CirrusData
Powered By Innovation

www.cdsi.us.com

Storage Migration for Virtual/Cloud Environments

April 2015

Executive Summary

As virtualized and cloud system platforms have become prevalent in the IT world, the need to migrate data from legacy storage to a virtual environment or to the Cloud is often necessary. This paper explores the pros and cons of various tools for data migration in the virtual/cloud environments, using VMWare as an example of popular hypervisors in the market. The Data Migration Server (DMS) is then introduced to highlight the specific features related to the virtual/cloud environments that make it the ideal tool for all combinations of Physical/Virtual to/from Physical/Virtual environments with minimum impact to existing production.

Reasons To Migrate Data

First let's put the term "cloud" into the right context. While "Cloud" means many things to many people, here we generally refer to a large scale virtualized data center where there are many clusters of VMWare or other hypervisor host servers connected to the SAN storage farm with petabytes of data. The environment may be part of the local data center for an enterprise-class company (private cloud), or a shared data center managed by a Cloud provider (public cloud). So in this context, the cloud environment is essentially the same as a virtual environment but at a larger scale.

Generally, the virtualized application servers and the virtualized storage are logically the same as in a physical environment: a virtualized CPU with memory, network, and block storage in the form of a VMDK disk file or a raw LUN (RDM). There are many reasons for migrating data within the virtual/cloud environment. The following is a short list:

- Storage refresh: the existing storage frame is at the end of its useful life or the lease is up. Tens or hundreds of terabytes need to be moved from the old to the new storage.

- Physical to Virtual Conversion: the original transformation from legacy physical environment to virtual environment may have been simply to map the existing storage LUNs verbatim using a mapping file, thereby saving a lot of hassle in copying a large amount of data into the virtual storage. However, keeping the original physical data in the original place negates the many benefits of virtual storage. As a second phase of virtualization, there is a need to migrate the raw/direct mapped storage into VMDK disk files stored in the datastore of the virtual environment (VMFS volumes). It would be nice to perform this kind of conversion with a Storage Refresh project so that while migrating to new storage, the Physical to Virtual Conversion is also done, killing two birds with one stone.
- Consolidation of storage: Moving data from one storage to another in order to consolidate the storage space. This is essentially the same as the Storage Refresh case but perhaps at a smaller scale.
- Performance tuning: Moving data from one storage to another in order to balance the distribution of active data onto separate storage spindles or controllers. This is essentially the same as the Storage Refresh case but at a smaller scale.

So essentially there are two basic use cases for migration: within a virtual environment/cloud for storage refresh or other needs where the source and destination are both virtual, or from a legacy physical environment to the virtual environment/cloud. In either case, the migration can be local (within the same datacenter's SAN) or remote (the destination is at a different location connected only by WAN with TCP/IP).

With worldwide Cloud initiatives gaining momentum, and with the amount of data growing at a rapid rate, the need to migrate data from one place to another is also growing. According to Gartner, business data continues to grow at 40%-60% rate, and the Computing Technology Industry Association issued a report detailing the results of its cloud computing survey. In the report's description of the cloud market, CompTIA also cited Gartner's prediction that "cloud storage will grow at 89.5% CAGR to $2.88 billion" through 2015.

Unfortunately, the downtime available for migration is simply becoming non-existent. The allowable impact to production while migration is being executed is also getting smaller and smaller. In these cases, the customer may become victim of high cost lease-extensions, or may suffer the inability to enjoy the newly-acquired faster, better storage.

Limitations of Native Data Migration Tools

There are many existing data migration tools available in the virtual environment, such as VMWare's Storage vMotion. Storage vMotion is simple, intuitive to use, and generally does not require any downtime. The only concerns are that it may have a significant performance impact on virtual machines where a large amount of data is to be migrated, and that it does not work if the destination storage is on a different SAN or a remote data center, making it unsuitable for data center move.

For a large scale migration project where essentially every datastore is moved from old storage to the new and if the mandate is to complete the project with a tight deadline, it requires invoking the maximum number of Storage vMotion sessions (see limitations below) to move as much data as possible in order to meet the deadline. The massive copying process places such a load on the virtual environment that the performance of the VMs may be compromised, and therefore it is often necessary to allocate application downtime in order to avoid issues. The main goal of virtualization is to make better use of the CPU/Memory/Storage bandwidth, and therefore a well-designed virtualized system SHOULD not have a lot of extra CPU/Memory/Storage bandwidth available for the massive copying of data. For a host platform with 20 VM and each with 5 to 10 disks, there might be hundreds of virtual disks in multiple datastores that needs to be moved. Starting Storage vMotion on all the disks at the same time will surely cause unacceptable impact to production. The migration project now has to be conducted either with schedule downtime, or by dividing it into a carefully designed multi-phase project, each phase moving a reasonable number of VMs.

Besides performance impact considerations, there are other limitations that makes it not possible to use Storage vMotion in a massive scale. The following limitations and requirements are stated in VMWare's technical references:

- Virtual machine disks must be in persistent mode or be raw device mappings (RDMs). For virtual compatibility mode RDMs, you can migrate the mapping file or convert to thick-provisioned or thin-provisioned disks during migration as long as the destination is not an NFS datastore. If you convert the mapping file, a new virtual disk is created and the contents of the mapped LUN are copied to this disk. For physical compatibility mode RDMs, you can migrate the mapping file only.
- The host on which the virtual machine is running must have a license that includes Storage vMotion.
- ESX/ESXi 3.5 hosts must be licensed and configured for vMotion. ESX/ESXi 4.0 and later hosts do not require vMotion configuration in order to perform migration with Storage vMotion.
- The host on which the virtual machine is running must have access to both the source and target datastores.
- A particular host can be involved in up to two migrations with Storage vMotion at one time.
- A maximum of eight simultaneous Storage vMotion accesses to a single VMFS3 datastore, and a maximum of four simultaneous Storage vMotion accesses to a single VMFS2 or NFS datastore.

Due to these limits, data migration is now no longer a simple exercise of copying data from A to B using Storage vMotion, even for a simple Storage Refresh project. Complying with these requirements (if possible) now demands a more complex, multi-phase project plan. Even then, only a small number of migrations can be performed at a time due to the limits of 8 or 4 accesses. This means the project may stretch out for weeks or months.

For a more complex use case, such as migration from legacy, physical application server clusters to the cloud/virtual environment, the native tools such as Storage vMotion are unable to address the source disks which are physical LUNs presented to physical application servers on the SAN. In this case, not only does the migration tool have to be able to read/access the source LUNs across different host server platforms (Linux, Windows, UNIX, etc.), but the tool must then be able to write to the new virtual storage. Lacking such a tool, the migration process may need to be a two phase project, first with a P-to-P copy, resulting in a RDM in the virtual environment, followed by a P-to-V process at the tail end, making the migration process even more complicated. An ideal solution is a migration process that can directly intercept the source storage (preferably live and without any downtime), and then online copy the source data into the virtual environment as VMDK files (or RDM if so desired). This will result in minimum downtime (needed only after migration is complete and cutting over production to the virtual environment), and lowest possible impact to the virtual environment (all the copying is performed by the migration appliance at the source).

A Fresh Approach: CDS DMS

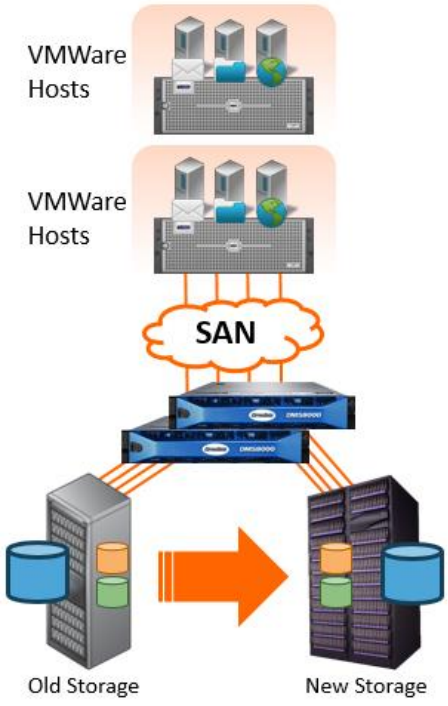
The Cirrus Data Solutions (CDS) Data Migration Server (DMS) Appliance is the industry's first disk-block level data migration solution designed from ground up for the purpose of efficiently, transparently and safely migrating data from any legacy or Cloud/Virtual environment to any other legacy or Cloud/virtual environment. Other tools such as Backup/Restore products, Data Mirroring products, Data Replication (Disaster Recovery) products, or native data mover tools can be used for migration. However, most of them either lack features that are specific to data migration into the virtual environment, or cause too much downtime or too great an application impact when migrating a large number of disks. The CDS DMS appliance is designed to perform migration on a massive scale (hundreds of LUNs and hundreds of terabytes) in such a way that zero extra load is added to the physical hosts or hypervisor hosts. With the availability of a

Virtual DMS appliance running native to hypervisor platforms, it is possible to include P-to-V as part of the migration process, even for long distance migration to remote datacenters over WAN. The DMS is a 100% transparent device that requires no changes to the existing environment. It is built on the patented Transparent Data Interception technology (US Patent No. 8,255,538) and when inserting it into the existing production system (legacy or Cloud/virtual), the process takes only 5 seconds per FC path, and does not require any reconfiguration….period.

Use Cases

The following use cases highlight the power of CDS DMS appliances and DMS-9000V virtual appliances.

Case 1: Cloud/virtual to Cloud/virtual migration



Use case 1 typifies a storage refresh project where a number of storage frames are at end of life or end of lease and need to be migrated to a new set of storage frames, and possibly to a different brand of storage. Because all datastore volumes must be moved and will affect hundreds of virtual disk files, and because of the limitations stated above, let's assume that using Storage vMotion or similar tools within the virtual environment is not feasible. A low-impact tool that can offload all the heavy lifting, such as the CDS DMS appliance, shall be used to massively migrate all disks without doing any work at the virtual environment and without any downtime until cutting over to the new storage.

Using DMS for this use case is quite easy. Simply insert the FC ports found on the DMS Appliance, one path at a time into the FC fabric of the source storage, zone the new storage to be accessible by the DMS, after which data migration jobs can be defined and executed.

This means that although the DMS is an appliance on the FC SAN, IT managers or Data Migration Service providers can avoid all the pain-points that are usually associated with deploying appliances on the SAN, such as certification of host drivers, changing of multi-path software, modification of FC zones and changing of LUN masking/presentations at the storage controllers.

The elimination of the above work tasks, the potential downtime, and all the risks associated with making changes translates to a significant amount of cost savings and great improvement on ROI.

During the massive copying of the datastore LUNs, DMS has positive impact control. For example, using the most nonintrusive throttle setting, DMS monitors the queue depth of each LUN being migrated, and will automatically suspend migration when there is I/O from the host. This way, DMS is using only those small windows of idle time (unused storage bandwidth) to copy data. Because the copying is done using CPU/Memory/Bus bandwidth of the DMS, the hypervisor hardware is completely out of the picture and zero load is added, thereby guaranteeing that production performance is preserved.

When the copying is complete, DMS enters a holding pattern to await for cut-over. Up to this point, no downtime is required. At the time of cut-over, a small amount of downtime is necessary to move the old datastores offline, rescan for the new disks, and bring up the datastores.

Case 2: Legacy to Cloud migration, Local or Remote

For use case 2, the source storage is part of a Fibre Channel SAN with physical application servers. The objective is to transform this legacy system to a virtual environment (private cloud) or to migrate to a hosted cloud environment (public cloud). The architecture of the application system after migration is identical to the original physical environment, except that each application server is now a virtual machine in the virtual/cloud, and each disk that was assigned to the host from the SAN is now a virtual disk in a datastore of the hypervisor. The destination may be a local or remote datacenter.

The migration from the Physical environment to the virtual/cloud environment is accomplished essentially by copying all the physical disks into the virtual environment. The destination virtual environment typically has virtual machines matching the source physical machines configured following the best practices prescribed by the application vendor (Oracle, SQL, Exchange…) For the migration project, the focus is on moving the data disks from the physical environment to the virtual. In this case, the native tool like Storage vMotion will not be able to perform the storage migration into the virtual environment. The migration tool for this case must be able to read the source disks, desirably while they are still in production, and at the same time, be able to write to a virtual disk that is part of the destination datastore (VMFS or NFS) that resides in the local datacenter SAN, or is at a remote datacenter in the case of a datacenter move to the cloud.

Due to the fact that the DMS-9000V is a virtual machine running natively in the virtual environment, it is quite simple to use the CDS DMS solution for this migration case, as it consists of both physical migration appliances (to intercept the Source disks without downtime) and virtual appliances (to write to virtual disks of the VMWare datastore in either VMFS or NFS volumes). Because the delivery of the source data to the DMS-9000V is through standard TCP/IP connections and since as many DMS-9000V's as necessary can be deployed in order to scale the performance requirements and to satisfy datastore access requirements, this

combined physical/virtual migration server appliances solution is ideal for the Physical-to-Virtual (Legacy-to-cloud) migration projects, both for local (private cloud) or remote (private or public cloud) datacenters.

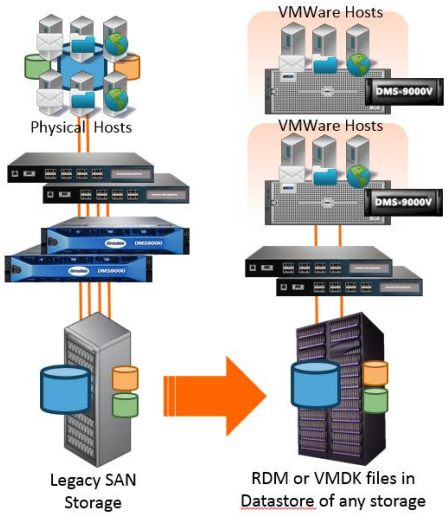


Figure 1

Figure 1 depicts a solution design for Legacy to Virtual (Private Cloud) migration within a local datacenter, where the source "Legacy SAN Storage" is intercepted live by physical DMS appliances. Data is read from the source and pushed into the virtual environment where DMS-9000V virtual appliances are deployed. The DMS-9000V virtual appliances receive the data, and write to the virtual disks assigned to them by VMWare. The virtual disks can be RDM, or VMDK disk files of NFS or VMFS datastore volumes.

Figure 2 depicts a solution design for remote migration from Legacy to Cloud datacenter. Because the connection between the physical DMS appliances and the virtual DMS appliances is over the WAN, the optional Encryption and Compression features of DMS may be turned on to secure the connection and to reduce bandwidth requirements by up to 5 times, assuming a typical 5:1 compression ratio.

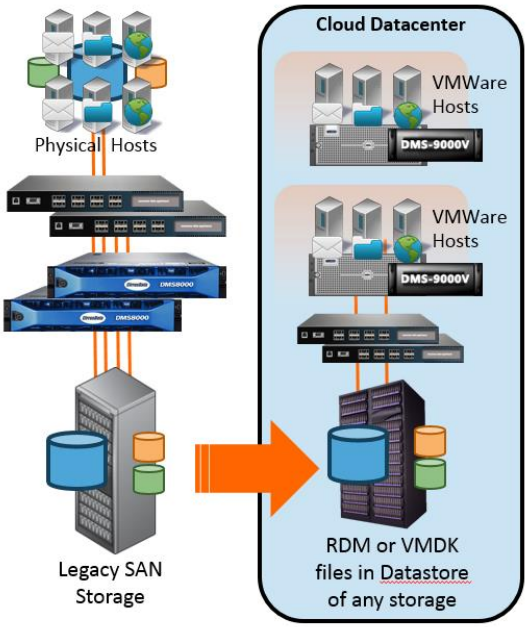


Figure 2

Summary

There are a number of reasons organizations migrate their data; to upgrade to new storage, to move to a new location, to consolidate and/or to reorganize data. Many tools are available to help achieve this goal provided there is enough downtime to work with. But when the migration needs to be done in a production environment, there is usually very little downtime available, and the cost associated with this downtime for a company can range from $90K to millions of dollars per hour depending on the organization.

How Cirrus Data Solutions Can Help

Cirrus Data Solutions, Inc. (CDS) developed the Data Migration Server (DMS) which is the industry's first block level data migration solution designed from ground up for the purpose of efficiently, transparently and safely migrating data. Without the need to make switch zoning changes, to modify LUN masking, or to change host drivers, the DMS appliance provides a solution for companies to migrate their production data without any downtime throughout the migration process. A few minutes of cut-over downtime may be required (after data is migrated successfully) at the user's convenience. The DMS saves time and money by eliminating work and risks associated with the work that is typically required by other solutions. The availability of the DMS as natively running virtual appliances allows the streamlining of P-to-V conversion as part of the migration process, making the DMS physical/virtual appliance combination an ideal solution for Legacy-to-Cloud (Physical to Virtual), or Cloud-to-Cloud, or even Cloud-to-Legacy migrations.

Cirrus Data USA
6800 Jericho Turnpike
Suite 213W
Syosset, NY 11791

Cirrus Data EMEA
2 Place de Touraine
78000 Versailles, France

Cirrus Data APAC
739 Sheng An Da Dao
Jiangning Qu, Nanjing Shi,
Jiangsu Sheng, China 211178

1.516.427.5210
info@cdsi.us.com
www.cdsi.us.com

Copyright © 2015 Cirrus Data Solutions, Inc. All Rights Reserved

The information contained in this whitepaper is the exclusive property of Cirrus Data Solutions, Inc. (CDS). This document may not be used for any purposes, reproduced or transmitted in any way, in whole or in part, without prior authorization in writing from CDS. The information contained within is provided as-is for informational purposes only and is subject to change any time without prior notice. CDS specifically disclaims all warranties, express or implied, including but not limited to warranties of merchantability and fitness for a particular usage.