RECOVERPOINT INSIDE VMWARE ENVIRONMENTS

Mohamed Gombolaty
Remote Technical Support Engineer
EMC
Mohamed.gombolaty@emc.com

EMC²

Table of Contents

Introduction

In July 2014, Gartner released its x86 Virtualized Server Magic Quadrant which estimated 70% of x86 servers are now virtualized, with VMware the leading vendor (https://www.gartner.com/doc/2788024?srcId=1-2819006590&pcp=itg). By now, most of you have a virtualized environment and most probably you are using VMware.

But with virtualization comes new challenges. While basic IT operations, such as monitoring, backup, and replication still need to be performed, they need to be done in a different manner to accommodate virtualization concepts and operations.

Replicating a virtual environment has been quite a challenge. In our experience we have found many common challenges. First, let's re-visit why replication is important. The notes below were inspired by the following link which discusses the need to spend on replication technology for virtual environments (http://www.slideshare.net/rackspace/vm-replication-webinar):

> 70% of reported DC outages are directly attributable to human error." *Source: Uptime Institute, Data Center Site Infrastructure Tier Standard: Operational Sustainability, 2010*. No recent reports on if that number has changed with automation being used, but for example a complete French Government system was down for four days due to a sub-contractor accidently triggering the extinguishing system (http://www.computerworlduk.com/news/public-sector/3454451/data-centre-outage-takes-out-french-government-payment-system-for-four-days/)
>
> Though many believe that backup alone will make them safe, it didn't happen the in French government example. Using replication will minimize downtime even if backups will work, the easier and more resilient your replication tool is will save money lost with every second of downtime. A Gartner blog discussion states an average of $5,600 per/minute (http://blogs.gartner.com/andrew-lerner/2014/07/16/the-cost-of-downtime/). You can calculate your own cost following Gartner Toolkit: Downtime Cost Calculator for Data Center Disaster Recovery Planning (Robert Naegle) (http://www.gartner.com/document/2674021)

When working with disasters or outages you need to be aware of two major concepts:

RPO (Recovery Point of Objective): RPO simply means the data you will bring back up whether on the Source or DR side, how far back will it be, i.e. will it be 15 minutes before a disaster occurred or 30 minutes, an hour, or more. The further back, the more data loss you will face, which will add up to your downtime cost. Thus, ensure you have lowest RPO possible.

RTO (Recovery Time Objective): this indicator is how much time you need to bring production servers back up serving customers from the data you replicated or backed up. Replication is much faster since they tend to update DR resources with the latest unlike Backups which will need more time to restore data on disks.

Those two combined are your compass when in downtime, disaster recovery, or corruption recovery situations. The lower both are, the more efficient you are in handling outages.

Picture from community.emc.com.

Now, let's move to the challenges facing VM replication. You need to consider the following points:

Complexity: You certainly have a lot to consider. Since these VMs share ESX servers, you must be very keen on deciding to do it on a VM (normally called Guest Level) or on an ESX level and weight the overhead of backup or replication on ESX resources. Additionally, you must be sure to replicate all data needed to bring VM up correctly on DR or to restore it. For example, you can

replicate the correct LUNs but still fail to bring the VM up because you didn't replicate or back up the snapshot data directory. Another consideration is how frequently can you replicate or backup and also test the solution without downtime or affecting production. All these questions require answers.

Scalability: Once you have decided how to replicate, how much can it scale as your environment grows? You need to know how long you can survive with your infrastructure until you need to add more investment. Certainly, the longer the better. Again, this depends on the replication method and technology you choose.

Meeting Expectations: Since you invested in replication, you are expected to achieve RPO and RTO. Thus, you need to keep making sure you can meet these requirements from the technology you used, or from frequent testing. This will help ensure that your staff and your replication technology can achieve targets, and be notified if any part still needs to be fine-tuned. It's worth mentioning that Disaster Recovery is not about software and hardware only; it is more of a Holistic operation that involves people, procedures, and technology. You need to ensure you can meet targets whether it be for performance overall in the process of actual failing over or recovery or also when not in need for a failover or recovery.

Cost: A solution that offers RPO of zero and RTO of zero as well will certainly be expensive and you might not be able to afford it. Even if you can, your RTO needs may not require zero time. Maybe 15 minutes would be the same and an RPO of 5 minutess will also be OK with management. Thus, you may need to think about cost vs RPO and RTO combined and increase it to lower numbers later from an RPO point of view. Start simply with a solution that can grow with you.

The remainder of this article will showcase how RecoverPoint as a replication technology works inside your VMware environment, enables you to achieve the best RPO and RTO times, scales as your environment grows, and ensures you will meet your replication targets and test as frequently as you want.

Why RecoverPoint is smart Replication

RecoverPoint's replication technology has a lot of unique features that make it a valuable resource in a data center. RP4VM, released in November 2014, is a tool aware of its virtual existence and provides administrators with remarkable power options.

Let's examine and explain RecoverPoint piece by piece.

Starting as software installed on a specifically designed U-Server, RecoverPoint can now be deployed as an OVF. What sets it apart is the code and theory of operation; here are some points that make RecoverPoint an efficient replication tool:

Block-Level Replication

RecoverPoint requires either FC or ISCSI access to storage, depending on topology used. Because RecoverPoint works on a block level it only understands 0s and 1s, which means that RecoverPoint need not understand your OS or application. Whatever it is, RecoverPoint will be able to provide a crash-consistent image of the LUN at a point in time you desire, so it's a unified replication for any OS or File system type. Thus, Windows, AIX, Solaris, and all Linux flavors can be replicated with the same replication tool and provide a crash-consistent image which means this is the LUN with the exact 0s and 1s at the point in time you choose. Application level might not be consistent, especially with databases. This is because to consider an image consistent, a DB normally expects writes to database data and redo log data is done. However, your image might have the write completed to the data database but not to the redo-log. Thus, database conceives it as inconsistent. Now those errors can be resolved from the database itself. Probably, it will discard transactions not found in redo-log. A DB Admin can fix these errors or you can search from the wealth of images you can choose from until you find an image application that is consistent. Normally, these images will be the lowest size when you search the images list. If you require application-consistent images, you will find tools and procedures to periodically create bookmarks on application consistent images. They are very quick and don't require putting your DB on hot-back up mode for long. If it is scripted well, it can take an average of 2 minutes tops in large environments.

Eventually, you will have a replication tool for any type of OS, and with simple tweaking, make it application consistent if necessary, though with the next point you might not consider even trying to fine tune.

Any Point in Time

RecoverPoint depends on having a mechanism in the environment – called a splitter – which has one function; to see packets moving between host and storage related to LUNs that RP has been configured to replicate. The reason that we care about packets with write SCSI codes is because these packets will make a change on the LUN they are destined for. Thus RecoverPoint must replicate that packet as well, so the splitter copies that write packet and makes the RP server the destination for the new copy packet. To ensure consistency, the host will not receive a write confirmation until the splitter receives confirmation from both the Storage and RecoverPoint appliance. This is because if storage fails or RecoverPoint for any reason there will be no consistency between Source LUN and Replica LUN.

While this splitter technology might seem intrusive, it's not. It only cares about writes. Only reads are unharmed and reads are what most of I/O load is about. There is no delay or higher latency. we have RecoverPoint working in data centers with large, complicated I/O loads and the splitter is not adding overhead to the process.

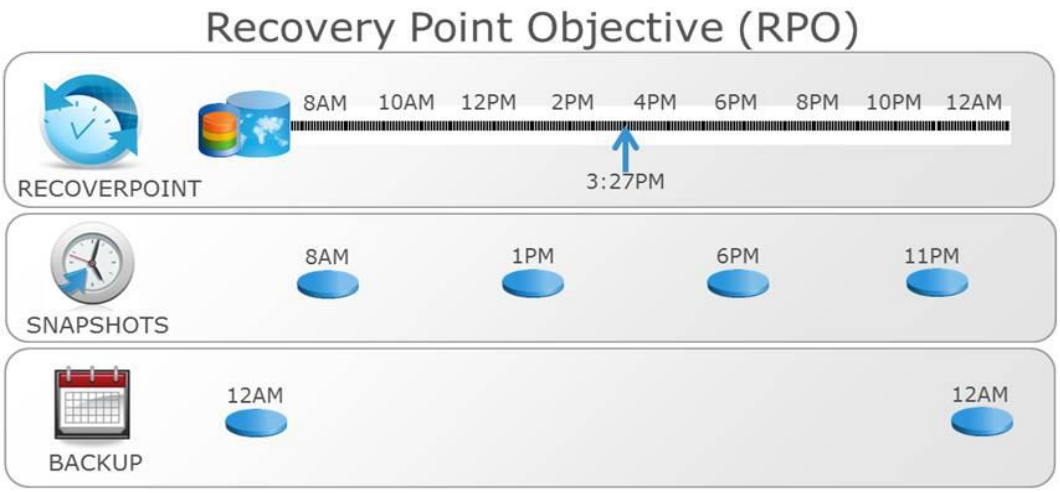
A splitter now can be residing in one of the following:

- Storage: VNX® and VMAX® and VNXe® all have RecoverPoint splitters embedded and only need an active license, suitable when Source LUNs are using EMC Storage.
- VPLEX®: if Source are on non-EMC storage, you can still replicate them using VPLEX as a splitter. VPLEX will mask the LUNs from third-party storage through it and do the splitting for RecoverPoint.
- ESX Splitter: Now in 5.5 Update 1, splitter software can be installed on ESX. This can work on any storage as long as traffic passes by the ESX servers it is installed on.

You can see multiple options according to your specific topology, and you can mix different splitters in the same site as well.

Perhaps the most Important value gained from the splitter is the ability to make images as the writes come. Even down to each write, an image can be made of it (Sync Replication) or group a number of writes into a single image (size or time). All of these attributes can be controlled and specified if you wish. This is how detailed options empowers a customer to control his replication RPO.

Many people describe the images they see on RecoverPoint like a Tevo TV backward option. You can go back second by second or write by write depending on replication you choose. This is the most trusted RPO decrease you can ever get or control.



Estimated Protected Period

The difference between replication and backup is needed to keep images in your possession. Backups are mainly for long durations – months, maybe even years. Meanwhile, replication is short term – a day or a week. To bypass any sudden production downtimes or performance or corruption, you need to know how much time your replication can hold data for a LUN or a group of replicated LUNs. RecoverPoint Appliances can estimate how much based on Incoming writes rate. This is possible because RecoverPoint uses Journal LUNs dedicated to keeping the images. Based on capacity of the journal with the incoming writes rate, RecoverPoint can quickly calculate the predicted or estimated period it can keep, constantly changing with the incoming I/O rate. If you need more time, you can add more Journals.

Consistency Groups and Group Sets

Logically, you always have a relationship between a number of LUNs or disks. For example, three disks consist or contain database information data and configuration, so logically you need your replication tool to be aware of that, and even accommodate having them replicated together and DR access to be done for them all at the same time. RecoverPoint achieves this by configuring LUNs into what resembles a Container named Consistency Group (CG). Each CG contains any number of replication sets, and a replication set consists of a source LUN and replica luns data.

RecoverPoint will replicate any number of replication sets configured under one consistency and make sure that write fidelity order is maintained across them. This ensures a crash consistent image across all these LUNs at that specific point of time.

Not only can bookmarks of images be taken at the same time across a number of different CGs, being able to have an exact image across a number of CGs can be beneficial if you want to subdivide LUNs in different CGs but need also to have a consistent point of time across them. The frequency starts from 30 seconds and above.

Replicating over IP and FC, locally and/or remotely

RecoverPoint can have local replication and/or remote replication. A Consistency Group is not limited to just one of either local or remote. It can replicate locally and remotely at the same time, and the remote copy can be more than 1. In RecoverPoint version 4.0 onward you can have one source and four replicas maximum, and those 4 can be in different sites as well. Additionally, a RecoverPoint System now can have from one to five sites.

You can replicate over IP or FC depending on RecoverPoint Appliances and topology. When using IP replication, RecoverPoint can facilitate compression and deduplication over WAN links between sites.

API and Scripting

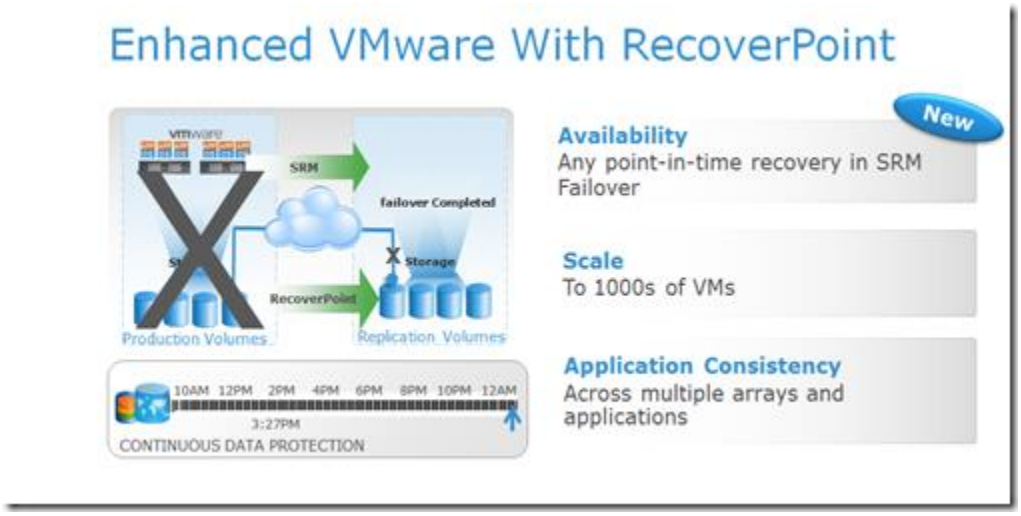
RecoverPoint operations can be managed and controlled by creating your own scripts or third-party tools such as VMWare Site Recovery Management (SRM) or EMC Replication Manager; this will help automate activities to eliminate accidental errors or

changes. Any Administrator with scripting knowledge can quickly write a script to enable image access or to disable it or to failover.

RecoverPoint and VMware

Virtual Environments save costs but as they grow they can become a complicated, large intertwining tree of components. Consequently, identifying all requirements or needs might not be difficult and time consuming. Certainly, identifying LUNs for replication will not be an easy task. RecoverPoint, with its many advantages as a replication technology and platform wanted to be more VMware friendly and easy to use.

The major RecoverPoint advantage with VMware is the ease of identifying and configuring replication for VMs, in any topology, enabling you to identify which VMs are totally protected, half protected, and not protected. VMs can be replicated by a click of a button.



RecoverPoint topologies in VMware

RecoverPoint offers three flavors of deployment to replicate VMware environments. Let's examine the difference and benefit of each.

Classic Physical RP deployment

RecoverPoint started as a Physical U Server-sized machine with EMC RecoverPoint proprietary software installed on it. This is the classic topology, requiring a minimum of two RPAs at each site for redundancy reaching up to eight RPAs in one site for one RercoverPoint Cluster in the site. The physical

Hardware of RPAs can only use FC connectivity to Storage, so such a topology can be used if Storage is EMC or if VPLEX exists and can be utilized for splitting.

The main advantage of this topology is that you don't have to waste resources on the virtual environment since replication of RecoverPoint platform has its own dedicated CPU, memory, and network and FC interfaces.

Virtual RecoverPoint Deployment

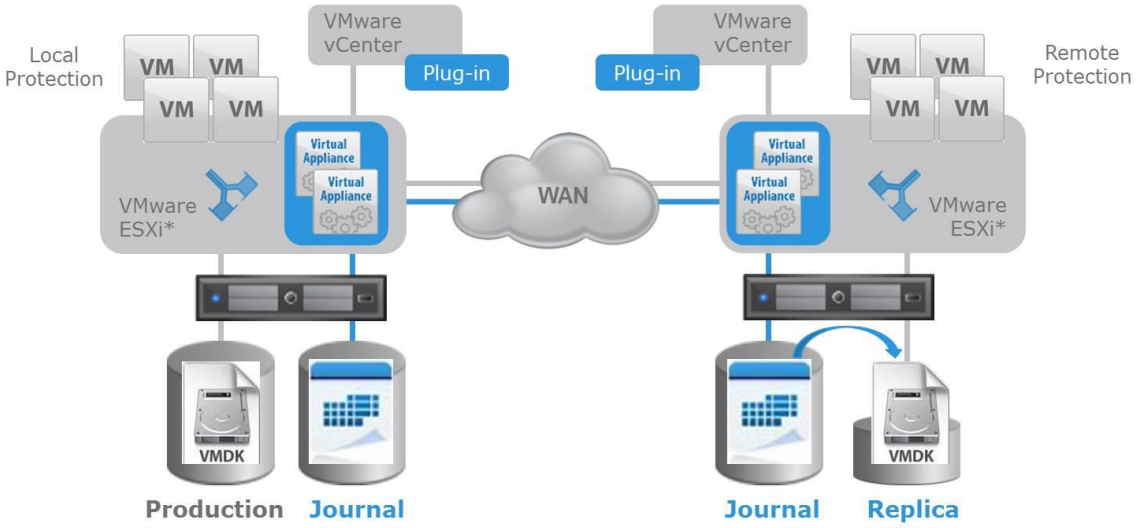
This started in RecoverPoint version 4.0 where you don't need to get physical servers. In this, you deploy OVF in your Virtual environment and you will have Virtual RecoverPoint Appliances as VMs. in this virtual topology, RecoverPoint will access Storage using ISCSI with EMC Storage and can also work with non-EMC storage via VPLEX splitter.

The main advantage of this topology is that you use your virtual environment resource though OVF reserve resources always, so no more extra power and using options of VMs for RPAs as well.

RecoverPoint for VM

RecoverPoint version 4.2, released at the end of Q4 2014, is based on Virtual RecoverPoint Deployment but it uses and only works with ESX splitter installed on ESX servers. This is totally customized for VMWare specifically, meaning in the first two topologies you need to configure a CG and specify LUNs you want to replicate and their replicas. However, in the new release you just need to choose a VM and it will configure the CG automatically and will also create VM in DR site and power it on as well.

The main advantage of this is that you are using your own virtual resources and easier functionality specific to and integrated with VMware.



* RecoverPoint for VMs vApps can be installed on the same or different ESXi servers as the protected VMs.

RecoverPoint for VMs significantly lowers CDP and DR infrastructure costs with embedded I/O splitters into vSphere, RecoverPoint vApps installed on existing ESXi servers, integrated management thru vCenter, and support for any type of storage.

Three deployments and the differences

Element	Physical Deployment	Virtual Deployment	RecoverPoint for VM
Connectivity	FC to storage	ISCSI to storage	ISCSI to storage
Splitters supported	EMC Storage splitters (VMAX and VNX) and VPLEX	VNX Splitter	ESX Splitter
Number of RPAs per site	8 Physical RPAs	8 VMs	8 VMs
Resources	Physically outside Virtual environment and dedicated	Supplied from Virtual environment and dedicated	Supplied from Virtual Environment and dedicated
Replication Types	WAN and FC	WAN	WAN
Number Of Copies	4	4	1

Design Considerations

While Installation and Deployment Guides are available for each topology, here we discuss some pre-requisites you need to design and take into consideration to ensure a stable replication environment:

Network

For any physical RPA there are two IPs to be configured; a LAN IP for management and GUI access purposes, and a WAN IP for replication over a WAN. If using FC replication, you still need to add a WAN and connect RPAs back-to-back on their Ethernet of WAN to avoid messages regarding WAN link is down. Thus, depending on the number of RPAs you will have per site the equation will be "(Number of RPAs X 2) + 1". The addition of one extra IP is for LAN IPs and this is a cluster IP. The cluster IP will be held only by RPA controlling RP cluster and there are only 2 RPAs that can control a cluster, RPA1 or RPA2.

For vRPAs, each one will need four IPs, two that are similar to physical RPAs WAN and LAN, along with one more IP for cluster on the LAN network. The extra two IPs are for ISCSI interfaces on the V. Those ISCSI interfaces are for communication between vRPAs and storage, so the equation will be "(Number of RPAs X 4) +1".

Another concern is it's better to have each interface on the RPA on a different subnet, i.e. WAN is a different subnet than LAN, and also each ISCSI interface be on a different subnet. This will help mitigate routing complexity on RPAs and avoid any network issue affecting an interface the affect other Ethernet Interfaces on the RPA.

On the WAN Interfaces, if you are using a Maximum Transmission Unit (MTU) size other than the default of 1500, you need to set it correctly while installing RPAs. MTU on WAN can differ due to a number of reasons, the most important being re-occurrence of the usage of IP Security Tunnels. You need to verify if MTU on WAN link is default or not. If WAN and LAN exist on the same subnet you will need to set both interfaces to avoid communications problems that can disrupt installation and cause it to fail. If changed after installation, replication will not work.

For ISCSI networks we suggest using Jumbo frames which have a MTU size of 9000 between RPA and storage through the entire path to have healthy I/O over ISCSI between RPAs and storage. Another factor for ISCSI is the bandwidth you will

expose to vRPAs via the vSwitches. The number of port NICs or VLAN tagging among multiple vNICs you provide RPAs will affect performance of vRPAs

Resources

The Physical GEN 5 RPAs have the following resources on them :

Item	GEN 5
Hardware	Intel R1000 (Kylin)
Form factor	1 U
CPU	Sandy Bridge
CPU Speed	1.8 GHz
Number of Cores	Quad Core
Number of CPUs	2
Memory	16 GB
HDD	2 x 300 GB 10K RPM 2.5 SAS
HBA Type	8 GB quad HBA
IP Ports	6 x 1 GE ports (RJ-45) for WAN, LAN & Remote management + 3 ports are unused

vRPAs resources needed are shown below from Simple Support Matrix for RecoverPoint 4.1 for Virtualized RPAs:

	CPU	Memory	ISCSI Adapter
Minimum	2 CPUs	4 GB	1 G
Maximum	8 CPUs	8 GB	10 G

RecoverPoint for VM has 3 profiles as described in Installation and Deployment document for 4.2 :

RPA Profile Feature	2xCPU/4GB	4xCPU/4GB	8xCPU/8GB
Virtual CPUs	2	4	8
RAM	4 GB	4 GB	8 GB

A vRPA best practice is to reserve all RAM for RecoverPoint. Plan your resources according to the Topology you will use.

Incoming writes

A crucial part of sizing environment that number of RPAs and WAN size required to replicate all depends on Size of Source LUNs and Incoming write changes which is number of writes sent to storage, otherwise known as Change ratio.

The more time you spend to get averages of Incoming writes, including times where high peaks of change occur, the better off you are in identifying the needs of your environment by number of RPAs and WAN size.

Journal Sizes

As discussed earlier, Journal is where RP stores information needed to move backward or forward with images, saves writes until they are committed to an image, and some configuration settings.

If you have a period of time required to keep images, you will need to size your Journal correctly. Here is a simple equation from the RecoverPoint Administrator Guide:

MinJournalSize = 1.05 * [(D data per second)*(required rollback time in seconds)/(1 – image access log size)] + (reserved for marking)

For example - if:

D data per second = 5 Mb/s

required rollback time = 24 hr = 86 400 s

image access log = 0.20

reserved for marking = 1.5 GB

Consequently, the minimum journal size would be:

1.05 * 5 Mb/s * 86 400 s / (1 - 0.20) + 1.5 GB = 579 000 Mb

579 000 Mb = 579 000/8 MB = 72 375 MB = 72.4 GB

You can consolidate images hourly, daily, and weekly as well as to increase Journal capacity.

Conclusion

Benefits for RecoverPoint with VMWare environment include:

- VM protection at VM-level granularity will let you know if the VM is fully or partially protected or not not protected at all and totally integrated with vCenter in VMware ESXi 5.1 Update 1 and VMware ESXi 5.5 with vCenter vSphere Web Client environments. So less complexity to verify replication is configured correctly or not.

- Replicates VMs using VMDK and RDM devices with any type of storage connectivity with its different topologies. For example, there is no longer need to worry if RDM is physical or virtual.

- Provides either local or remote replication protection or both depending on topology you used.

- Supports both synchronous and asynchronous replication and, with Syncreplication can change automatically to Async according to defined thresholds you choose or specify. For example, if latency increases it replicates in Async; when latency drops, it automatically resumes Sync replication.

- All Scenarios for DR workflows exist, including testing, failing over, failing back, and recovering production of a single Consistency Group or a group of Consistency Groups to and from any point in time.

- WAN compression and deduplication to optimize bandwidth consumption lowers cost of connectivity to DR.

- REST API for the developer community with RecoverPoint for VM topology.