Viewing the reclaming status of a thin device

Once the reclamation process is complete, the status of the device will return to bound. Refreshing the view in Figure 5 will reflect the change in status upon completion.

Application considerations

At times of initialization or as a part of cleanup operations, applications may write contiguous areas of zeros to datafiles or file systems. The following sections discuss when commonly used applications will write zeros and whether those zeros can be reclaimed with Virtual Provisioning space reclamation.

VMware ESX Server

In a VMware environment, the zeroing out of virtual disks can lead to unnecessary physical storage allocation on Symmetrix thin pools, a situation that is inherently detrimental to the space-saving benefit of Symmetrix Virtual Provisioning. The ability to perform space reclamation on thin devices hosting these zeroed-out virtual disks alleviates this issue. In VMware® Virtual Infrastructure 3 and VMware vSphere™ 4.0, there are multiple situations where performing space reclamation on thin devices will be helpful.

VMware Virtual Infrastructure 3

In VMware Virtual Infrastructure 3, the usefulness of space reclamation comes in to play for three main scenarios: virtual machine cloning, deploying a virtual machine from a template, and importing a virtual machine using VMware Converter. All three operations result in virtual disks that use the "eagerzeroedthick" allocation mechanism. "Eagerzeroedthick" allocates all of the space and initializes all of the unused blocks with zeros. Consequently, the use of these three different types of virtual machine deployment methods causes the new virtual machines to consume far more space than is used by actual data. Since the extra space is taken up by zeros, it can be reclaimed to free space up on the thin pool.

The fourth method of deploying virtual machines is through the use of the New Virtual Machine wizard. This method uses the "zeroedthick" allocation mechanism for virtual disks. With the "zeroedthick" allocation scheme, the storage required for the virtual disks is reserved in the datastore but the VMware kernel does not initialize all the blocks. The blocks are initialized by the guest operating system as write activities to previously uninitialized blocks are performed. Therefore, the space consumed on the thin pool is only actual data, not zeros. For virtual machines created in this method, space reclamation is not necessary.

VMware vSphere 4

In VMware vSphere 4, the ability to choose virtual disk allocation mechanisms allows administrators to deploy virtual machines using the thin, zeroedthick, or eagerzeroedthick mechanisms. Since newly provisioned virtual machines are not forced into the eagerzeroedthick format like in Virtual Infrastructure 3, virtual machines created in vSphere are typically Symmetrix Virtual Provisioning "friendly."

As shown in Figure 6**,** if a new virtual machine is created with the box **Support clustering features such as Fault Tolerance** selected, the virtual machine will be created using the "eagerzeroedthick" allocation mechanism and if placed on a thin device, will result in a complete allocation due to the zeroing of the virtual disk. Space reclamation can be used in this instance to alleviate the consumed space on the thin pool, but if the virtual machine is targeted to be used with VMware Fault Tolerance, do not reclaim those zeros. VMware fully allocates virtual disks of Fault Tolerant machines purposefully and the zeros are not meant to be removed and could lead to issues with the Fault Tolerant virtual machine configuration.

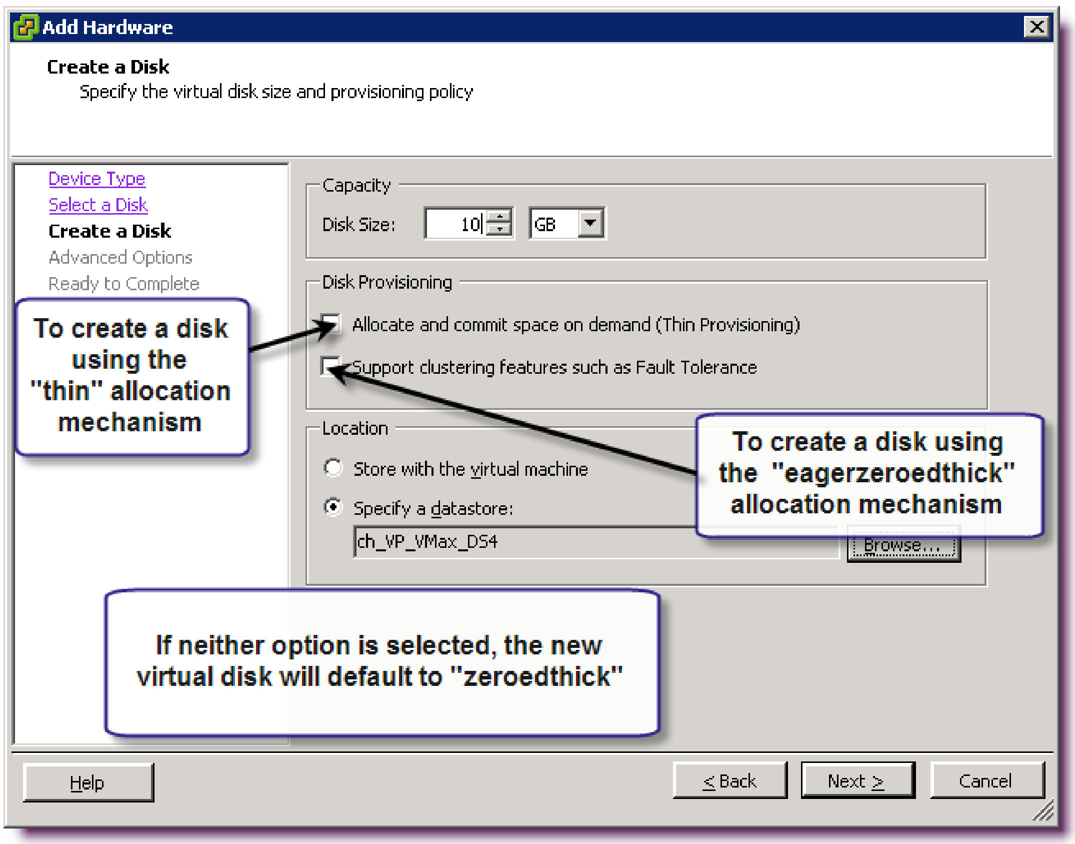


Figure 6. Virtual disk allocation mechanism options in vSphere 4.0

Migrating from Virtual Infrastructure 3.x to vSphere 4.0

The migration of a VMware environment from Virtual Infrastructure 3 to vSphere 4 follows all of the same recommendations and practices set forth in the preceding paragraphs. Templates, virtual machines deployed from templates, and virtual machines produced from the cloning process in Virtual Infrastructure 3.x all result in fully allocated virtual disks. In vSphere environments the zeroed blocks can be reclaimed by using the integrated Storage VMotion® feature. But this requires two VMotion processes, one to convert the virtual machine to "thin" to reclaim the zeros, and a second to VMotion it to "zeroedthick" to follow best practices when storing VMs on Symmetrix thin devices. In the cases of large virtual machine farms it is laborious and inefficient to manually convert each virtual machine from eagerzeroedthick to "thin" to the optimal "zeroedthick" format for Symmetrix Virtual Provisioning. Instead, space reclamation can be used on each Symmetrix thin device to remove the zeroed out allocations while simultaneously offloading the host CPU cycles and memory requirements consumed by Storage VMotion.

Considerations for space reclamation in VMware environments

For both Virtual Infrastructure 3.x and vSphere 4.0, there are important considerations that should be evaluated before performing reclaims.

If a thin device stores multiple virtual machines, a zero reclaim function cannot be performed on just one of the virtual disks. It is currently not possible to determine the specific blocks each virtual disk consumes on the thin device and therefore it is not possible to limit a reclaim to a single virtual disk out of many on a given VMware file system[1]. Consequently, a reclaim should only be performed on a thin device if all virtual machines hosted by the VMware file system volume on that thin device can and should be reclaimed. In particular, in vSphere environments, virtual machines that are set up in a Fault Tolerant configuration should not share Symmetrix thin devices (or the VMware file systems that are hosted by them) with other virtual machines. This will reduce the risk of reclaiming the zeros from a Fault Tolerant virtual machine.

Additionally, zeroing out by guest operating systems must be taken into account before performing a zero reclaim. If the virtual machines' operating system or its hosted applications zero out files for a particular reason, care must be taken to consider any implications of removing those zeros before reclaiming them.

The use of EMC Storage Viewer (available for Virtual Infrastructure 3.5 and vSphere 4.0) is highly recommended in order to easily map virtual machines and their corresponding VMFS volumes to the correct underlying Symmetrix thin device. Double-checking storage mapping information with the EMC Storage Viewer will eliminate the possibility of performing a zero reclamation on the incorrect thin device.

Oracle

REDO log files

Oracle REDO log are pre-allocated files that store all the database change records before they are applied to datafiles. Oracle writes redo records to REDO log files in a circular fashion and therefore there is no reason for space reclamation on these devices.

Datafiles

When an Oracle datafile is first created, Oracle initializes the space in the file. Each initialized datafile contains a series of empty data blocks populated with a block header and other metadata. Due to the nature of file initialization in Oracle, there are no contiguous ranges of zero blocks with potential for space reclamation when targeting storage allocated by an initialized Oracle datafile.

In order to get improved capacity utilization benefits from a Virtual Provisioning infrastructure, a strategy for sizing datafiles, pools, and thin devices should be developed in accordance with EMC best practices.[2]

ASM considerations

When an ASM diskgroup is created, Oracle writes just a small amount of metadata but does not initialize with zeros or write to the whole newly created diskgroup space. Thus ASM is thin provisioning friendly and allows better storage utilization when used with thin LUNs (as described in EMC Virtual Provisioning best practices for Oracle and ASM). Therefore, there is no potential of zero space reclamation for devices provisioned to Oracle ASM outside of other scenarios described later in this white paper.

[1] EMC is actively working with VMware on the vStorage initiative that may allow in the future for zero reclaim to be performed on a single virtual disk.

[2] *Implementing Virtual Provisioning on EMC Symmetrix VMAX with Oracle 10g and Oracle 11g* white paper

Reclaiming space from deleted Oracle objects

Whenever an Oracle object is dropped, the underlying storage is not zeroed out and cannot be reclaimed directly. ASM will, however, attempt to reuse space when new extents are added. This behavior reduces the need to reclaim deleted space in ASM. See also the "File system considerations" section for reclaiming space on deleted Oracle objects on file systems.

Summary of Oracle considerations

Virtual Provisioning space reclamation improves storage utilization by releasing space consumed by contiguous zero blocks on the thin devices. Oracle file initialization and metadata limits such reclamation potential for Oracle database objects. However, Oracle applications can still improve storage capacity utilization by following the specific EMC best practices for Oracle in Virtual Provisioning environment, incorporate external tools to zero out previously used devices before they are given back to Oracle, or benefit from the TimeFinder/Clone thick-to-thin feature as described in the "Migration and replication" section of this white paper.

Microsoft SQL Server

Transaction log files

Active log files are formatted when a Microsoft SQL Server database is created. Every single page of the log files is written to at the time of creation so that the log files become fully provisioned when they are initialized. Because the log file will mostly contain contiguous zeros when first written, its storage has the potential to be reclaimed by Virtual Provisioning reclamation. Over time, as transactions are written to the log files and later truncated based on log backup or checkpoint operations, the log file will no longer contain all zero data. With this behavior in mind, it should not be expected that a log file allocated to a proper size and used for a given period of time will contain contiguous zeros sufficient enough to be reclaimed.

Should a transaction log file be sufficiently oversized to where areas of the log were never used by SQL Server, there is the potential that these areas on disk could be reclaimed. While it is possible to reclaim unused portions of the SQL log file it is generally not recommended to reclaim zero space dedicated to a log device. Pre-written space on thin devices provides optimum write performance, important for the SQL log file. Additionally, pre-written space will guarantee thin pool allocations are dedicated and available for the SQL log file.

Database files

Prior to SQL Server 2005, SQL Server would fully initialize all datafile and transaction log file components, writing to every page in all database files and transaction log files. The process of writing to every page during initialization writes contiguous zeros to areas of the database not otherwise used by file metadata pages. If these zero pages remain untouched by SQL Server there is the potential that the areas on disk could be reclaimed.

As table and index information is written to a fully initialized datafile, areas of the database become allocated and used by non-zero user data. Assuming a database file was overallocated to where its initialized size is larger than the amount of user data in tables and indexes it may be possible to reclaim the remaining zero space, assuming it was never touched by SQL Server outside of the initialization process during creation or growth operations. Figure 7 provides a logical depiction of a file system and database file in this state.

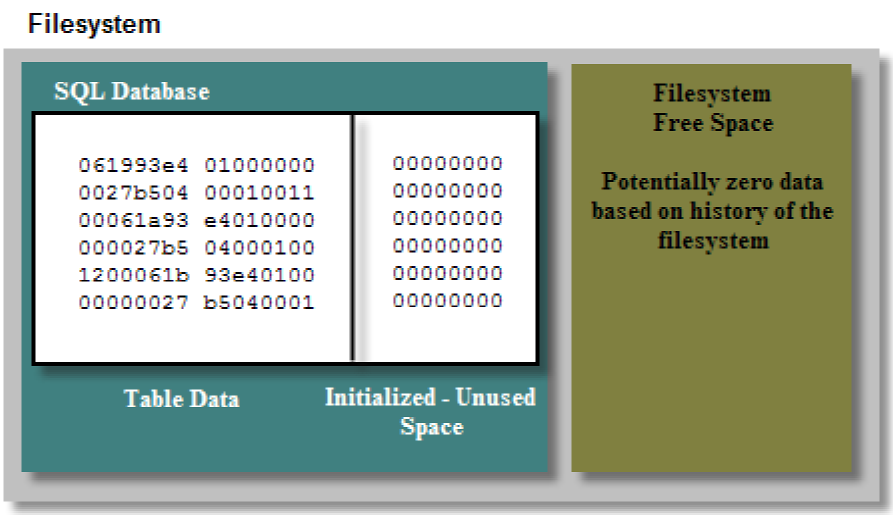


Figure 7. File system and SQL Server database file with initialized and unused space

Microsoft SQL Server 2005 introduced new functionality that altered the behavior during the database creation phase. As of SQL Server 2005, the database creation phase was no longer required to initialize all datafiles if Instant File Initialization could be utilized by the SQL Server environment.

Transaction log files will continue to be fully allocated with zeros even when using Instant File Initialization.

Instant File Initialization will be used automatically by SQL Server provided the service account, under which the SQL Server service is running, has the "Perform volume maintenance tasks" user rights assignment under the local security policy. By default, only users marked as Administrators have this permission.

Areas of a disk under which a sparse file is defined, as created by Instant File Initialization, are not zeroed. Whatever data existed on these areas of the disk, be it zeros or perhaps non-zeros based on the history of the file system and LUN, will remain. In the context of migrations, should a LUN-based copy be replicated from a source volume where a database was created with Instant File Initialization, to a target thin device, the ability to reclaim zero written data will depend on the history of the LUN and file system. Figure 8 depicts this scenario.

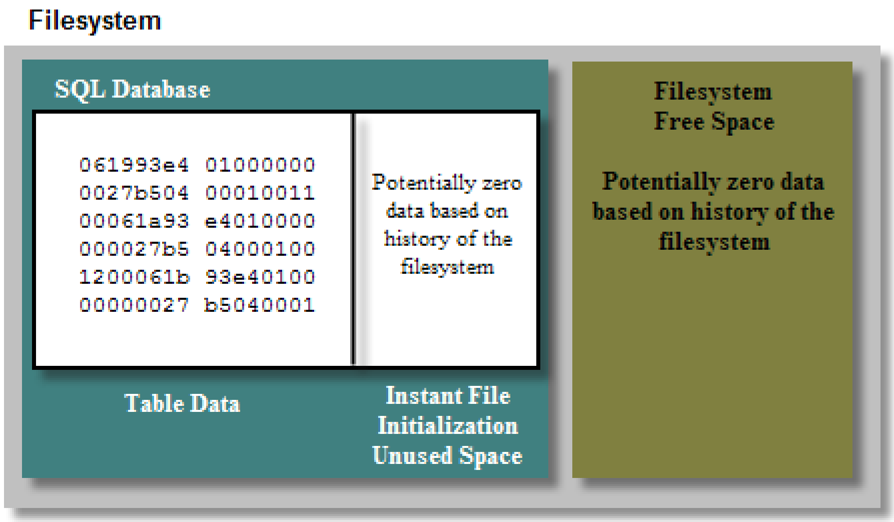


Figure 8. File system and SQL Server database created with Instant File Initialization

Should user data have ever resided in areas of the database file now considered free based on row or table delete operations, those areas will be non-zero and will not have the potential to be reclaimed. With SQL Server 2008 there is a stored procedure called sp_clean_db_free_space, which is intended to clean areas of a database file that are considered free. This procedure is also included with SQL Server 2005 SP3 but is undocumented. The sp_clean_db_free_space command will zero not only previously used areas of the database file, but also areas of the file that have never been used. This operation will subsequently cause writes and thin pool allocations on areas of the database file that have not been allocated where Instant File Initialization was used. As a part of the cleaning process each page touched by the procedure will leave behind database header information. While the majority of the page is zeroed, the page header left behind makes it impossible to reclaim the space cleaned by this utility. Because of this behavior, it is not recommended to use sp_clean_db_free_space as a means to reclaim space within a SQL database file. Additional information regarding Microsoft SQL Server and Symmetrix Virtual Provisioning can be found in the white paper *Implementing Virtual Provisioning on EMC Symmetrix with Microsoft SQL Server – Applied Technology* available on Powerlink®.

Microsoft Exchange Server

Microsoft Exchange log files

By default, Microsoft Exchange allocates log space in the form of 5 MB files or 1 MB files, depending on the Exchange version. As the log files are filled, new log files are created. Over time, log files that have been committed to the database files will be deleted (truncated). This is normally performed following a successful backup operation. Log files are also deleted automatically should circular logging be enabled. As log files are deleted, space is freed within the NTFS volume. However, the space is not freed within the thin data pool. The current implementation of Virtual Provisioning does not have a mechanism to reclaim storage from the thin pool when an object is deleted by the application using the thin device. For the thin pool space to be utilized efficiently, NTFS needs to reuse areas of the volume that had been previously written to and subsequently freed via the truncation process.

NTFS reuse rates may vary depending on operating system and service pack levels. Testing with Windows 2003 SP2 and Windows 2008 have shown more efficient use of newly freed log volume space on NTFS volumes when compared with earlier versions of Windows.

Microsoft Exchange database files

By default, an Exchange database file will start between 2 MB and 4 MB in size and will incrementally grow by 1 MB or 2 MB (depending on version) as additional space is needed. This auto-extend behavior is efficient from a thin pool perspective as only space immediately needed by the database file will be allocated. Based on this behavior, contiguous free or zero value space will generally only exist in the file system that holds the Exchange database file. How much of the file system and subsequently the LUN supporting the volume will contain zeros is dependent on the size of the database files and whether any other file system objects existed previously and were subsequently deleted.

One instance in which an Exchange database file may contain zero data is if page zeroing functionality is enabled. If page zeroing is enabled, once a deleted database object (mailbox messages, user mailboxes, and so on) has surpassed the date specified by the deleted item retention policy, the page on which it resided will be zeroed. Deleting data in this fashion ensures it is not recoverable, which may be a requirement in environments with sensitive data. Depending on the version of Exchange, page zeroing can be enabled and subsequently occur during a streaming backup operation, or as a part of the online maintenance window.

Testing with Exchange 2007, utilizing database page zeroing as a part of online maintenance, has shown that the pages cleaned by this process are not subsequently recoverable by Virtual Provisioning space reclamation. It should not be expected to be able to reclaim any data within an Exchange database file following page zeroing operations, even when large amounts of data are deleted.

Microsoft Hyper-V

Two primary representations of storage devices may be configured for a virtual machine in Microsoft Hyper-V environments. Either a storage device will be provisioned as a Virtual Hard Disk (VHD) and reside within an NTFS file system on the Hyper-V parent, or the device will be connected as a physical hard disk (a pass-thru storage device). The efficiency of pass-thru storage with regards to Virtual Provisioning as well as space reclamation will be fully dependent on the use of the hard disk by the virtual machine.

The efficiency of virtual hard disks from a Virtual Provisioning perspective, along with their use by a virtual machine, will be dependent on the kind of VHD format used. Three different types of VHD disks are available when configuring new or additional storage devices, represented in Figure 9. The choice between a “Fixed size” and “Dynamically expanding” format will generally be based on the storage utilization requirements, in as far as there is a difference in how storage is allocated for these two types. For this reason, the two selections will subsequently affect storage provisioning functionality such as that provided by Symmetrix Virtual Provisioning. A fixed sized VHD device will be fully written to at creation time. The writes that occur to a fixed size VHD will be contiguous zeros and can be reclaimed using Virtual Provisioning space reclamation dependent on the use of the VHD by the virtual machine.

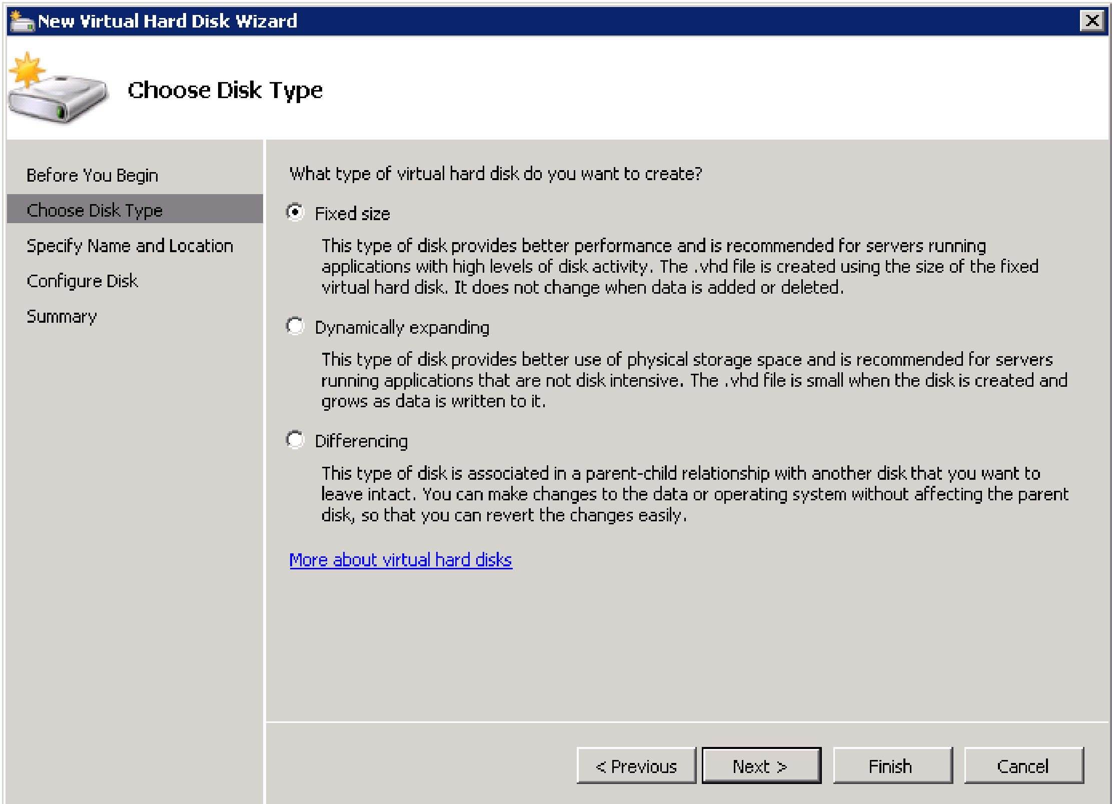


Figure 9. Hyper-V virtual hard disk options

Dynamically expanding VHD devices will not pre-allocate all storage defined for them; however these devices may suffer a slight degradation in performance as a result of the need to allocate actual storage when the operating system or applications within the virtual machine require more allocations. These storage allocations are simply those to acquire additional storage from the parent partition for the VHD device.

Beyond the format of the VHD files, the amount of space reclamation possible in a given Hyper-V environment will also depend on the NTFS volume that supports these files and its historical use. Please see the “File system considerations” section for additional details.

Migration and replication

File system considerations

The efficiency of space reclamation following a migration process will be highly dependent on the nature of the data written to the original source file system. Most file systems are Virtual Provisioning friendly in that they will not write non-zero data to the entire contents of a volume during creation. Most allocations that occur will be due to user and application data being written to the partition or file system. Depending on the use of the file system, objects may be added and deleted on a regular basis. Generally speaking, when a file is deleted from most file systems it is removed from being referenced or otherwise available for access; however the data on disk is still non-zero and represents the original contents of the file. Over time, as files are repeatedly added and deleted, it may be possible for the entire file system to contain non-zero data, even though from the operating system perspective there could be a large amount of free space available.

When devices containing file systems that exhibit a large number of add and delete operations are migrated, depending on the migration tool used, it is possible for the non-zero data representative of objects no longer referenced by the file system to be copied as a part of the migration process. The act of copying deleted but non-zero data will lead to lower space savings on thin devices targeted for migration and will also prevent Virtual Provisioning reclamation from freeing this space.

The ability of space reclamation to work efficiently following a migration from a thick to a thin device depends on the amount of contiguous zero blocks that may have been copied from the thick LUN. To maximize the zero blocks, external tools such as sdelete for Windows can be used on the thick LUN prior to a migration. Tools like sdelete will ensure a given area of a file system that was previously used and subsequently freed will contain zero data that can be reclaimed.

Open Migrator/LM

EMC Open Migrator/LM (OM) enables online data migration of Microsoft Windows, UNIX, or Linux volumes between any source and EMC storage. Open Migrator/LM host-based software helps to enhance data center efficiency by automating and simplifying data migration. Regardless of the purpose for the migration, be it consolidating servers, upgrading storage, or tuning performance, volumes remain online and fully available to critical applications during migration.

At the time of publication, versions of Open Migrator/LM supported on UNIX and Linux platforms perform a full volume copy from a source device to a target device. Should the target device be virtually provisioned, it should be expected that the target thin device will become fully allocated within the thin pool. The ability to reclaim the allocations done during the migration will be dependent on the amount of contiguous zero blocks that originally existed on the source device. The “File system considerations” section has more details and considerations.

Open Migrator/LM for Microsoft Windows can also perform full volume copies but has an additional feature referred to as “Sparse” copy. The sparse copy functionality of OM for Microsoft Windows allows for the replication of only known file system objects, specifically only the clusters of data marked as in use by NTFS. Previously deleted objects and other areas of the file system considered empty will not be migrated. This makes sparse migrations a compelling option when moving data from a thick environment to a thin environment. The only contiguous areas of zero data that will be replicated with OM using sparse copy will be those that exist within the files and areas of the file system marked as used by NTFS. To use our SQL Server example in Figure 7, this would mean the entire SQL datafile, including the initialized and zeroed regions of the file, would be migrated. The area of free space in the file system would not be copied. In this case, the zeroed region of the datafile could be reclaimed following the migration using Virtual Provisioning space reclamation

EMC Open Replicator

EMC Open Replicator for Symmetrix (ORS) enables remote point-in-time copies to be used for data mobility, remote vaulting, and migration between EMC Symmetrix and qualified storage arrays with full or incremental copy capabilities. Open Replicator leverages a SAN environment for copying data between Symmetrix and qualified storage arrays. Open Replicator can:

- Pull from source volumes on qualified remote arrays to a Symmetrix volume
- Push any live source Symmetrix volume to a target volume on a qualified array with incremental updates
- Perform online data migration from qualified storage to a Symmetrix with minimal disruption to host applications

Open Replicator can be used to perform remote replication between thin devices or between thin and regular devices. Managing thin device replication with Open Replicator is exactly the same as managing the replication of regular devices.

Thin devices can be used as control devices for hot and cold pull and cold push Open Replicator copy operations. If a push operation is done using a thin device as the source, zeros will be sent for any regions of the thin device that have not been allocated, or that have been allocated but to which have not been written.

Open Replicator can also be used to copy data from a regular device to a thin device. If a pull or push operation is initiated from a regular device that targets a thin device, with Enginuity 5874 or earlier, then a portion of the target thin device, equal in size to the reported size of the source volume, will become allocated.

As has been previously discussed, the ability to reclaim portions of a thin device, following a migration with ORS, will be dependent on the ranges of contiguous zeros that existed on the source device prior to migration. If the source device contained contiguous ranges of zeros, they can be reclaimed using Virtual Provisioning zero space reclamation.

With Enginuity 5875, Open Replicator pull operations have been enhanced to detect contiguous ranges of zero data. The Symmetrix VMAX front-end adapter will scan incoming tracks for all zero data. If the 12 tracks that comprise a thin provisioning extent all contain zero data, that extent will be reclaimed immediately, thus maintaining the thin nature of the target device. This enables unused space on thick devices to be reclaimed while being migrated to a thin device on a Symmetrix VMAX. The feature works for both hot and cold pull operations and is supported for all systems that are currently supported by Open Replicator. Further, performing on-the-fly zero detection mitigates the need to run zero space reclamation as a post-migration activity.

TimeFinder/Clone thick to thin replication

Previous versions of Solutions Enabler and Enginuity have allowed for thin to thin replication. In the context of TimeFinder/Clone, what has been allowed is the replication from a source thin device to a target thin device within a Symmetrix. With the latest release of Enginuity 5874, TimeFinder/Clone can replicate standard volumes to thin volumes "sparsely" to ensure only written tracks are copied, reducing capacity requirements and TCO. Thin volumes also can be replicated to standard volumes, further improving mobility into and out of thin pools.

When a clone copy is made between a regular source device and a thin target device, device extents that have never been written to by a host will not be copied to the target volume. Following the clone copy, any thin device extents that were allocated on the clone target that contain all zeros can be reclaimed and added back to the target thin device’s pool.

Symmetrix Remote Data Facility (SRDF)

SRDF is a frequently used option for migrating data between Symmetrix storage arrays. Beginning with Enginuity 5875 it is now possible to use SRDF to migrate data from standard devices in a Symmetrix array running Enginuity 5671, 5773, or 5875 to thin devices in a Symmetrix VMAX running Enginuity 5875. In the case where the source array is running 5773 or later, the SRDF adapters (RAs) on the target Symmetrix VMAX will perform zero data detection on the fly on a per-track basis. Zero data detection is accomplished by looking at each track of data as it arrives to determine if it meets one of these two conditions:

- If the never written by host (NWBH) indicator is set for the track

  Beginning with Symmetrix VMAX and Enginuity 5874 the special NWBH indicator is maintained in the metadata stored in cache for every track on every drive in the array. When this indicator is set for a track, it signifies that a host has never written any data to the track. If Enginuity receives a read request for a track with the NWBH indicator set, it can skip performing a read I/O to the disk and simply return all zeros to the host. If the track is part of a device participating in a local or remote replication session, the NWBH indicator will be set on the corresponding track of the target device.

- If the track contains all zero data

  If the RA determines that an incoming track contains all zero data, by scanning that track's worth of data, it will indicate this in the track's metadata. If the 12 tracks that comprise a thin provisioning extent all contain zero data, that extent will be reclaimed immediately, thus maintaining the thin nature of the target device.

Note: Zero data detection for SRDF is supported only in adaptive copy mode.

PowerPath Migration Enabler

EMC PowerPath Migration Enabler is a migration tool that enables nondisruptive or minimally disruptive data migration between storage systems or between logical units within a single storage system. Migration Enabler resides on the host and allows applications continued data access throughout the migration process. When pseudo device names are used, the migration can be nondisruptive. When native device names are used, a migration is minimally disruptive because applications must be reconfigured to use the new target device name that contains the migrated data. Migration Enabler works independently of PowerPath multipathing and can be used whether or not PowerPath is used for multipathing. If PowerPath is not already installed on the host, on some platforms a disruption may be required to install it. PowerPath Migration Enabler integrates with other technologies to minimize or eliminate application downtime. Supported technology types are:

- EMC Open Replicator for EMC Symmetrix. The online pull method is supported, in which data is copied through the SAN from a remote logical unit to a Symmetrix storage system.
- EMC Invista® in which a Virtual Volume encapsulates a Symmetrix or EMC CLARiiON source logical unit.
- EMC TimeFinder/Clone (Symclone), which migrates data between devices within a single Symmetrix system.
- Host-based copy (Host Copy), which works in conjunction with the operating system to migrate data from the source logical unit to the target logical unit.

The Host Copy technology type supports migrating virtually provisioned, or *thin*, devices. Host Copy automatically detects virtually provisioned CLARiiON and Symmetrix devices. When a thin target is detected PPME will scan the source disk for zeros and will not replicate these areas of the disk as a part of the migration process.

To again use our SQL example from Figure 7, PPME would only copy the non-zero regions of the database and may potentially replicate the free space region of the LUN depending on whether it contains non-zero data based on previous use. Due to this functionality, it should not be expected to be able to reclaim any space within a thin pool with Virtual Provisioning reclamation, following a thin-friendly migration using PPME Host Copy.

With technology types other than Host Copy, Migration Enabler support for thin devices is determined by the underlying technology's support. Please see the previous sections to understand TimeFinder/Clone and Open Replicator in the context of Virtual Provisioning and space reclamation.

Veritas Storage Foundation SmartMove

Veritas Storage Foundation from Symantec offers a "SmartMove" capability where, during migrations from thick to thinly provisioned storage, only used space from the file system perspective is migrated. Previously deleted objects and other areas of the file system considered empty will not be moved. SmartMove therefore helps to enable efficient, thick-to-thin migrations with the reclamation of unused and otherwise free space. More information regarding the use of Veritas Storage Foundation SmartMove with EMC Symmetrix can be found in the white paper *EMC Symmetrix VMAX with Veritas Storage Foundation* available at http://powerlink.emc.com and http://www.symantec.com.

Space reclamation from previously used devices

In some cases devices are no longer needed and can be re-used for other applications. Such devices would continue to retain stale non-zero data from prior uses and will not allow reclamation of unused storage space. Better capacity utilization and cost savings will be realized if the devices are zeroed out first using operating system tools prior to re-provisioning. As a caution, prior to using such tools administrators must make sure that the devices don't contain any active application data.

On UNIX and Linux systems, a command like "dd" can be used to zero out the disk blocks. If an entire partition is re-provisioned then "dd" can be used to zero out the whole partition. Once allocated space on the device has zeros in contiguous blocks, space reclamation can be used to reclaim the zeroed (and un-used) storage in the thin pool. For example, dd if=/dev/zero of=/dev/emcpowerdf1 bs=1024 count=1024000, will zero out 1 GB space on the first partition on device /dev/emcpowerdf.

On Windows 2003 and 2008 systems the diskpart "clean all" command can be issued to zero out all areas of a specified disk. With Windows 2008 only, a non-quick or full format will also write zeros to the targeted volume.

Additionally, as detailed in the "File system considerations" section, external tools such as sdelete for Windows can be used to zero out contiguous ranges of a device.

The use case for zeroing out a standard device is most interesting in the context of future replication from thick to thin volumes, like with Open Replicator or TimeFinder/Clone. The target thin devices in these scenarios will be able reclaim the now zero space as replicated from the source.

If the source is a thin device and the entire device is re-provisioned, then the most efficient method for reclamation would be to unbind and re-bind to the thin pool to automatically reclaim space allocated by the device. For example, if the thin device 0297 on Symmetrix 191 is added to a thin pool Oracle1 and masked to a host then the following commands can be used to reclaim space.

(1) Change device state to "not_ready" to prevent host access:
    symdev –sid 191 –RANGE 0297:0297 not_ready:
(2) Unbind the device from the thin pool:
    symconfigure –sid 191 –cmd "unbind tdev 0297 from pool Oracle1 ;" commit
(3) Rebind the device to the thin pool:
    symconfigure –sid 191 –cmd "bind tdev 0297 to pool Oracle1 ;" commit

Online reclamation of previously used space

Enginuity 5875 includes support for new T10 industry-standard SCSI commands including UNMAP and WRITE_SAME unmap. These SCSI commands allow for operating systems and applications to communicate to the Symmetrix VMAX a range of logical block addresses (LBAs) that are no longer needed. If the LBA range covers an entire Virtual Provisioning extent (12 tracks or 768 KB), then the extent can be reclaimed.

Applications that utilize the UNMAP or WRITE_SAME unmap capabilities of Symmetrix VMAX help to enable the reclamation of previously used, but now deleted, areas of a thin device, without the need to zero out areas of a file system or otherwise unbind the thin device.

For Windows environments, the *storreclaim* utility has been written by EMC to take advantage of the WRITE_SAME unmap command. The *storreclaim* command is intended to be run against NTFS file systems that contain previously written but now deleted files. The utility will create sparse files within the free space of the file system, map out the extents of the newly created files, and subsequently issue a WRITE_SAME unmap command for those ranges of extents. Symmetrix VMAX will then reclaim the ranges of previously allocated and written space within the thin pool. The *storreclaim* command is nondisruptive and can be run while the file system being affected is mounted and in use.

The *storreclaim* command will initially be released as part of the Global Services ToolKit. EMC employees can find more details regarding the *storreclaim* command on the internal EMC One website. Customers should contact an EMC representative for more information.

Veritas Storage Foundation also offers a reclamation facility that utilizes the WRITE_SAME unmap specification inherent in the Symmetrix VMAX with Enginuity 5875. The reclamation capabilities of Storage Foundation span a wide range of operating systems and file systems. For more information regarding the supported platforms for thin reclamation with Veritas Storage Foundation, please see the appropriate hardware compatibility lists on http://www.symantec.com.

Conclusion

Solutions Enabler 7.1 and Enginuity 5874, with the addition of space reclamation, enhanced the benefits of Virtual Provisioning in Symmetrix VMAX environments. The ability to reclaim unnecessary ranges of zero data from a thin pool enhanced storage efficiency and potential cost savings. Enginuity 5875, with the addition of on-the-fly zero reclamation with Open Replicator and SRDF, as well as the implementation of the WRITE_SAME unmap T10 standard, enables even greater efficiencies for virtually provisioned environments. The ability to utilize native application functionality as well as common migration tools, with less worry of wasted storage space, further enhances administrative efficiencies, including the time it takes to provision or migrate storage resources. Increased efficiency and savings, from both a storage and administrative perspective, will lead to an overall higher return on investment for critical data center resources.

References

White papers

- *New Features in EMC Enginuity 5874 for Symmetrix Open Systems Environments*
- *New Features in EMC Enginuity 5875 for Symmetrix Open Systems Environments*
- *Implementing Virtual Provisioning on EMC Symmetrix with Microsoft Exchange Server – Applied Technology*
- *Implementing Virtual Provisioning on EMC Symmetrix with Microsoft SQL Server – Applied Technology*
- *EMC Symmetrix with Microsoft Hyper-V Virtualization – Applied Technology*

Technical notes

- *Best Practices for Fast, Simple Capacity Allocation with EMC Symmetrix Virtual Provisioning*

Feature sheet

- *Symmetrix Virtual Provisioning Feature Specification*

Product documentation

- *EMC Solutions Enabler Symmetrix Array Management CLI Product Guide*
- *EMC Solutions Enabler Symmetrix Array Controls CLI Product Guide*
- *EMC Solutions Enabler Symmetrix CLI Command Reference*