Removing Data Waste from Virtualized Storage

Forgotten data objects from virtual machines can clog up virtualized storage. Cleaning up this waste will reclaim storage that can then be reused.

WHITE PAPER BY BRAD BONN AND ALEX ROSEMBLAT

Introduction

Optimizing resource usage in a virtual environment is significantly more challenging than in a physical environment when it comes to efficiency. The ability to create virtual machines rapidly, while a key driver for enhanced agility and return on investment; is also the main cause of this challenge. To further intensify the problem, related data objects such as snapshots or additional VM images are sometimes created for each VM, increasing the amount of storage usage created by virtualization. Unused snapshots, templates, abandoned VM images and Zombie VMs all contribute to wasted CPU, memory, throughput and most importantly, storage resources. Yet locating and reclaiming these resources is not always a simple task. This whitepaper walks through each of the data objects that can create waste and describes how to clean up unused data objects so as to increase data center ROI.

Types of Waste that Occur in Virtual Infrastructure

Abandoned VM Images

When a VM is deleted from VMware vCenter, Microsoft Systems Center, Red Hat Enterprise Management (RHEM) or another VM management console, it must also be deleted from the disk. Otherwise, the VM listing is no longer present in the management console, but the accompanying VM image still exists in storage. If this scenario occurs, the result is an abandoned VM image which resides in storage and consumes space but is no longer in use. Theoretically, proper operating procedures should ensure that each time a system administrator deletes a virtual machine from the management console, that administrator also repeats the same action with the storage array. But this is not always the case.

Due to a variety of reasons, a VM image may not get deleted from storage although it has been deleted from the management console. This can occur when:

- A VMware vMotion storage fails and a file is not completely moved to another datastore. In this case, it is possible that either the old VMDK file will be left where it was, or that the new partially copied VMDK file will be placed on the new datastore. In either case, vCenter will not know about the file. vMotion storage can fail if the new host doesn't have the same configuration as the old one did, or if there wasn't sufficient disk space. Additionally, some users configure vMotion to be completely automated, so unless an error log is checked, it would be difficult to know that a vMotion had failed.
- System administrators manually copy and paste VM images to move them and forget to delete the old file.
- VM images are copied in lieu of using templates and a VM image that isn't necessary is not deleted.
- A third party backup or storage snapshot taking tool is duplicating VM image files. The management console will not know about additional VM image files created in this way.

Finding this Waste:

Detecting abandoned VM images is accomplished through a reconciliation of VMs listed in the management console and the VM images reported within storage. These files are not easy to identify manually since importantly, the VM image name may not always match the VM name. These VM image files are often called orphaned files or orphaned VMs.

Sources of Savings:

Finding and deleting orphaned VMs is important to reclaim and free up storage. Also, deleting these files will liberate software licenses which can be reused.

Powered-Off VMs

Powered-off VMs are just what the name implies: VMs that are powered-off. There is nothing wrong with a VM powered-off unless it is an indication of a VM that is no longer required. The longer a VM has been powered down and not used, the more likely it is no longer needed. It is also possible that a zombie (also known as an idle) VM is identified and rather than being deleted is powered-off to be dealt with later. As this file is no longer in use, it will then become a powered-off VM and still take up storage resources which could be reclaimed.

Finding this Waste:

Reporting on the number of days a VM has been powered-off is the starting point to determining whether a VM is still required. Once the candidate list of powered-off VMs that are no longer needed is identified, some amount of detective work and cross referencing the VM with an inventory report will be required prior to deleting the file.

The key to detecting powered-off virtual machines that are no longer used is the ability to exclude a VM from the analysis going forward if a VM administrator determines that the powered-off image is needed. Otherwise, any reporting mechanism on powered-off VMs will eventually become cluttered with long unused powered-off VMs obscuring the next level of detail and hiding powered-off VMs that need to be deleted.

Sources of Savings:

Deleting powered-off VMs is important to not only free up storage space, but also to release any software licenses that the VM is taking up.

Unused Snapshots

Snapshots are a state of a virtual machine at a particular time and are used for backup and recovery purposes. A snapshot is similar to a desktop recover point. VM administrators typically will make a snapshot of a VM image prior to updating or changing a particular VM to provide for rollback should the upgrade not go according to plan.

Once the patch has been successfully applied, system administrators are supposed to incorporate the snapshot back into the VM configuration which will remove the snapshot and make the changes permanent.

Theoretically, this sounds easy. But there are a few key issues with snapshots that can make managing snapshots particularly troublesome for administrators:

- Snapshots can take up all the available storage without an administrator knowing it.
- Some environments have multiple snapshots for a particular VM

- Finding what snapshots are in storage, when they were made, and if they are still being used requires additional work and tools to report on.
- Snapshots for a VM that has been deleted may remain, and will be difficult to cross-reference to the VM that no longer exists.
- The SAN can take its own snapshots with its software.
- System administrators simply forget about snapshots that were taken, or the snapshot was taken by another user and not reported.

Finding this Waste:

The key to finding unused snapshots is to look at the age of the snapshot and then remove the older snapshots for a particular VM when it has been deemed that that snapshot is out of date.

Sources of Savings:

Deleting unused snapshots will free up storage space.

Unused Template Images

A template is a base image that is used to quickly create virtual machines that are identical. Templates drive compliance for operating system images, patch levels and installed software. Each time the operating system or application changes, a new template needs to be created. If a template that is out of date is not deleted, it will consume storage resources unnecessarily. Unused template images can become a significant source of wasted storage.

Finding this Waste:

The key to finding unused templates is to look at the age of the template and then check to see that the template is still valid.

Sources of Savings:

Deleting unused templates will free up storage space.

Zombie (also known as Idle) VMs

Zombie VMs are virtual machines that are still running but have reached the end of their production lifecycle. This is most likely to happen in volume in environments where:

- VMs are created by end users themselves.
- Where a communication loop has not been closed between an end user and a system administrator and the end user has not informed the system administrator that a particular VM is no longer being used.

- QA or development teams can spin up VMs on their own without central administration oversight.
- End customers can deploy VMs automatically in cloud initiatives.
- Applications running on a VM are offline and the application owner has not yet noted that the applications are down.

Finding this Waste:

Zombie VMs are tough to detect. They are powered on and appear to have a load on them. However, as a Zombie, the load may be low. But even low load is not a reliable way to close in on a potential Zombie VM. The deviation of the load over time is the best way to separate a Zombie, for example, from a DNS server that simply rarely has a load on it.

Once the list of potential Zombie VMs is identified, cross referencing the VM information with inventory information is the next step to determine if a VM is truly a Zombie or is doing useful work. VMs that appear to be Zombies but are not should then be tagged as such to prevent them from being identified over and over again as potential Zombies. Importantly, if the Zombie VM is simply powered-off instead of being deleted, it will still take up storage resources, and will be noted as a powered-off VM.

Sources of Savings:

Cleaning up Zombies frees up CPU, memory, throughput, and storage for reuse. In addition, each operating Zombie VM consumes licenses for operating systems and other software that could be used elsewhere.

Conclusion

Standard operation of a data center can generate waste in virtual environments. Even when procedures to clean up wasted data objects are put in place, there are use cases where such waste may still occur or the procedures may not always be fully followed. Instead, by instituting periodic waste clean-up initiatives through an environment, data center managers and system administrators can actively search for wasted data objects and throw them out when they are found. Establishing these procedures is important to maintaining the highest possible return on investment for virtual environments.