EMC²
where information lives®

Reclaiming Storage in a SAN Environment: the Good, the Bad, and the Ugly

Brian Dehn

EMC Proven Professional Knowledge Sharing 2009

EMC²
PROVEN
PROFESSIONAL

Brian Dehn
Champion – Resource Management Software
Technology Business Consultant
dehn_brian@emc.com

Table of Contents

Disclaimer: The views, processes, or methodologies published in this compilation are those of the authors. They do not necessarily reflect EMC Corporation's views, processes, or methodologies.

Introduction

"My MBO is based on saving $5M worth of storage."
"We are cutting the budget for storage so you need to reuse capacity."
"No additional storage will be purchased until utilization is increased."

Reclaiming storage capacity can reduce IT capital expenditures, increase storage utilization, contribute to "green computing" initiatives, and make a major step toward addressing issues like those listed above. Proper planning for and execution of a storage reclamation effort is key to realizing maximum benefit and avoiding problems. The goal of this article is to help you achieve storage reclamation objectives ("the good") while reducing time and cost spent on less beneficial efforts ("the bad") and avoiding significant problems ("the ugly").

Before embarking on such an effort, you must determine what storage is available for reclamation, whether the benefit offsets the cost, and which tools will facilitate the effort.

Understanding Reclaimable Storage

Making better use of existing storage assets is a daunting task. These are likely among your concerns:

- I don't know exactly what storage we have today.
- How do we figure out what storage is really being used and what is available?
- I'm sure we have capacity that's "lost" or "orphaned" that we can reclaim but I don't know how to find it.
- How do I begin?

IT departments, specifically storage administration teams, usually have a clear idea of how much storage they have available or "free" for allocation. More difficult to find and identify is capacity that may appear to be used but is not. This article focuses on the latter of these two points. The definition of reclaimable storage is capacity that appears to be in use by one or more hosts, but is not used at all. This capacity is difficult to find and sometimes referred to as "lost" or "orphaned."

Storage may be lost or orphaned for many reasons, usually as the result of inadequate or broken processes. For example, one group could decommission an application and/or host without notifying the storage team that the storage is no longer used. Alternatively, the storage team may be notified that storage is no longer being used; however, they do not perform some or all of the tasks required to return that storage to the free pool.

Another scenario that may lead to orphaned storage involves the methods used to reserve storage for specific or future use. A storage administrator may decide to flag devices as reserved by configuring them in a certain manner, such as mapping them to front-end ports on an array and/or masking devices to a specific invalid host bus adapter (HBA) port worldwide name (WWN). Many storage resource management (SRM) reporting tools will discover these configuration settings and automatically classify the devices as allocated, when they are simply reserved. The devices can be lost if the storage administration team does not keep accurate records of what they reserved using this type of process. The devices may also be lost if only one person knows the reservation methodology and leaves the company.

To understand how to find and reclaim orphaned storage, you must first understand all the states of configuration in which storage may exist.

The Storage Configuration Hierarchy

Understanding the storage configuration hierarchy is the first step toward identifying reclaimable capacity. Storage exists in various states of configuration, and it may be "lost" or "orphaned" in one or more of these states. The states of configuration span from unconfigured raw disk in an array, to usable capacity containing data generated by applications, and various states in between, as represented in *Figure 1*.

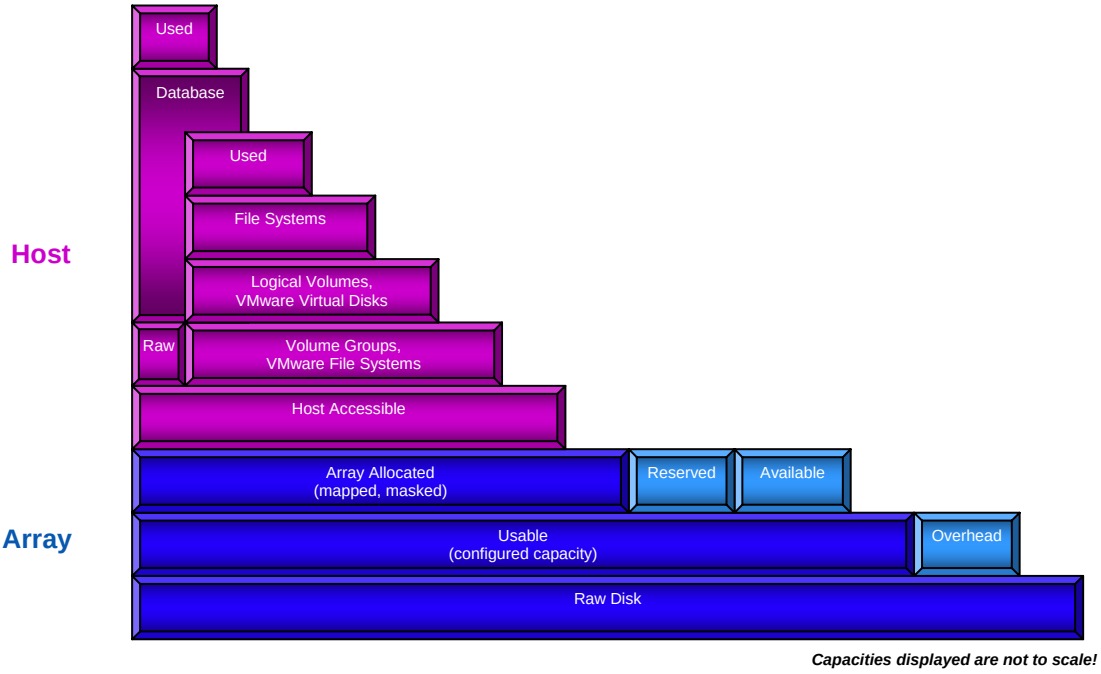


Figure 1 - Storage Configuration Hierarchy

The first layer of the hierarchy, Raw Disk, represents the capacity of all physical disks in an array in their unformatted state. Capacity in this state is not usable because it has not been configured into logical devices.

The second layer, Usable, represents capacity that has been configured into logical devices or formatted into RAID groups, depending on the type of storage array. At this layer, the majority of the storage is usable from a host perspective while a percentage is consumed by protection-level overhead (such as mirrors or RAID parity) for devices and RAID groups. In addition, storage administrators may leave small slices of disks as raw capacity because the small slices are not large enough to configure into standard-sized devices, or because managing the small slices is not cost effective.

After configuring raw capacity into a usable format, you can allocate it for host usage. Allocated capacity is the amount of configured storage that has been reserved for use, either for hosts connected to the array, or by the array itself. Array allocation of usable capacity generally involves mapping and masking (refer to definitions in the next paragraphs) after which a device is considered allocated from the array perspective, represented by the third layer in the diagram.

At this layer in the storage configuration hierarchy, storage administrators may also logically reserve devices for specific projects (e.g. new applications, future projects, technology refreshes, etc.) Administrators use various tools to identify usable devices as available for allocation, sometimes referred to as “free pools” of capacity, including SRM reporting, array configuration software, and/or spreadsheets.

As a definition, mapping is required when access controls are enabled for front-end ports on arrays. It consists of associating a device to a front-end port on the array where that device is configured, thereby providing access to that device via that array port. Mapping is not required when access controls are not enabled for array front-end ports.

Also, device masking is the capability that allows a specific device (also referred to as a LUN) to be exclusively available to some hosts and unavailable to other hosts. It is implemented at the host us adapter (HBA) port level.

Capacity is considered “Host Accessible” when hosts can directly use devices for I/O operations. For devices to be accessible to a host they must be array allocated (i.e. mapped and masked) and the storage array front-end ports to which the devices are mapped must be zoned to the HBA port WWNs. When devices first become accessible to a host, the host sees them as raw devices. The Host Accessible layer in the hierarchy represents this capacity.

Once array devices are accessible to a host, server administrators usually build one or more data structures on those raw devices. One typical data structure is a database, such as Oracle or Sybase, built directly on the raw devices. Another typical data structure is a volume group, defined and configured using volume management software. A volume group is a collection of raw devices, upon which logical volumes can be created. Logical volumes typically are not created with a one-to-one correspondence to the raw devices in the groups but rather span across devices within a volume group.

A VMware File System (VMFS) is a third type of data structure that may be built on host-accessible devices. It is a file system optimized to store ESX Server virtual machines. A VMFS is configured on one or more VMFS volumes, corresponding to array allocated devices mounted to an ESX Server, upon which VMware Virtual Disks are created.

A VMware Virtual Disk (VMDK) is a file in a VMFS and always appears to a virtual machine as a mounted SCSI device. The virtual disk hides the physical storage layer from the virtual machine's operating system.

At the next layer in the hierarchy, file systems are typically created on host logical volumes and are used for various purposes including storage of files, application data, and databases. Physical hosts and virtual machines both require and use file systems.

Storage can exist in one or more of the states of configuration described above, from unconfigured capacity in arrays, to file systems and databases containing data generated by users and applications. Familiarity with the storage configuration hierarchy provides a basis for understanding the types of capacity potentially available for reclamation at the various layers of configuration.

Types of Reclaimable Capacity

You can reclaim storage at many layers in the storage configuration hierarchy, but you must first understand the types of capacity at each layer as identified in *Figure 2*.

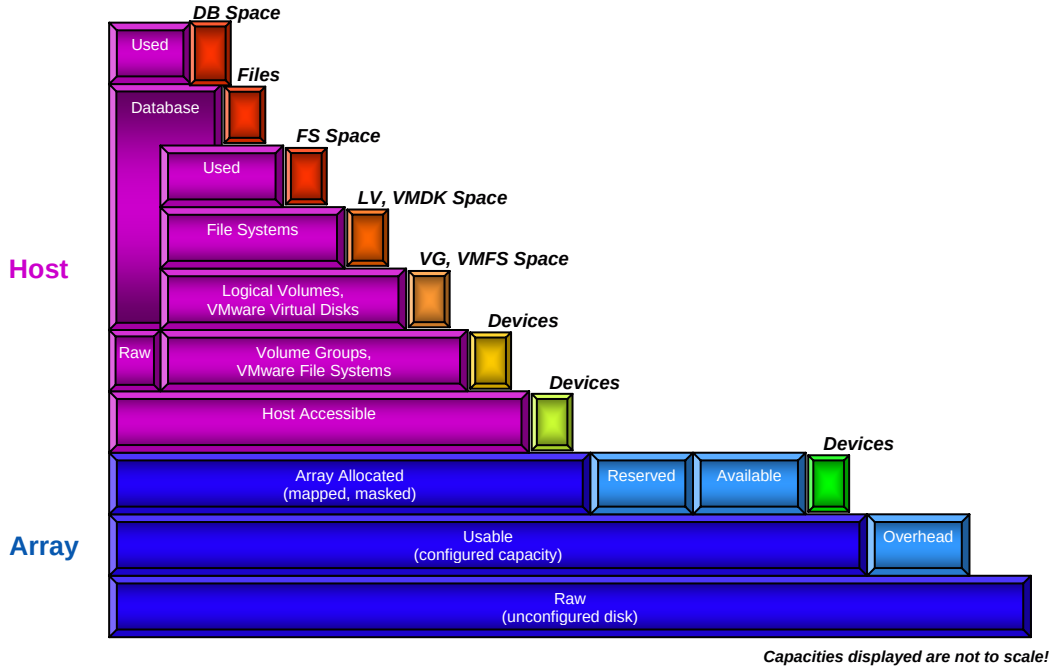


Figure 2 - Types of Reclaimable Capacity

Capacity at the first two layers in the hierarchy – raw, unconfigured disk and usable, configured capacity – do not need to be reclaimed. Capacity in a raw state has not yet been configured into a usable format; therefore, it cannot be used and does not need to be reclaimed. If enough raw capacity exists, storage administrators may simply configure that capacity into a usable state and allocate it as needed. Capacity in a configured state, identified by the second layer, does not need to be reclaimed either since it has not been allocated.

Reclaimable capacity at the next three layers exists in the form of usable devices. Any efforts to reclaim storage at those layers will result in the freeing of usable devices that can then be reallocated as needed.

Beginning at the layer of Logical Volumes and VMware Virtual Disks, efforts to reclaim capacity will result in space being freed within the data structure defined at the next lower layer. For example, removing file systems will result in space becoming available in the underlying logical volumes. Removing logical volumes will result in space being freed in the underlying volume groups. Removing VMware Virtual Disks will result in available space in the underlying VMware File System.

Understanding the storage configuration hierarchy and the types of capacity that may be reclaimed at each layer provides the foundation for the next step in the reclamation process. Your knowledge of the hierarchy will allow you to determine where you want to focus your reclamation efforts based on cost/benefit analysis.

Identifying Reclamation Targets

The level of effort (LOE) required for reclaiming storage, and the return on investment (ROI), varies at each layer in the storage configuration hierarchy. In general, reclamation efforts focused on higher layers of the hierarchy require more time and resources and yield less benefit when compared to efforts focused on lower layers (refer to *Figure 3*).

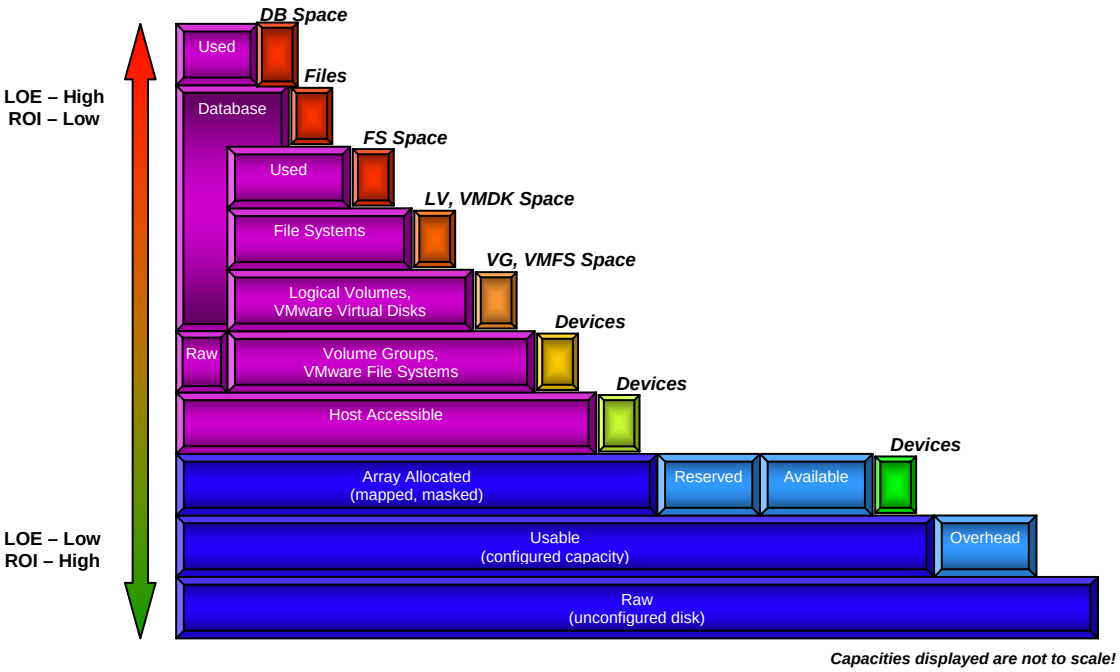


Figure 3 - Level of Effort & Return on Investment

Since reclamation efforts at the lower layers of the hierarchy provide a higher ROI, this section of the article will examine the lowest layers first and progress upward through the hierarchy. As a reminder, capacity at the first two layers in the hierarchy does not need to be reclaimed; therefore, the first layer examined in this paper is the third layer from the bottom.

Array Allocated Devices

Lost or orphaned devices at the Array Allocated layer are the low-hanging fruit that you should target first in any reclamation effort because they require the least LOE and deliver the highest ROI. These devices are not accessible or in use by any host(s); therefore, no effort is required from a host perspective to return these devices to the free pool and make them available for allocation elsewhere.

Definition

As stated previously, devices are considered array allocated when they are both mapped and masked. In addition, some IT departments consider devices Array Allocated when they are either mapped **or** masked, depending on internal policy. In either case, array allocated devices may be reclaimed if they are not actually allocated.

Devices that are both mapped and masked but not allocated to any host are potentially reclaimable.

Devices that are mapped to a front-end array port with access controls enabled but not masked to a HBA WWN are inaccessible to any host. Some storage administrators automatically map all devices to front-end array ports but leave them unmasked as a technique for creating free pools of devices. Other IT departments consider their free pools to consist of devices that are unmapped **and** unmasked; in these situations, unmasked but mapped devices are potentially reclaimable.

Devices that are masked to a HBA port WWN but not mapped to a front-end array port that has access controls enabled are also inaccessible by any host; therefore, masked but unmapped devices are potentially reclaimable.

Possible Causes

Possible scenarios leading to devices at this layer being potentially reclaimable include:

- Decommissioning of applications or hosts – the devices were never unmapped and/or unmasked
- Device reservation techniques – mapping or masking was used as a technique for logically reserving devices for future use
- Temporary usage – the devices were used for some temporary purpose but never unmapped and/or unmasked

Reclamation Process

Tasks required to reclaim these devices involve modification or removal of mapping and/or masking configurations. These tasks generally require the time and expertise of storage administrators.

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the devices are not being used.

Host Accessible Devices

Devices lost or orphaned at the Host Accessible layer should subsequently be targeted second in a reclamation project, following reclamation of array allocated devices. Reclaiming these devices requires slightly more effort (as compared to array allocated devices) but still yields a strong ROI. To return these devices to the free pool, relatively minor effort is required from the server administration team and the storage administration team must remove mapping and/or masking configurations.

Definition

Devices are considered host accessible when hosts can directly use those devices for I/O operations. In this state of configuration, a device must be array allocated, zoned, and mounted to a host. Devices orphaned at this layer are array allocated to one or more hosts but are not accessible to those hosts, possibly due to SAN zoning issues.

Possible Causes

Possible scenarios leading to devices at this layer being potentially reclaimable include:

- Invalid or missing zones
- Lack of physical connectivity to the host

Reclamation Process

Tasks required to reclaim these devices involve:

- Possibly unmounting the devices from the host by a server administrator
- Modification or removal of zones by a SAN administrator
- Modification or removal of mapping and masking configurations by a storage administrator

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the devices are not being used.

Volume Groups

After targeting devices orphaned at the Array Allocated and Host Accessible layers, you should examine volume groups for potentially reclaimable devices. Reclaiming devices from volume groups requires additional effort (as compared to reclaiming host accessible devices) with a slightly lower ROI. To return these devices to the free pool, moderate effort is required from the server administration team, and the storage administration team must remove mapping and/or masking configurations.

Definition

A volume group is configured on a host and is a collection of host accessible storage devices (or volumes) from which logical volumes can be created. Host accessible devices are considered "used" at the Volume Group layer when they are included in a volume group. In this state of configuration, a device must be array allocated, host accessible, and included in a volume group. Devices orphaned at this layer are accessible to one or more hosts but do not have any volume groups or other data structures built on them and are not used in any other way.

Possible Causes

Possible scenarios leading to devices at this layer being potentially reclaimable include:

- Devices reserved but forgotten – devices may have been allocated to a host for a project that was subsequently cancelled but the devices were not unallocated
- Devices removed from volume groups but not used for any other purpose – an application may have been decommissioned and its volume groups removed but the devices were not unallocated

Reclamation Process

Tasks required to reclaim these devices involve:

- Unmounting the devices from the host by a server administrator
- Modifying or removing SAN zones, mapping, and masking configurations by a storage administrator

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the devices are not being used.

VMware File Systems

You may examine VMware File Systems for potentially reclaimable devices if VMware ESX Servers are implemented in the IT environment and are using SAN-based storage. As with volume groups, you should examine VMware File Systems after targeting devices orphaned at the Array Allocated and Host Accessible layers. Reclaiming devices from VMFSs requires additional effort on ESX servers (as compared to host accessible devices) with a slightly lower ROI. To return these devices to the free pool, moderate effort is required from the ESX server administration team, and the storage administration team must remove mapping and/or masking configurations.

Definition

A VMware File System is configured on an ESX server and is a special file system built on a collection of VMware volumes that correspond to mounted devices. Host accessible devices are considered “used” at the VMware File System layer when they are included in a VMFS. In this state of configuration, a device must be array allocated, host accessible, and included in a VMFS. Orphaned devices at this layer are accessible to one or more ESX servers but do not have any VMware file systems or other data structures built on them and are not used in any other way.

Possible Causes

Possible scenarios leading to devices at this layer being potentially reclaimable include:

- Devices reserved but forgotten – devices may have been allocated to a host for a project that was subsequently cancelled but the devices were not unallocated
- Devices removed from VMFSs but not used for any other purpose – an application may have been decommissioned and the VMFSs removed but the devices were not unallocated

Reclamation Process

Tasks required to reclaim these devices involve:

- Removing unused volumes (i.e. devices) from the VMFS, or deleting the entire VMware file system by a server administrator
- Unmounting the devices from the host by a server administrator
- Modifying or removing SAN zones, mapping, and masking configurations by a storage administrator

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the devices are not being used.

Logical Volumes

Reclaiming capacity at the Logical Volume layer in the hierarchy may be significantly more difficult than at lower layers, and you may decide that the effort to reclaim devices is not worth the benefit.

Following the creation of volume groups on a server, server administrators normally use the volume group space by configuring logical volumes that can be used to create data structures for applications, file systems, and databases. Free space may exist in volume groups if the entire volume group is not utilized with logical volumes, but the devices on which that free space exists may not be easily reclaimed. Due to how volume groups are configured or how logical volumes are configured within a volume group (e.g. striped across the group), reclaiming devices may require deleting all logical volumes in the group and removing the volume group itself. If any data structures exist on logical volumes in a group, removing those volumes and the group may require host-level migration of the data to a different volume group.

The LOE associated with reclaiming devices from the Logical Volume layer in the hierarchy is significantly higher, and yields a much lower ROI, when compared to reclaiming devices at the Host Accessible layer. To return devices at the Logical Volume layer to the free pool, application or system downtime may be required, significant effort is required from the server administration team, and the storage administration team must remove mapping and/or masking configurations.

One alternative to reclaiming devices at the Logical Volume layer is to increase utilization of volume group space by creating additional logical volumes. This alternative provides more space for data structures on the same host but does not allow reclamation of those devices for reallocation to other hosts.

Definition

A logical volume is a virtual device that appears to applications, databases, and file systems like a physical disk device. Logical volumes are defined and exist within volume groups but are not restricted to a specific disk or disks in the group. Logical volumes can be striped or concatenated within the volume group, and provide a more flexible layer of abstraction above physical devices and volume groups. A volume group is considered "used" at the Logical Volume layer in the hierarchy by one or more logical volumes configured in the group. Reclaimable capacity at this layer exists in the form of space in a volume group on which no logical volumes are configured and is not used in any other way.

Possible Causes

Possible scenarios leading to capacity at this layer being potentially reclaimable include:

- Volume groups not fully utilized by logical volumes – an application may have been using logical volumes that were removed when the application was decommissioned, but the volume group was never removed or the space was never reused

Reclamation Process

Tasks required to reclaim devices occupied by volume groups include:

- Deleting remaining logical volumes from the volume group by a server administrator (be sure that a logical volume is no longer in use before reclaiming its storage space)
- Removing unused devices from the volume group, or deleting the entire volume group by a server administrator
- Unmounting the devices from the host by a server administrator
- Modifying or removing SAN zones, mapping, and masking configurations by a storage administrator

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the devices are not being used.

VMware Virtual Disks

Reclaiming capacity at the VMware Virtual Disk layer in the hierarchy is as difficult as reclaiming logical volume capacity and may not be worth the effort. Increasing VMware file system utilization may be a more beneficial approach.

Following the creation of VMware file systems on an ESX server, server administrators normally use the VMFS space by configuring virtual disks that can be used by virtual machines. Free space may exist in VMware file systems if the entire VMFS is not utilized with VMDKs, but the devices on which that free space exists may not easily be reclaimed. Reclaiming may require deletion of the entire VMFS. If any virtual machines are using VMDKs in a VMFS, removing those VMDKs and the VMFS may require migrating the VMDKs to different VMware file systems.

As with logical volumes, the LOE associated with reclaiming devices from the VMware Virtual Disk layer in the hierarchy is significantly higher with a much lower ROI as compared to reclaiming devices at the Host Accessible layer. To return devices at this layer to the free pool, application or system downtime may be required, significant effort is required from the server administration team, and the storage administration team must remove mapping and/or masking configurations.

Increasing utilization of VMFS space by creating additional VMDKs is one alternative to reclaiming devices at the VMware Virtual Disk layer in the hierarchy. This alternative provides more space for virtual machines but does not allow reclamation of those devices for reallocation to other hosts.

Definition

A VMware Virtual Disk is a file in a VMFS that always appears to a virtual machine as a mounted SCSI device. A VMware file system is considered "used" at the VMware Virtual Disk layer in the hierarchy by one or more VMDKs configured in the VMFS. Reclaimable capacity at this layer exists in the form of space in a VMFS on which no VMDKs are configured and is not used in any other way.

Possible Causes

Possible scenarios leading to capacity at this layer being potentially reclaimable include:

- VMFSs not fully utilized by VMDKs – a virtual machine may have been using VMDKs that were removed when the VM was decommissioned, but the VMFS was never removed or the space was never reused

Reclamation Process

Tasks required to reclaim devices occupied by VMware file systems involve:

- Deleting VMDKs from the VMFS by a server administrator (be sure that a VMDK is no longer in use before reclaiming its storage space)
- Removing unused devices from the VMFS, or deleting the entire VMFS by a server administrator
- Unmounting the devices from the ESX server by a server administrator
- Modifying or removing SAN zones, mapping, and masking configurations by a storage administrator

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the devices are not being used.

File Systems

Reclaiming capacity at the File System layer in the hierarchy is even more difficult than at the Logical Volume layer, and you may decide that increasing file system utilization is a more beneficial approach than attempting to reclaim devices.

Following the creation of volume groups and logical volumes on a server, server administrators usually create data structures on the logical volumes, such as file systems. Free space may exist in logical volumes if the entire volume is not utilized with file systems or other data structures, but the devices on which that free space exists cannot be reclaimed for use on other servers without extensive effort. Reclaiming devices may require deleting all file systems and other data structures in all logical volumes that exist in a volume group, followed by removing those logical volumes and the related volume group. If data exists in those file systems that must be retained, removing those file systems, logical volumes and the volume group may require host-level migration of the data to file systems built on logical volumes in a different volume group, possibly on a different host.

The LOE associated with reclaiming devices from the File System layer in the hierarchy is even higher and yields an even lower ROI when compared to reclaiming devices at the Logical Volume layer. To return devices at the File System layer to the free pool, application or system downtime may be required, significant effort is required from the server administration team, and the storage administration team must remove mapping and/or masking configurations

Reclaiming file system space for new projects or data before allocating more devices is one alternative. This can increase file system utilization but does not allow reclamation of those devices for reallocation to other hosts. One method for reclaiming file system space involves identifying and removing duplicate and temporary files. Another method is to identify aged and dormant files, and archive them. A third method involves identifying file types that should not be stored on enterprise storage (e.g. personal audio files) and removing those files.

Definition

A file system is a structure for organizing and storing data in files and is represented at the File System layer in the storage configuration hierarchy. File systems are usually created on lower-level structures such as logical volumes. A file system is considered "used" at the File System layer when data files are created and saved. Reclaimable capacity at this layer exists in the form of space in a logical volume on which no file systems or other data structures are configured and is not used in any other way.

Possible Causes

Possible scenarios leading to devices at this capacity being potentially reclaimable include:

- File system size requirement is not as large as originally estimated – logical volumes may have been initially configured for an application whose space requirements subsequently decreased, leaving unused space in the logical volumes

Reclamation Process

Tasks required to reclaim these devices involve:

- Potentially deleting some or all file systems configured on logical volumes by a server administrator
- Potentially deleting some or all logical volumes from the volume group by a server administrator
- Removing the unused devices from the volume group, or deleting the entire volume group by a server administrator
- Unmounting the devices from the host by a server administrator
- Modifying or removing SAN zones, mapping, and masking configurations by a storage administrator

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the file systems, other data structures, logical volumes, volume groups, and devices are not being used.

Databases

Reclaiming capacity at the Database layer in the hierarchy is extremely difficult. You may decide that increasing database utilization is a more beneficial approach than attempting to reclaim devices.

Database administrators usually configure databases on file systems or on raw devices mounted to a host. Free space may exist in databases if the entire tablespace is not utilized with user or application data. As with file systems, the devices on which that free space exists cannot be reclaimed for use on other servers without extensive effort. Reclaiming devices occupied by a database would require the deletion of the database and all file systems, logical volumes, and volume groups occupied by the database.

If data exists in the database and/or at any lower layer in the hierarchy must be retained, removal of the database, file systems, logical volumes and volume groups may require exporting the data from the database, and host-level migration of all the required data to a database and file systems built on logical volumes in a different volume group, possibly on a different server.

The LOE associated with reclaiming devices from the Database layer in the hierarchy is higher than at any other layer and yields the lowest ROI. To return devices at the Database layer to the free pool, application or system downtime may be required, significant effort is required by the database and server administration teams, and the storage administration team must remove mapping and/or masking configurations.

Using free database space for new projects or data before allocating more devices to the server is one alternative to reclaiming devices at the Database layer in the hierarchy. This alternative increases database utilization but does not allow reclamation of the underlying devices for reallocation to other hosts.

Definition

A database is a structured collection of records or data, managed by database management system software. You can configure databases on lower-level structures such as file systems and logical volumes, and on raw devices mounted to hosts. A database is considered "used" when records or data are created and stored. Reclaimable capacity at this layer exists in the form of space in a database tablespace in which no data exists.

Possible Causes

Possible scenarios leading to capacity at this layer being potentially reclaimable include:

- Method of reserving space for future database requirements – some database administrators create tablespaces but do not use them.
- Forgotten space – situations may occur that lead to the creation of tablespaces that are subsequently forgotten

Reclamation Process

Tasks required to reclaim these devices involve:

- Deleting the database and tablespace by a database administrator
- Potentially deleting some or all file systems configured on logical volumes by a server administrator
- Potentially deleting some or all logical volumes from the volume group by a server administrator
- Removing the unused devices from the volume group, or deleting the entire volume group by a server administrator
- Unmounting the devices from the host by a server administrator
- Modifying or removing SAN zones, mapping, and masking configurations by a storage administrator

Warning: Before reclaiming any devices, I strongly recommend additional research to verify that the databases, file systems, other data structures, logical volumes, volume groups, and devices are not being used.

Summary

Before you embark on a project to reclaim storage, you should choose your targets wisely. Various layers in the storage configuration hierarchy in your environment may include potentially reclaimable capacity, but the level of effort may not be worth the return on investment.

Every IT department configures their storage environments differently and has different policies, so optimal targets in one environment may not be optimal in others. You need to choose the targets that make the most sense in your specific environment, and you will require extensive and accurate information to make effective decisions.

An SRM reporting tool is a critical component to provide the information you need, not only to choose optimal targets but also to monitor the progress of your reclamation efforts over time.

Using StorageScope to Identify Reclaimable Storage

Many different pieces of information are required during a storage reclamation project. You must examine various items to find potentially reclaimable storage at each layer in the storage configuration hierarchy. You must then examine the state of configuration of that space or device to know whether it is reclaimable. Finally, you will need to track utilization and free capacity against a baseline to provide a measure of success.

Although you can gather all this information manually, an SRM reporting tool will significantly speed up the process of gathering, manipulating, and presenting the information, further reducing overall level of effort and increasing the return on investment. ControlCenter StorageScope®, a component of the EMC ControlCenter® suite of products, is a powerful and flexible reporting tool that reports what storage resources you have in your infrastructure, how those resources are used, and who is using them. More specifically, StorageScope provides all the information you need to identify potentially reclaimable capacity at all layers in the storage configuration hierarchy and to monitor the success of your reclamation effort.

StorageScope "Canned" Reports

StorageScope, an integral piece of the EMC ControlCenter suite of storage management software, provides a variety of features to help you:

- View configuration, status, and usage information for individual objects, user-defined groups, or the entire enterprise so you can assess your current storage environment
- Determine future storage needs based on historical usage and trending reports
- Reclaim storage resources by identifying unused or underutilized storage, as well as duplicate, rarely accessed, or non-business files
- Facilitate billing and chargeback operations by location, line of business, or application
- View point-in-time snapshots of high-importance areas of your storage environment on StorageScope's customizable Dashboard page
- Generate custom database queries unique to your enterprise

To facilitate the effort to reclaim storage, StorageScope provides various SRM Views, Queries, and file-level reports that you can use to identify potentially reclaimable capacity. In addition, you can use Query Builder to generate custom queries specific to the reclamation effort.

SRM Views

The SRM Views provide a consolidated view of the configuration, capacity, and utilization of the storage assets in your enterprise. For identification of a subset of the potentially reclaimable storage across all arrays at the Array Allocated layer in the hierarchy, StorageScope provides the Reclaimable Capacity listing in the Arrays View, as shown in *Figure 4*.

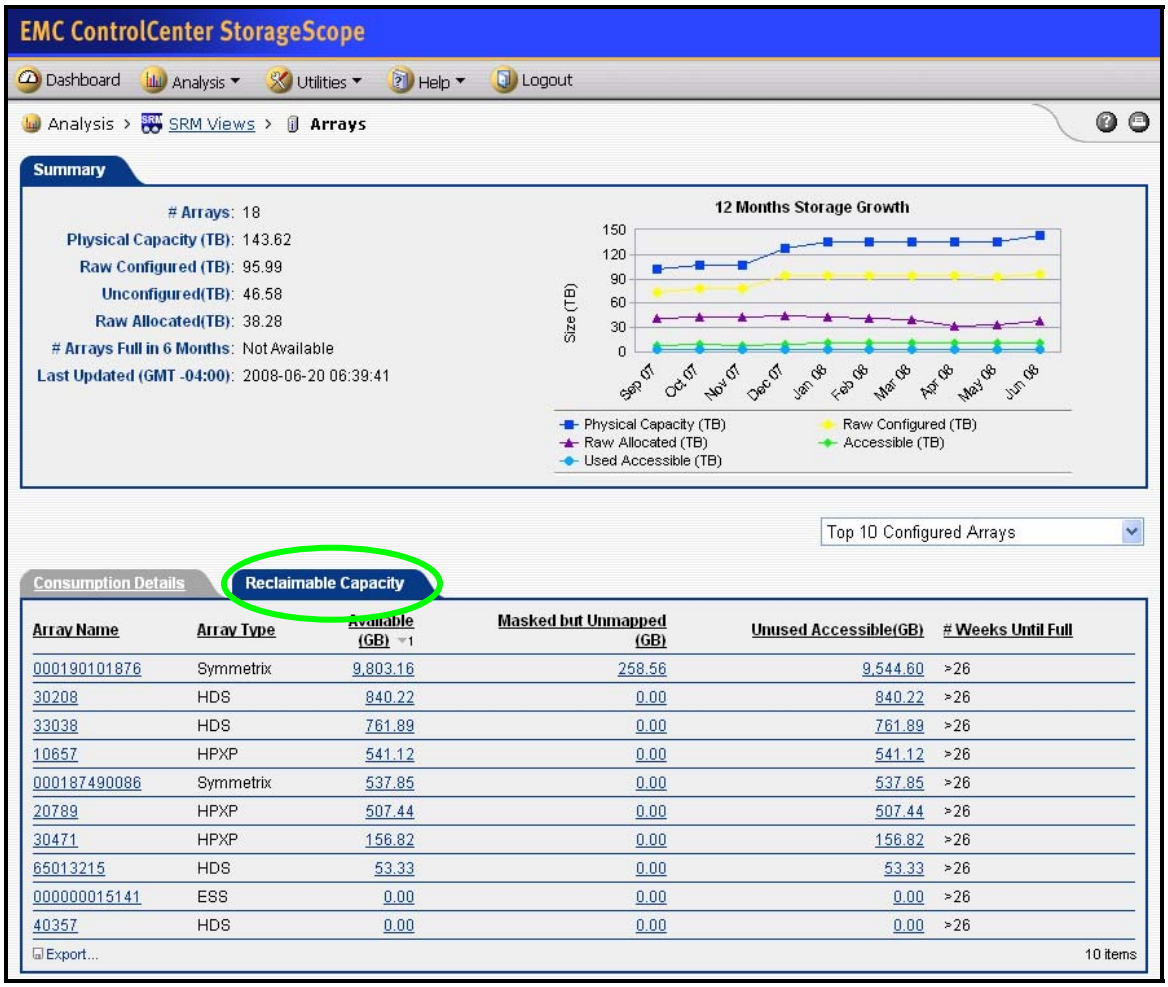


Figure 4 - Reclaimable Capacity, Arrays View

The “Masked but Unmapped” column identifies total capacity per array that is masked to a HBA port WWN but not mapped to any array front-end ports. In addition, the “Unused Accessible” column identifies total capacity per array that is accessible to hosts but is not used by any type of data structure (e.g. volume groups, file systems, raw databases, VMFS volumes, RDMs, etc.).

If you click on an amount for an array in the “Masked but Unmapped” column, you will see all the devices on that array that are masked but not mapped, as shown in *Figure 5*. This view helps to identify reclaimable devices at the Array Allocated layer in the hierarchy. As a reminder, the devices in this list are potentially reclaimable but not all the devices should be reclaimed (e.g. gatekeeper devices for Symmetrix arrays, System Resource device types, etc.)

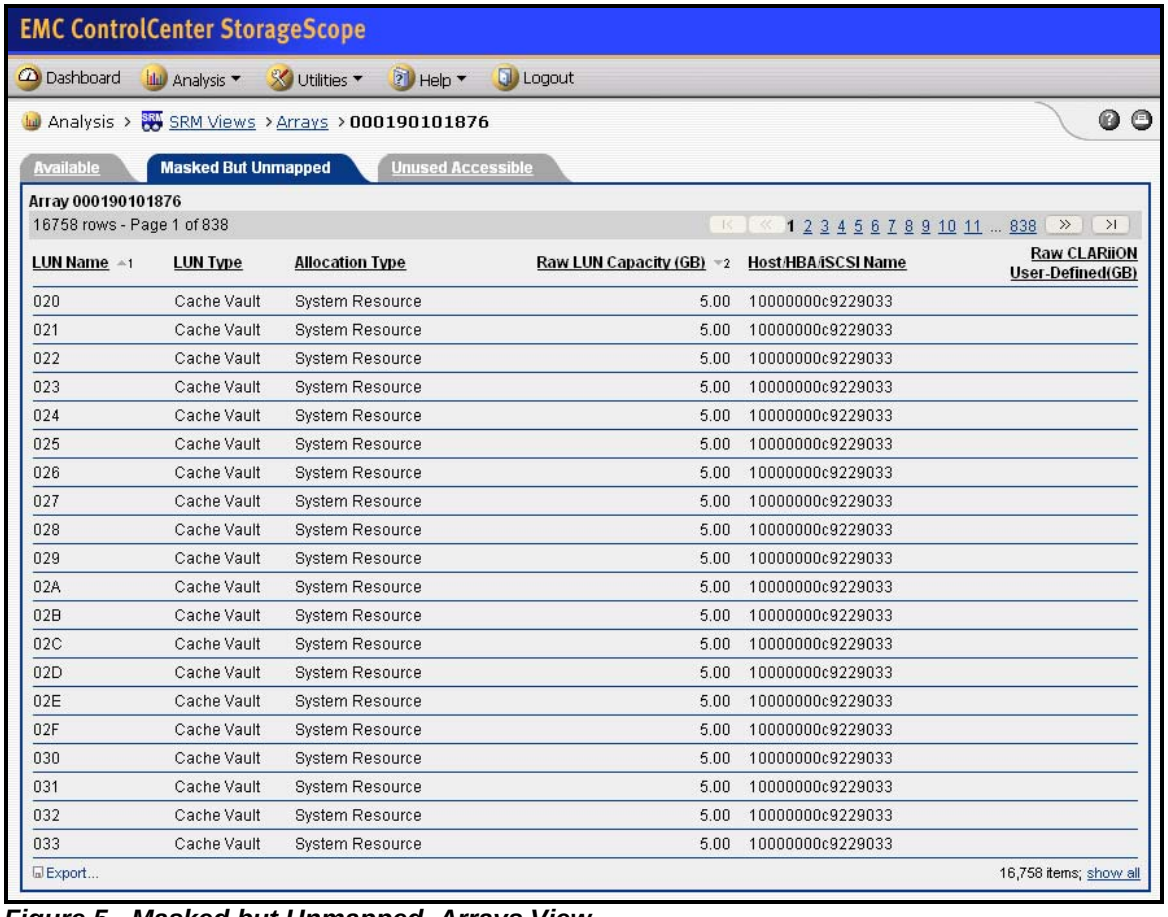


Figure 5 - Masked but Unmapped, Arrays View

From this view, you may also click on the “Unused Accessible” tab to view all the devices on that array that are accessible to hosts but not used in any way, as shown in *Figure 6*. This view helps to identify reclaimable devices at the Volume Groups, VMware File Systems layer in the hierarchy. As a reminder, the devices in this list are potentially reclaimable – you should perform additional analysis to verify whether you can reclaim the devices without causing any problems.

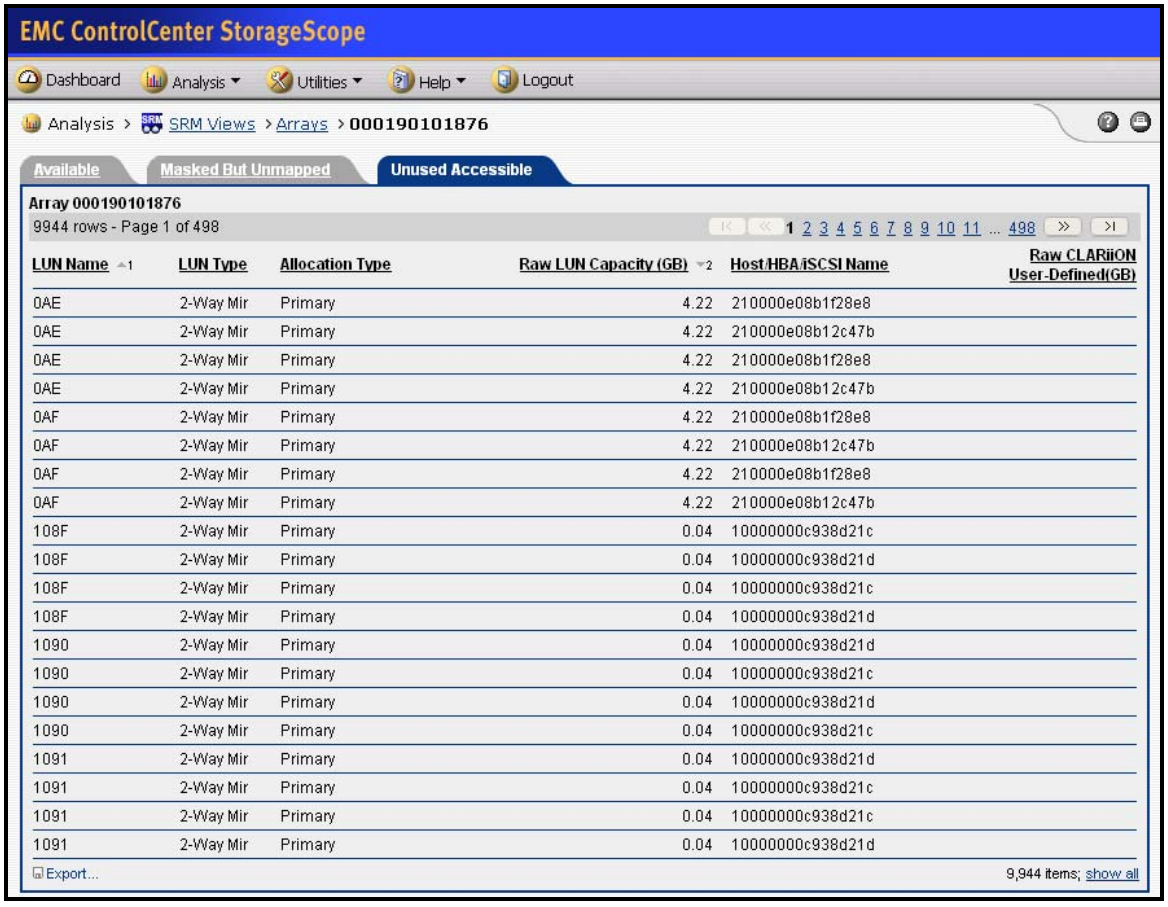


Figure 6 - Unused Accessible, Arrays View

Reclaiming file system space and reusing that capacity for new applications or data is one option to reclaiming devices at the File System layer in the hierarchy (which is a very difficult effort). To facilitate such an effort, StorageScope provides the File Level Storage View as shown in *Figure 7*.

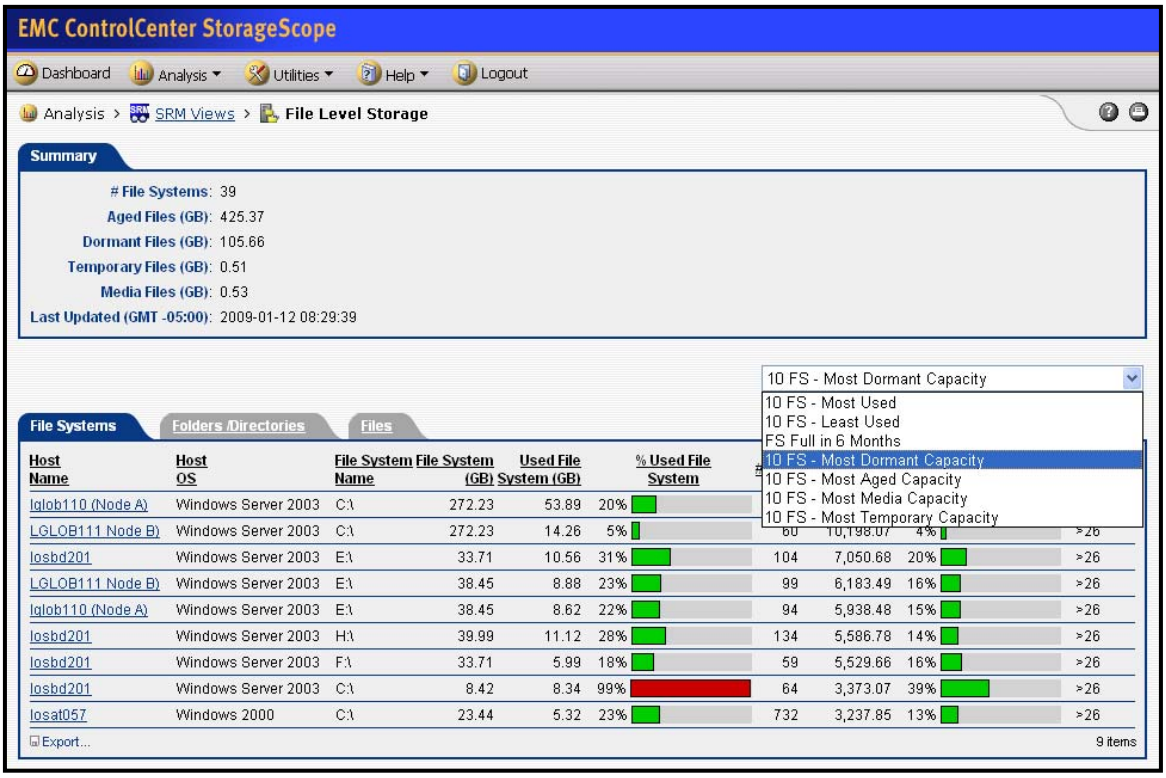


Figure 7 - Reclaimable Capacity, File Level Storage View

The File Systems tab in this view includes several filters that identify the ten file systems with the most dormant capacity, most aged capacity, most media capacity, and most temporary capacity. The Folders/Directories and Files tabs in this view include filters that identify the top ten folders and hosts, respectively, with the same types of capacity.

Note: File-level discovery and reporting is provided with StorageScope File Level Reporting, a separately licensable feature.

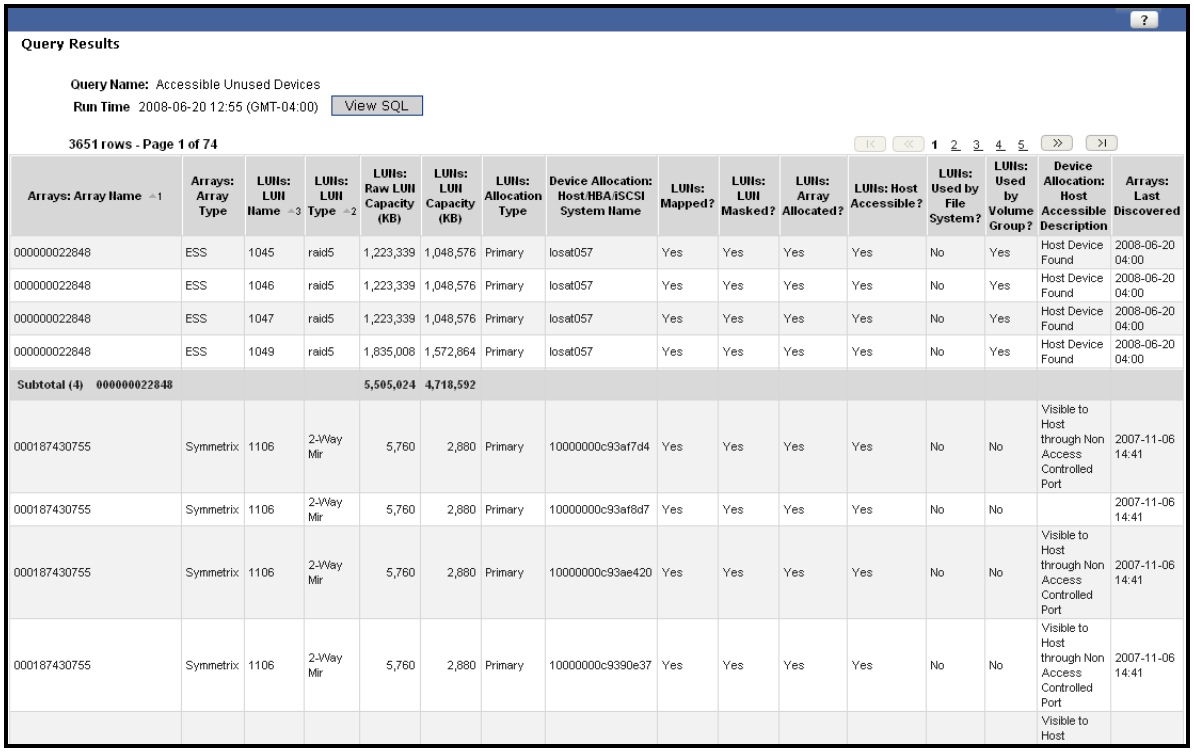
Queries

StorageScope allows the execution of queries against the StorageScope Repository to produce reports to meet most requirements. You can run them interactively or schedule them to run on a periodic basis. StorageScope provides many "canned" queries that you may execute or modify to meet your needs. Query Builder, a wizard-based SQL utility in StorageScope, allows you to create and modify queries.

Several of the “canned” queries identify potentially reclaimable capacity:

- Accessible Unused Devices
- Masked, but Not Connected
- Masked, Mapped, but Not Accessible

The “Accessible Unused Devices” query provides a listing per array of devices that are accessible to hosts but unused, as well as other columns that provide more details regarding the specific devices and the reasons they might be reclaimable, as shown in *Figure 8*. This query helps to identify reclaimable devices at the Volume Groups and VMware File Systems layer in the hierarchy.



Query Results

Query Name: Accessible Unused Devices
Run Time 2008-06-20 12:55 (GMT-04:00) View SQL

3651 rows - Page 1 of 74 1 2 3 4 5

Arrays: Array Name	Arrays: Array Type	LUNs: LUN Name	LUNs: LUN Type	LUNs: Raw LUN Capacity (KB)	LUNs: LUN Capacity (KB)	LUNs: Allocation Type	Device Allocation: Host/HBA/iSCSI System Name	LUNs: Mapped?	LUNs: LUN Masked?	LUNs: Array Allocated?	LUNs: Host Accessible?	LUNs: Used by File System?	LUNs: Used by Volume Group?	Device Allocation: Host Accessible Description	Arrays: Last Discovered
000000022848	ESS	1045	raid5	1,223,339	1,048,576	Primary	losat057	Yes	Yes	Yes	Yes	No	Yes	Host Device Found	2008-06-20 04:00
000000022848	ESS	1046	raid5	1,223,339	1,048,576	Primary	losat057	Yes	Yes	Yes	Yes	No	Yes	Host Device Found	2008-06-20 04:00
000000022848	ESS	1047	raid5	1,223,339	1,048,576	Primary	losat057	Yes	Yes	Yes	Yes	No	Yes	Host Device Found	2008-06-20 04:00
000000022848	ESS	1049	raid5	1,835,008	1,572,864	Primary	losat057	Yes	Yes	Yes	Yes	No	Yes	Host Device Found	2008-06-20 04:00
Subtotal (4) 000000022848				**5,505,024**	**4,718,592**										
000187430755	Symmetrix	1106	2-Way Mir	5,760	2,880	Primary	10000000c93af7d4	Yes	Yes	Yes	Yes	No	No	Visible to Host through Non Access Controlled Port	2007-11-06 14:41
000187430755	Symmetrix	1106	2-Way Mir	5,760	2,880	Primary	10000000c93af8d7	Yes	Yes	Yes	Yes	No	No		2007-11-06 14:41
000187430755	Symmetrix	1106	2-Way Mir	5,760	2,880	Primary	10000000c93ae420	Yes	Yes	Yes	Yes	No	No	Visible to Host through Non Access Controlled Port	2007-11-06 14:41
000187430755	Symmetrix	1106	2-Way Mir	5,760	2,880	Primary	10000000c9390e37	Yes	Yes	Yes	Yes	No	No	Visible to Host through Non Access Controlled Port	2007-11-06 14:41
														Visible to Host	

Figure 8 - Query: Accessible Unused Devices

The “Masked, but Not Connected” query provides a listing per array of devices that are masked to hosts but missing connectivity, as well as other columns that provide more details regarding the specific devices and the reasons they might be reclaimable, as shown in *Figure 9*. This view helps to identify reclaimable devices at the Host Accessible layer in the hierarchy.

Query Results

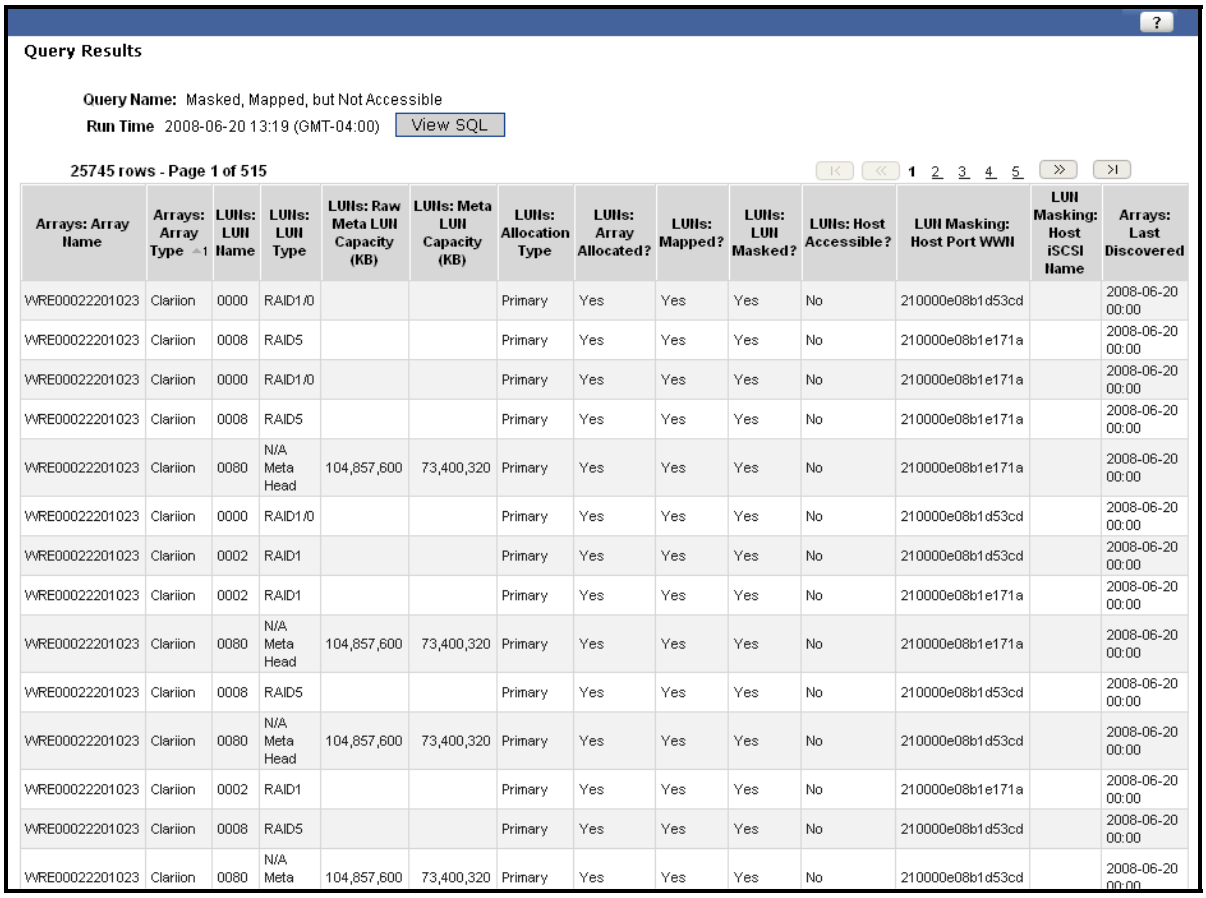
Query Name: Masked, but Not Connected
Run Time 2008-06-20 13:01 (GMT-04:00) View SQL

45489 rows - Page 1 of 910 1 2 3 4 5

Arrays: Array Name	Arrays: Array Type	LUNs: LUN Name	LUNs: LUN Type	LUNs: Raw Meta LUN Capacity (KB)	LUNs: Meta LUN Capacity (KB)	LUNs: Allocation Type	LUNs: Array Allocated?	LUNs: Emulation	LUNs: Mapped?	LUNs: LUN Masked?	LUNs: Host Accessible?	Ports: Array Port Name	LUN Masking: Host Port WWN	LUN Masking: Host iSCSI Name	Arrays: Last Discovered
APM00063106024	Clariion	0007	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	210100e08b3cccb5		2008-01-11 00:00
APM00063106024	Clariion	0007	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c944c654		2008-01-11 00:00
APM00063106024	Clariion	0017	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c9488a3a		2008-01-11 00:00
APM00063106024	Clariion	0017	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c9488a3b		2008-01-11 00:00
APM00063106024	Clariion	0017	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	210100e08b3cccb5		2008-01-11 00:00
APM00063106024	Clariion	0017	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c944c654		2008-01-11 00:00
APM00063106024	Clariion	0019	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c944c654		2008-01-11 00:00
APM00063106024	Clariion	0019	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	210000e08b1cccb5	N/A	2008-01-11 00:00
APM00063106024	Clariion	0019	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c944c655		2008-01-11 00:00
APM00063106024	Clariion	0019	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c93d465b		2008-01-11 00:00
APM00063106024	Clariion	0023	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c9488a3a		2008-01-11 00:00
APM00063106024	Clariion	0023	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c9488a3b		2008-01-11 00:00
APM00063106024	Clariion	0023	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	210100e08b3cccb5		2008-01-11 00:00
APM00063106024	Clariion	0023	RAID1/0			Primary	Yes	Snapshot Source	Yes	Yes	Yes	losay020 Port 0	10000000c944c654		2008-01-11 00:00
WRE00022201023	Clariion	0008	RAID5			Primary	Yes		Yes	Yes	No	172.23.141.182 Port 0	210000e08b1d53cd		2008-06-20 00:00

Figure 9 - Query: Masked, but Not Connected

The “Masked, Mapped, but Not Accessible” query provides a listing of devices, per array, that are masked to hosts and mapped to array front-end ports but not accessible to any hosts, as well as other columns that provide more details regarding the specific devices and the reasons they might be reclaimable, as shown in *Figure 10*. This view helps to identify reclaimable devices at the Host Accessible layer in the hierarchy.



Query Results

Query Name: Masked, Mapped, but Not Accessible

Run Time 2008-06-20 13:19 (GMT-04:00) View SQL

25745 rows - Page 1 of 515 1 2 3 4 5

Arrays: Array Name	Arrays: Array Type	LUNs: LUN Name	LUNs: LUN Type	LUNs: Raw Meta LUN Capacity (KB)	LUNs: Meta LUN Capacity (KB)	LUNs: Allocation Type	LUNs: Array Allocated?	LUNs: Mapped?	LUNs: LUN Masked?	LUNs: Host Accessible?	LUN Masking: Host Port WWN	LUN Masking: Host iSCSI Name	Arrays: Last Discovered
WRE00022201023	Clariion	0000	RAID1/0			Primary	Yes	Yes	Yes	No	210000e08b1d53cd		2008-06-20 00:00
WRE00022201023	Clariion	0008	RAID5			Primary	Yes	Yes	Yes	No	210000e08b1e171a		2008-06-20 00:00
WRE00022201023	Clariion	0000	RAID1/0			Primary	Yes	Yes	Yes	No	210000e08b1e171a		2008-06-20 00:00
WRE00022201023	Clariion	0008	RAID5			Primary	Yes	Yes	Yes	No	210000e08b1e171a		2008-06-20 00:00
WRE00022201023	Clariion	0080	N/A Meta Head	104,857,600	73,400,320	Primary	Yes	Yes	Yes	No	210000e08b1e171a		2008-06-20 00:00
WRE00022201023	Clariion	0000	RAID1/0			Primary	Yes	Yes	Yes	No	210000e08b1d53cd		2008-06-20 00:00
WRE00022201023	Clariion	0002	RAID1			Primary	Yes	Yes	Yes	No	210000e08b1d53cd		2008-06-20 00:00
WRE00022201023	Clariion	0002	RAID1			Primary	Yes	Yes	Yes	No	210000e08b1e171a		2008-06-20 00:00
WRE00022201023	Clariion	0080	N/A Meta Head	104,857,600	73,400,320	Primary	Yes	Yes	Yes	No	210000e08b1e171a		2008-06-20 00:00
WRE00022201023	Clariion	0008	RAID5			Primary	Yes	Yes	Yes	No	210000e08b1d53cd		2008-06-20 00:00
WRE00022201023	Clariion	0080	N/A Meta Head	104,857,600	73,400,320	Primary	Yes	Yes	Yes	No	210000e08b1d53cd		2008-06-20 00:00
WRE00022201023	Clariion	0002	RAID1			Primary	Yes	Yes	Yes	No	210000e08b1e171a		2008-06-20 00:00
WRE00022201023	Clariion	0008	RAID5			Primary	Yes	Yes	Yes	No	210000e08b1d53cd		2008-06-20 00:00
WRE00022201023	Clariion	0080	N/A Meta	104,857,600	73,400,320	Primary	Yes	Yes	Yes	No	210000e08b1d53cd		2008-06-20 00:00

Figure 10 - Query: Masked, Mapped, but Not Accessible

Built-In Reports

StorageScope provides built-in reports based on Crystal Report descriptions. They can be run interactively or as part of a scheduled job. StorageScope provides at least twelve (12) built-in reports.

For example, the Duplicate Files built-in report provides a listing of all duplicate files, and their locations and sizes, as shown in *Figure 11*. This view helps to identify file system space that may be freed up at the File System Used layer in the hierarchy by archiving or deleting unnecessary files.

EMC ControlCenter™ StorageScope

Duplicate Files Report

Listing of duplicate files by host

2008-06-20 14:11:43 (GMT-04:00)

Page 1 of 33

EMC² where information lives®

Host Name	Path	Allocated File Size (MB)	Actual File Size (MB)	Last Accessed
File Name: __temp.zip		**1.35**	**1.34**	
losat057	C:\Documents and Settings\Administrator\Local Settings\Temp\unz4D7.tmp\			9/20/2006 2:11:24 PM
losat057	C:\WINNT\Temp\unz4D8.tmp\			9/20/2006 2:11:26 PM
Excess Size for __temp.zip:			**1.34**	
File Name: __temp.zip		**1.35**	**1.34**	
losat057	C:\Documents and Settings\Administrator\Local Settings\Temp\unz45F.tmp\			9/20/2006 2:08:55 PM
losat057	C:\WINNT\Temp\unz460.tmp\			9/20/2006 2:09:01 PM
Excess Size for __temp.zip:			**1.34**	
File Name: _sys1.cab		**2.55**	**2.55**	
losat057	C:\drivers\530v\			10/ 5/2001 12:51:53 PM

Figure 11 - Built-in report: Duplicate Files

Note: File-level discovery and reporting is provided with StorageScope File Level Reporting, a separately licensable feature.

StorageScope Custom Queries

As mentioned previously, StorageScope provides queries to produce reports that meet most requirements. In some situations, you may need to create your own custom queries to satisfy your reporting requirements. Query Builder, a wizard-based SQL query builder utility in StorageScope, allows you to create and modify queries without having to develop SQL statements from scratch.

StorageScope provides hundreds of fields that can be included in reports; you may use a subset of them to identify potentially reclaimable storage. Query Builder provides access to various fields in the StorageScope Repository tables that are helpful to identify reclaimable storage at each layer in the storage configuration hierarchy.

Table 1 lists each layer in the hierarchy and the related categories and columns available in Query Builder. The listing is not all-inclusive, so you may find other columns helpful as well.

Storage Configuration Hierarchy Layer	StorageScope Query Builder Category	StorageScope Query Builder Column
Raw	Arrays	Arrays.Array Name Arrays.Physical Capacity Arrays.Unconfigured
Usable	Arrays	Arrays.Array Name Arrays.Physical Capacity Arrays.Configure Arrays.Unconfigured Arrays.Raw Configured
Array Allocated	LUNs	LUNs.LUN Name LUNs.Array Allocated? LUNs.HDS Allocated? LUNs.Host Allocated? LUNs.LUN Masked? LUNs.Mapped?
Host Accessible	LUNs	LUNs.LUN Name LUNs.Host Allocated? LUNs.Host Accessible? LUNs.Used by File System? LUNs.Used by Volume Group? LUNs.Used by Virtual Machine?
	Host Devices	Host Devices.Host Device Name Host Devices.Used by File System? Host Devices.Used by Volume Group?
Volume Groups, VMware File Systems	Host Devices	Host Devices.Host Device Name Host Devices.Used by File System? Host Devices.Used by Volume Group?
	Volume Groups	Volume Groups.Volume Group Name Volume Groups.# Logical Volumes Volume Groups.Free Volume Group Volume Groups.Used Volume Group

Storage Configuration Hierarchy Layer	StorageScope Query Builder Category	StorageScope Query Builder Column
	ESX Server	ESX Server.Host Name ESX Server.# Accessible LUNs ESX Server.# Allocated LUNs ESX Server.Accessible ESX Server.Allocated ESX Server.VMFS Capacity ESX Server.Used VMFS
Logical Volumes, VMware Virtual Disks	Logical Volumes	Logical Volumes.Logical Volume Name Logical Volumes.Free Logical Volume Logical Volumes.Used Logical Volume
	ESX Server	ESX Server.Host Name ESX Server.VMDK Capacity ESX Server.VMDK on Arrays
File Systems	File Systems	File Systems.File System Name File Systems.Free File System File Systems.Used File System File Systems.VMFS?
	Files	Files.File Name Files.Actual Size Files.Date Created Files.File Name Extension Files.Last Accessed Files.Last Discovered Files.Last Modified Files.Owner
Database	Databases	Databases.Database Name Databases.Database Capacity Databases.Data Capacity Databases.Free Database Databases.Used Database
	Oracle Tablespaces	Oracle Tablespaces.Tablespace Name Oracle Tablespaces.Tablespace Size Oracle Tablespaces.Used Tablespace

Table 1 - Helpful Query Builder Categories and Columns

StorageScope provides “canned” reports in the forms of SRM Views, built-in reports, and queries, as well as a feature that allows creation of custom queries to meet specific reporting requirements. An SRM reporting tool, such as StorageScope, will significantly speed up the process of gathering, manipulating, and presenting the information you need to execute a successful storage reclamation project.

Conclusion

IT departments are increasingly looking for ways to reduce costs and expenditures and to increase efficiency. The storage domain in an IT environment presents an opportunity to increase utilization of existing storage assets, and reduce the need to purchase and support new storage. One specific opportunity is the reclamation of lost or orphaned storage, capacity that appears to be used but is not.

You should develop an overall strategy with specific tactical targets before embarking on a storage reclamation project. You must understand how storage is configured to develop a plan that will require the lowest possible level of effort and yield the highest return on investment.

Finally, an SRM reporting tool, such as ControlCenter StorageScope, will greatly facilitate the tasks of identifying potentially reclaimable storage and monitoring the ongoing reclamation effort.

Ideally, the information presented in this article will allow you to achieve “the good”, reduce “the bad”, and avoid “the ugly” aspects of reclaiming storage.

Biography

Brian Dehn has more than 20 years experience in technical pre- and post-sales consulting, management, training, and business development. He is currently an EMC Technology Business Consultant and a Resource Management Software Champion for the Telco, Media, and Entertainment Division in the United States.

He attended California State University Northridge and earned his Bachelor of Science degree in Business Administration with a focus in Management Information Systems. While at EMC, Brian achieved his Proven Professional Storage Technology Foundations Associate (EMCPA) certification, his Storage Management Technology Architect Specialist (EMCTA) certification, and is now a Storage Management Technology Expert.