AVAMAR SOLUTIONS FOR VIRTUALIZED INFRASTRUCTURE

Balaji Panchanathan
Principal Software Quality Engineer
EMC
Balaji.Panchanathan@emc.com

Satchidananda Patra
Software QA Group Team Lead
EMC
Satchidananda.Patra@emc.com

Pravin Kumar A
Senior Software Quality Engineer
EMC
Pravin.ashokkumar@emc.com

EMC²

Table of Contents

Introduction

Virtualization is widely adopted in the IT industry to simplify IT operations and allow IT organizations to respond faster to changing business demands. The IT industry has witnessed a growth trend in virtualization infrastructure. Vendors such as VMware and Microsoft Hyper-V are the key players in this space.

EMC Avamar® is tightly integrated with VMware and Microsoft Hyper-v which in turn have widened the data protection space for virtual infrastructures and provided flexibility for customers to use the Avamar data protection solution for their virtualized IT infrastructure.

Why data protection for virtual infrastructure?

Exponential data growth, regulatory requirements, and 24X7 availability have all increased the difficulty in completing both virtual and physical backups successfully within shrinking backup windows. In virtual environments, storage challenges are exacerbated due to virtual server sprawl that virtualization almost inevitably brings, creating hundreds of duplicate server and/or desktop images. Virtual server consolidation further adds to the backup window problem.

Improving backup operations and reducing storage costs involved in server virtualization are key to successful expansion of virtualization across an organization. Leveraging deduplication in virtualized environments can address not only the pain points and barriers to expansion, but can also offer a strong return on investment (ROI) and significantly reduce the total cost of ownership (TCO) of virtual infrastructure data protection.

EMC Avamar solution to address data protection challenges of virtual infrastructure

EMC Avamar integration with VMware, Hyper-V, and Data Domain® provides a complete data protection solution for virtual infrastructure. Combination of all of the following provides the data protection solution to virtualized data centers and cloud infrastructure.

1. Data protection solution for VMWare
2. Data protection for Hyper-V
3. Avamar virtual edition
4. Avamar integration with Data Domain
5. Data protection for Azure cloud

Why choose Avamar for virtual infrastructure?

Avamar software quickly and efficiently protects virtualized infrastructure environments by using agents on the virtual machines or virtual storage hypervisor to reduce the size of backup data within and across virtual machines. For virtual machine backups, Avamar eliminates traditional backup bottlenecks caused by the large amount of redundant data that must pass through the same set of shared resources, i.e. the physical server's CPU, Ethernet adapter, memory, and disk storage.

Capabilities of Avamar

- Up to 10X faster daily full backups
- Up to 95 percent reduction in required global backup storage media
- Up to 99 percent reduction in required daily network bandwidth
- Encryption of backup data in flight and at rest
- Scalable grid architecture
- Daily server integrity and data recoverability checks
- Simple one-step recovery
- Flexible deployment options, including EMC Avamar Data Store and EMC Avamar Virtual Edition for VMware and Hyper-V (a virtual appliance)

Avamar Virtual Edition

Avamar Virtual Edition (AVE) is a single-node Avamar server that runs as a virtual machine on VMware ESX or Microsoft Hyper-V. AVE is similar to existing single-node Avamar backup and recovery software providing fast daily full backups by deduplicating the data at the client to eliminate backup bottlenecks caused by constrained or shared infrastructure and networks. Thus, AVE is an ideal data protection solution for small, medium, or remote offices that have standardized on virtual infrastructure.

Benefits of AVE

- Rapid, cost-effective deployment.
- Simplified management by virtualizing all aspects of the backup and recovery offering.
- Lower cost by leveraging shared server and storage infrastructure.
- Replication in virtual environments as well as between virtual and physical.
- Rapid return on investment.

- Starting with Avamar 7.0, each Avamar Virtual Edition supports up to 4TB of licensed disk storage capacity provisioned to a virtual machine on a VMware and Hyper-V. Other capacities are available: 500GB, 1TB, and 2TB.

Choosing the right hypervisor for AVE

VMware and Hyper-V are highly efficient and both support almost the same functionality with respect to Avamar data protection functionality. Thus, before implementing or recommending a AVE- based data protection solution, it is important to do an in-depth performance analysis on the virtual infrastructure to support the business requirement.

- AVE system performance benchmarking test
- AVE backup/restore benchmarking test

AVE system performance benchmarking

AVE system performance testing should be done to analyze the storage I/O subsystem to ensure that AVE can function at an acceptable level in the virtual environment. Avamar AVE has recommended throughput levels for each AVE configuration.

- Example:

  .5TB AVE : 400 seeks/sec
  1TB/2TB/4TB AVE : 500 seeks/sec

16GB RAM | 2TB Hyper-V Benchmark				
Pass	(eager	2	3	Avg 2&3
Total Minimal Write Throughput [MB/sec]	307	304.3	311.6	**307.905**
Total Minimal Read Throughput [MB/sec]	417.8	395.2	398.4	**396.8**
Total Write Throughput (sum) [MB/sec]	308.5	305.1	312.2	**308.64**
Total Read Throughput (sum) [MB/sec]	427	399	403	**401.02**
Total Seek Minimal Throughput for 4 Threads (sum)	528.5	321	327.4	**324.195**
Total Seek Throughput for 4 Threads (sum)	532.2	326	329	**327.5**

16GB RAM | 2TB VMWare Benchmark				
Pass	(eager	2	3	Avg 2&3
Total Minimal Write Throughput [MB/sec]	318.4	318.2	305.2	**311.68**
Total Minimal Read Throughput [MB/sec]	364.8	373.7	334.9	**354.27**
Total Write Throughput (sum) [MB/sec]	324.7	322.9	313.4	**318.19**
Total Read Throughput (sum) [MB/sec]	389.7	392.7	372.7	**382.7**
Total Seek Minimal Throughput for 4 Threads (sum)	549	491	522.1	**506.56**
Total Seek Throughput for 4 Threads (sum)	553.8	510.6	551.1	**530.82**

Analysis

Total seek throughput of AVE on Hyper-V is not in the acceptable range, so storage subsystem compatibility with Hyper-V needs to be checked and improved before implementation.

AVE backup/restore benchmarking

This type of backup/restore performance analysis provides some indicative information on choosing the right hypervisor solution for AVE. The performance results will vary depending on the backup data size, data type, application type and number of parallel backup, etc.

Note: Following table shows some sample data for example purpose only

In this table, the AVE configuration and data type is the same in both Hyper-V & VMware.

| 2TB | Hyper-V | 16 GB RAM				
Pass	Bench 1	2	3	Avg
Performance Result ID	135317	135318	135320	
aggregate_backup (GB/hr)	131.93	134.9	122.21	133.42
aggregate_restore (GB/hr)	115.15	117.06	116.51	116.11
cp_time(min)	0.32	0.27	0.3	0.295
cp_rate (stripes/min)	14578.1	17277.8	15550	15928
hfscheck_time (min)	101.03	100.93	100.6	100.98
hfscheck_rate (stripes/min)	44.6	44.64	44.79	44.62
hfscheck_throughput (GB/hr)	867.49	868.36	871.23	867.93
hfscheck_startup_time (min)	1	1	1	1
hfscheck_indexsweep_time (min)	7.59	7.43	7.35	7.51
hfscheck_datasweep_time (min)	90.44	90.57	90.29	90.505
hfscheck_paritysweep_time (min)	0	0	0	0
hfscheck_refcheck_time (min)	1.58	1.54	1.62	1.56
hfscheck_indexsweep_rate (stripes	34.26	34.99	35.37	34.625
hfscheck_datasweep_rate (stripes	46.97	46.9	47.05	46.935
gc_time (min)	57.28	56.12	56.07	56.7
gc_rate (GB/hr)	416.56	425.17	425.55	420.87
gc_bytes_freed (MB)	407222	407222	407222	407222
gc_refcount_time (min)	1.28	1.28	1.32	1.28
gc_chunk_deletion_time (min)	56	54.83	54.75	55.415
rollback_restart_time (min)	11	11	11	11
rollback_restart_rate (stripes/min	409.73	409.73	409.73	409.73

| 2TB | VMware | 16 GB RAM				
Pass	Bench 1	2	3	Avg
Performance Result ID	135325	135328	135330	
aggregate_backup (GB/hr)	128.93	133.76	135.74	131.345
aggregate_restore (GB/hr)	139.46	128.13	123.7	133.795
cp_time(min)	0.58	0.67	0.63	0.625
cp_rate (stripes/min)	8825.86	7631.3	8115.9	8228.6
hfscheck_time (min)	142.52	137.17	137.5	139.845
hfscheck_rate (stripes/min)	34.78	36.14	36.05	35.46
hfscheck_throughput (GB/hr)	680.53	707.07	705.37	693.8
hfscheck_startup_time (min)	2.52	2.53	2.77	2.525
hfscheck_indexsweep_time (min)	8.67	8.91	8.24	8.79
hfscheck_datasweep_time (min)	128.07	122.38	123.15	125.225
hfscheck_paritysweep_time (min)	0	0	0	0
hfscheck_refcheck_time (min)	2.51	2.39	2.55	2.45
hfscheck_indexsweep_rate (stripes/m	29.99	29.18	31.55	29.585
hfscheck_datasweep_rate (stripes/mi	36.71	38.42	38.18	37.565
gc_time (min)	68.68	68.83	69.42	68.755
gc_rate (GB/hr)	341.75	341.01	338.11	341.38
gc_bytes_freed (MB)	400578	400584	400584	400581
gc_refcount_time (min)	1.48	1.53	1.52	1.505
gc_chunk_deletion_time (min)	67.2	67.3	67.9	67.25
rollback_restart_time (min)	16	16	16	16
rollback_restart_rate (stripes/min)	309.88	309.88	309.88	309.88

Take away points from this analysis from the above table is:

1. Backup performance of AVE on Hyper-V is better than AVE on ESX
2. Restore performance of AVE on ESX is better than AVE on Hyper-V
3. HFSCHECK performance of AVE on Hyper-V is better than AVE on ESX
4. Garbage collection performance of AVE on Hyper-V is better than AVE on ESX.(meaning the maintenance window of ESX will be more than Hyper-V)

Conclusion with this analysis

- In this infrastructure, Hyper-V is the preferred option for faster backup and lower maintenance window.

- In this infrastructure, VMware ESX is the preferred option for cases where data restore frequency is high. Also, ESX would be the preferred option for replication source.

Recommendation on AVE configuration

The decision on AVE size should be determined according to the business requirement.

- .5TB/1TB AVE can be an ideal option where Data Domain is the preferred storage and Avamar will be used primarily for policy and management.
- 2TB/4TB AVE can be an ideal option for small business units where the rate of data growth rate is minimal and Hypervisor local storage is the preferred storage option. This is a cost-effective solution.

Avamar Integration with Hyper-V

Traditional data protection challenges

Hyper-V virtualization technology has brought us a number of benefits, such as faster virtual server deployment and easier management. When looking to protect these new rapidly growing server implementations, traditional backup methods become a challenge.

Business opportunity for Avamar with Hyper-V:

As per recent market analysis, MS Hyper-V holds **30.6% (**vs VMware 46.4%) market share in the private cloud space. Looking at Hyper-V market share, there is a big opportunity for EMC to sell Avamar as a data protection solution for customers that have Hyper-V virtual infrastructure.

How Avamar works with Hyper-V

There are two ways to back up and restore Hyper-V data with Avamar software:

1. Guest backup and restore
2. Image-level backup and restore

Guest backup and restore

With guest backup and restore, an Avamar file system client is installed on each virtual machine on the Hyper-V Server. Avamar considers each virtual machine to be a separate client to perform individual backups of each virtual machine.

If there are databases or specific applications – such as Microsoft Exchange or Microsoft SharePoint – that run on the virtual machine, an Avamar plug-in can be installed to protect the application data.

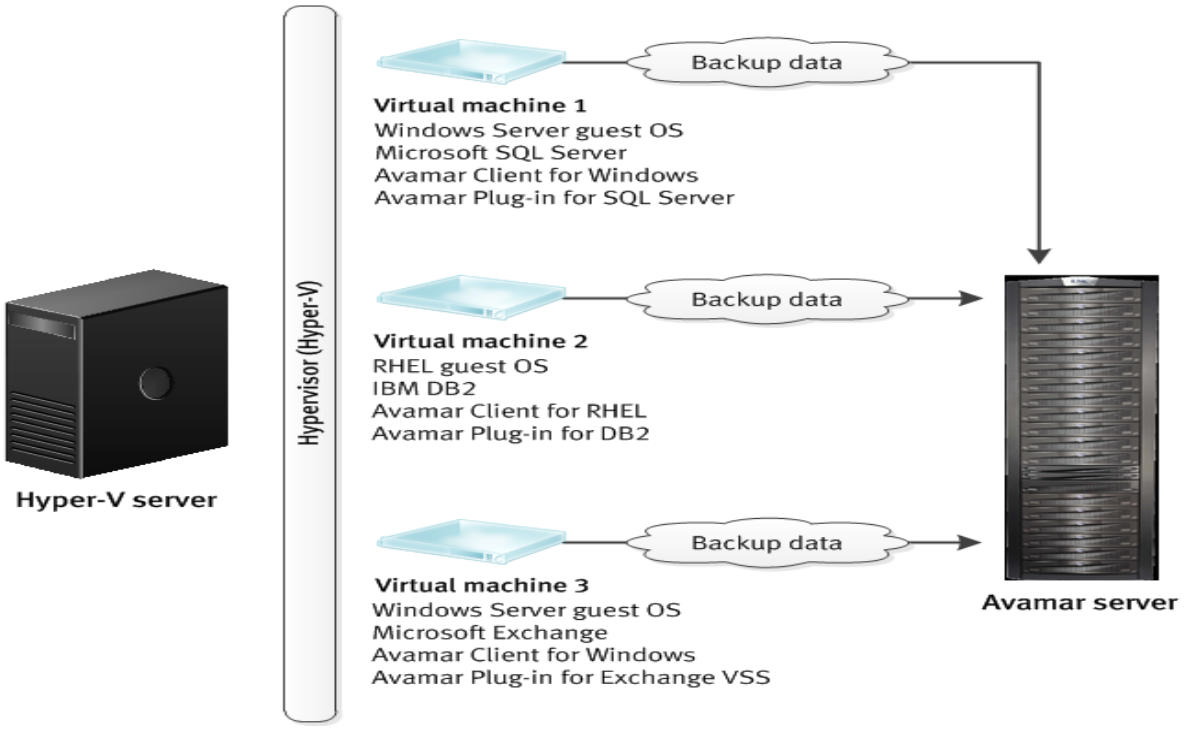
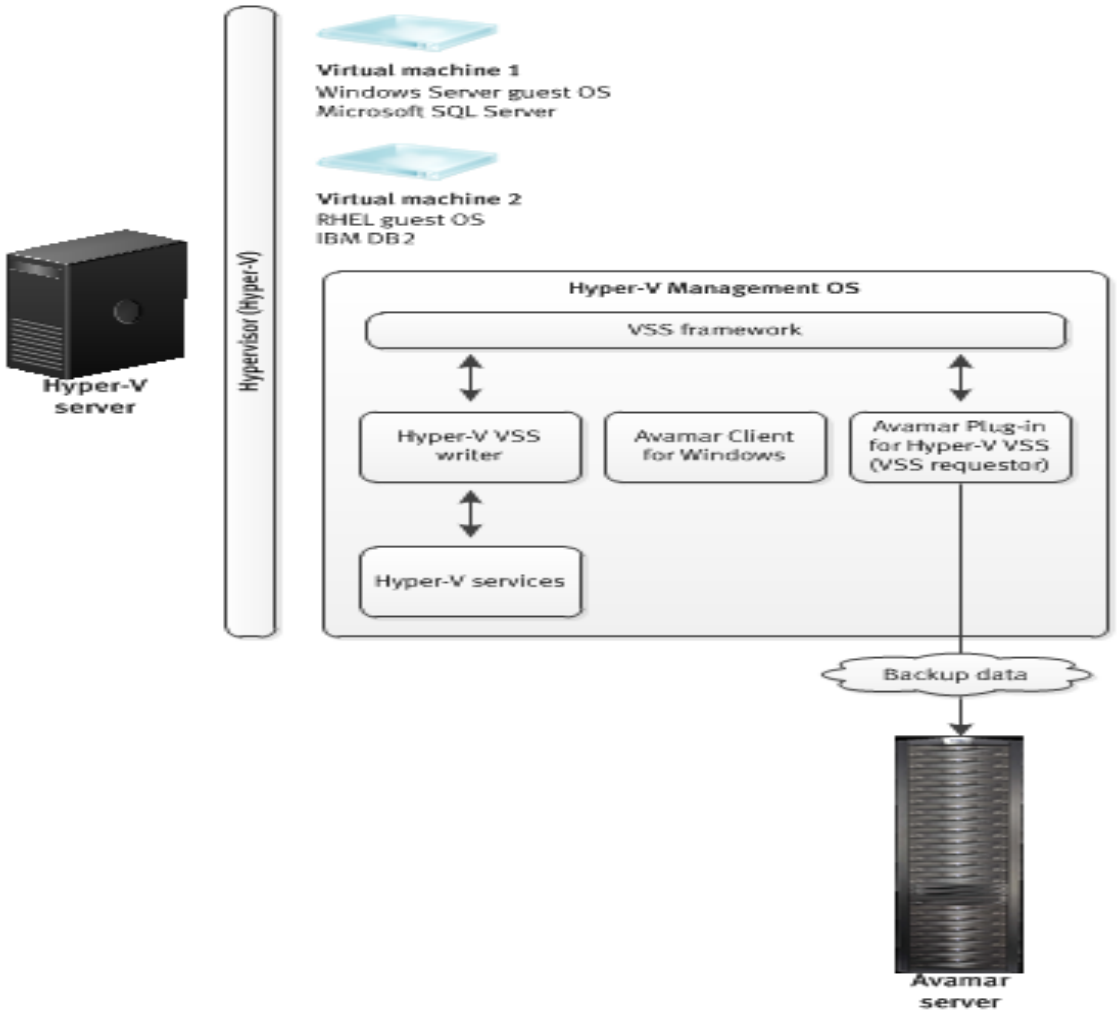


Image-level backup and restore

With image-level backup and restore, the Avamar Client for Windows and the Avamar Plug-in for Hyper-V VSS is installed on the Hyper-V management OS (parent partition).



The Avamar Plug-in for Hyper-V VSS uses Microsoft Volume Shadow Copy Service (VSS) Technology to perform backups. VSS enables volume backups while applications on a system continue to write to the volumes. The full image-level backups can be performed for individual virtual machines or the Host Component. Image-level backups occur from the Hyper-V management OS instead of the individual virtual machines.

Image level backup capabilities:

- Meets disaster recovery (DR) requirements
- Federated Hyper-V CSV backup and recovery
- Complete protection of virtual machines (VM's)
- Recover VM's to alternate Hyper-V servers
- Granular Level Recovery (GLR) for faster restores

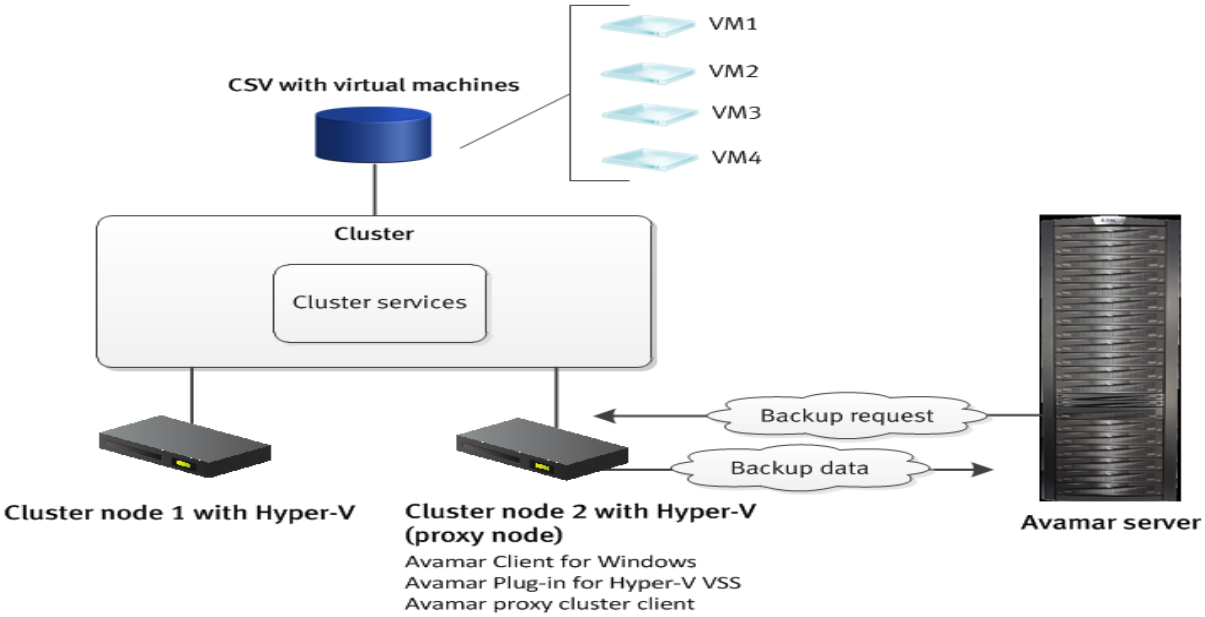
How Avamar works with Hyper-V in a cluster

Hyper-V has the option to create failover cluster to prevent single point of failure. In a failover cluster, all servers (nodes) run Hyper-V and can host one or more virtual machines. A virtual machine can run on multiple nodes in the cluster, but can only be active on one node at a time. Avamar can perform image-level backups of Hyper-V in a failover cluster environment.

Windows Server 2012 clusters with CSV

Using single proxy

Avamar uses proxy architecture to back up and restore virtual machines on a Clustered Shared Volume (CSV) in a Windows Server 2012 cluster. Avamar Client is installed for Windows and the Avamar Plug-in for Hyper-V VSS on one or more physical nodes in the cluster. The Avamar server communicates with the Avamar proxy cluster client to perform backups and restores of virtual machines on CSV, regardless of which node is managing the virtual machines at the time of the operation.



Using multiple proxy

Avamar Client is installed on Windows and the Avamar Plug-in for Hyper-V VSS on each physical node in the cluster that will act as a proxy. Cluster Configuration Tool is installed on the node that will serve as the primary proxy node to configure the Avamar proxy cluster client, which manages the backup process in the cluster. The Cluster Configuration Tool is run on the primary proxy node to configure the secondary proxy nodes.

Hyper-V Granular level recovery

Granular level recovery (GLR) enables a user to mount an image backup to a temporary file system and then browse and restore individual files and folders.

Benefits of GLR:

- Mount Hyper-V Image backups quickly from within Avamar and Data Domain to recover files or folders
- Faster recovery
- Easy to use

Avamar for Hyper-V based private cloud

A larger scale enterprise environment requires thousands of application servers to be deployed for all the supporting business critical data. This scale requires deployment of the Hyper-V foundation technology to be deployed in a cluster implementation to provide services at large and highly available scale. Microsoft leverages Cluster Shared Volumes (CSV) as the virtual machine storage mechanism for providing access to all VM files to multiple Hyper-V Servers. This allows for the ultimate in flexibility and redundancy enabling any Hyper-V Server to service compute and memory resources for any VM stored inside CSVs.

At the run time of backup, the plug-ins communicate the current states and locations of VMs and provide complete protection coverage for all VMs in the private cloud. With Hyper-V on Windows Server 2012, it is possible to designate multiple proxy node systems for backup, which allow for performance increases in environments hosting hundreds of VMs.

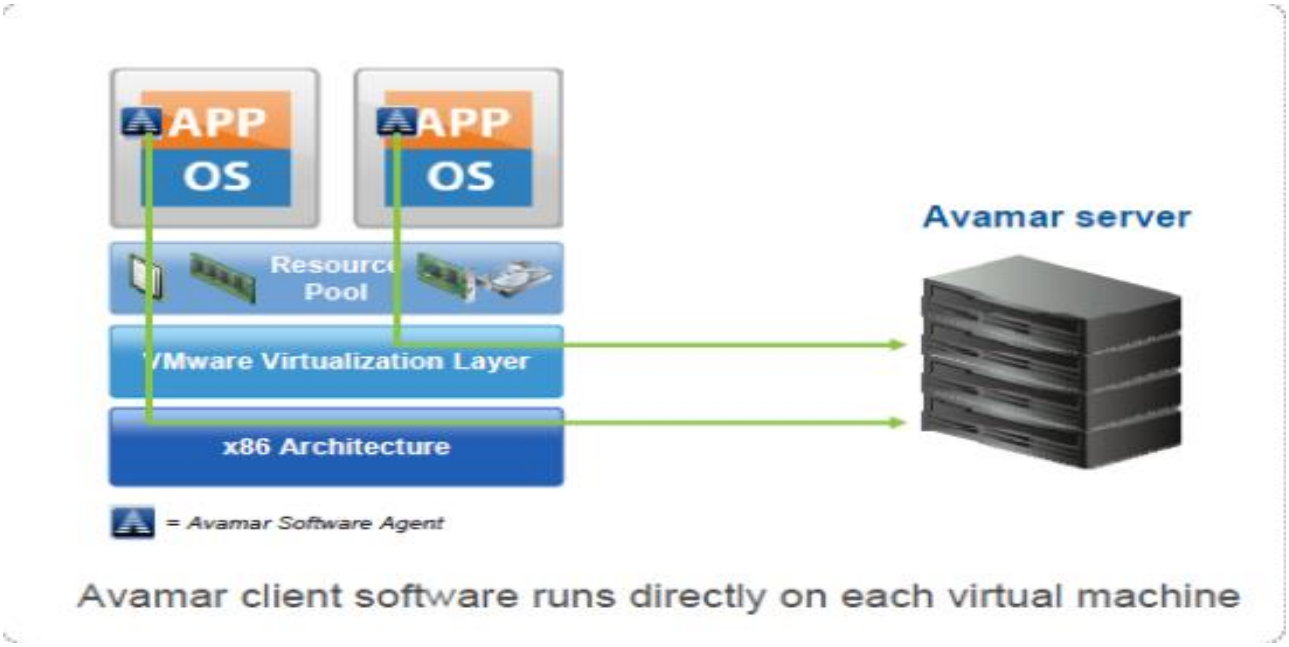
Key benefits of Avamar in Hyper-V based private cloud

- No impact to Hyper-V servers during backup
- Simplified configuration and management
- Increased operational flexibility
- File Recovery from Image Backup
- Utilizes existing LAN/WAN IP bandwidth and virtual infrastructure
- Centralized Policy Management
- Centralized Monitoring and Reporting
- Disaster Recovery and Replication

Avamar integration with VMWare

Guest-based backup

The basic Avamar client software is installed on each virtual machine. Backup configuration in this method is exactly the same as that of physical server. For specific application backup such as MS SQL, MS exchange etc., the respective plug-in software is installed on top of basic Avamar client software.



Guest-based backup capabilities

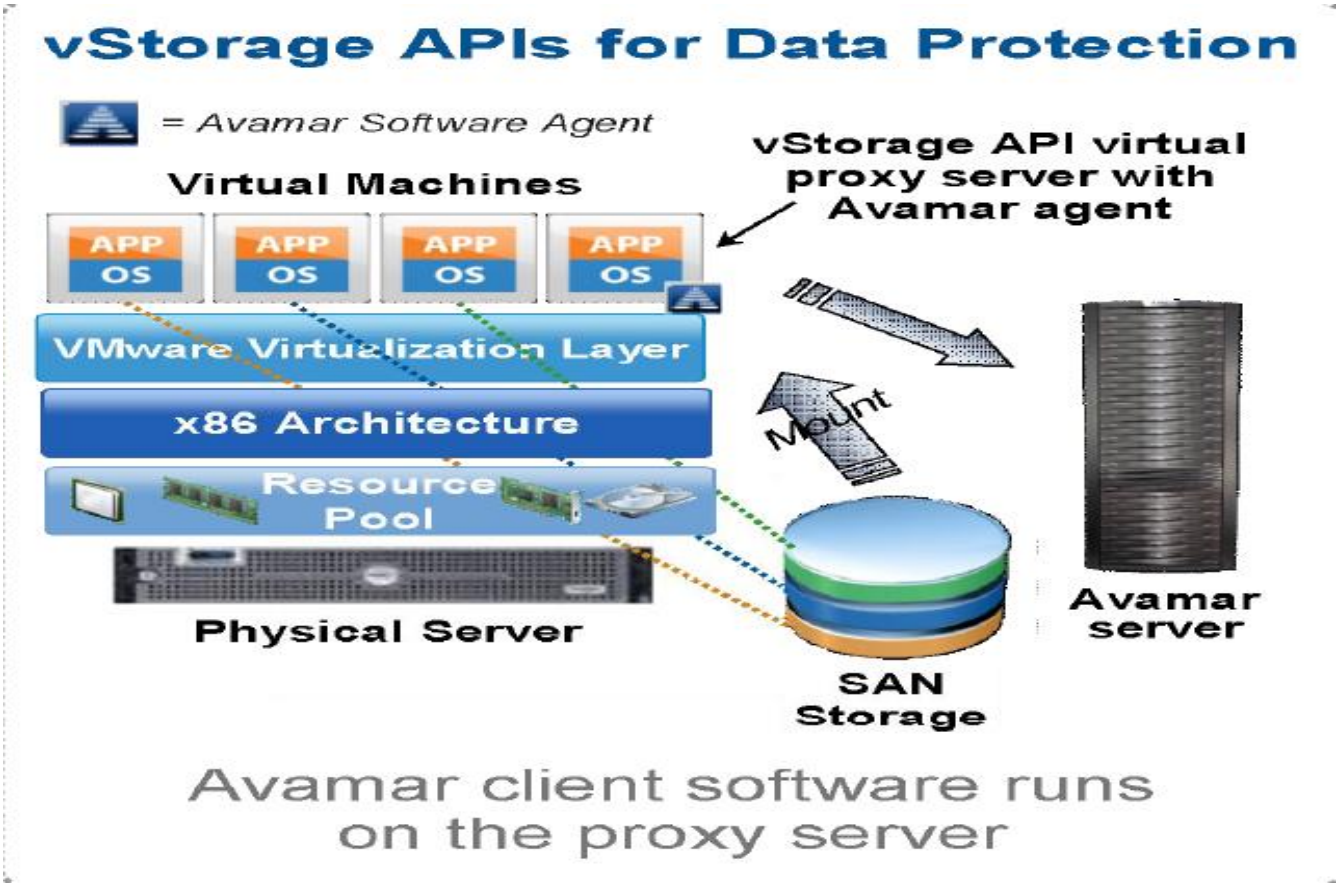
- Highest level of data deduplication.
- Support for backup of applications inside VM.
- Supports file-level restore.
- No prior VMware knowledge required.
- Unchanged day-to-day procedures for backup.

Guest-based backup limitations

- Since each virtual machine has a separate backup client installed, ESX servers with a large number of virtual machines may experience a strain on resources, especially memory, if all VMs are backed up at the same time.
- No Image-Level restores option.
- File-level restore is done in one step; however, full system restore is done in two steps.

Image-level backup via vStorage API for Data Protection

VMware vStorage APIs for data protection enables LAN-free backup and offloads the backup workload to a backup proxy server. The vStorage APIs' proxy server can mount a virtual machine's .vmdk files. Using the Avamar agent to back up the mounted VM disk, Avamar provides data deduplication at both the file level and the .vmdk level.



VMware Image Backup uses a proxy server to handle backup processing. Using VADP, Avamar can back up VMware virtual machines without using backup agents inside each virtual machine. This proxy server can communicate with the vCenter server to mount a snapshot of a particular VM's vmdk to perform image-level backup of that virtual machine. With this method, deduplication is provided on the file level as well as on the .vmdk level.

VMware Image-level backup capabilities

- Perform backup and restores without installing agents on individual virtual machines.
- Perform backups from a centralized location, thus offloading backup processing from the ESX hosts.
- Perform backups any time because backups are non-disruptive to the VMs being backed up.
- The ability to perform backups at any time provides greater flexibility in scheduling backups and increased backup window times.

VMware Image backup limitations

- Requires a moderate amount of VMware knowledge to set up and configure backups and restores.
- Backups consume ESX server resources, including CPU, RAM, and disk.
- Backup requires a temporary virtual machine snapshot.
- Backup is crash-consistent snapshot of full VM image and may not support a full system restore without data loss.

Avamar for VMWare Changed Block Tracking (CBT)

VMware Changed Block Tracking is an incremental backup technology for virtual machines, comparable to snapshot differential. Avamar takes advantage of VMware's Changed Block Tracking (CBT) to further speed up the backup and restore processes. VMware presents only changed blocks to the Avamar software, where each block is broken into variable length segments and further evaluated for uniqueness. Only the unique segments are sent for backup, achieving the fastest backup possible. Conversely, the restore process also leverages CBT for faster recovery.

Avamar for vCloud director deployment

Avamar has embedded its backup and recovery services into vCloud Director so they can be shared and distributed in a multi-tenant model, as a managed resource in vCloud Director. These resources are assigned to Org VDCs and through that assignment, exposed to tenants for consumption in protecting vApps. Avamar provides the first standard backup API for VMware cloud providers. Service Providers and enterprise admins can back up native vCD constructs by applying polices (e.g. Gold, Silver, Bronze) from a set of catalogs at the Org, vDC, and vApp levels, enabling service tiers of protection.

Avamar Data Domain integration

Avamar is tightly integrated with Data Domain to expand on existing Avamar capacity and make better use of data deduplication. Leveraging Data Domain Boost (DD Boost) software, Avamar can back up data directly to a Data Domain system or to the Avamar Data Store. For small/medium infrastructure, Avamar Virtual Edition (AVE) and Data Domain can be used as a cost-effective and valuable data protection solution.

How does backup work in an AVE+ DD environment?

The Avamar client installed on each host will work as a typical Avamar client performing client-side deduplication. Whether the backup data goes to the Avamar Data Store or to the Data Domain system is completely transparent to the client. Metadata is sent to the Avamar server and the backup data will be directed to either the Avamar Data Store or the Data Domain system depending on how the relevant dataset is configured in Avamar.

Avamar clients that support backup and restore to and from Data Domain include:

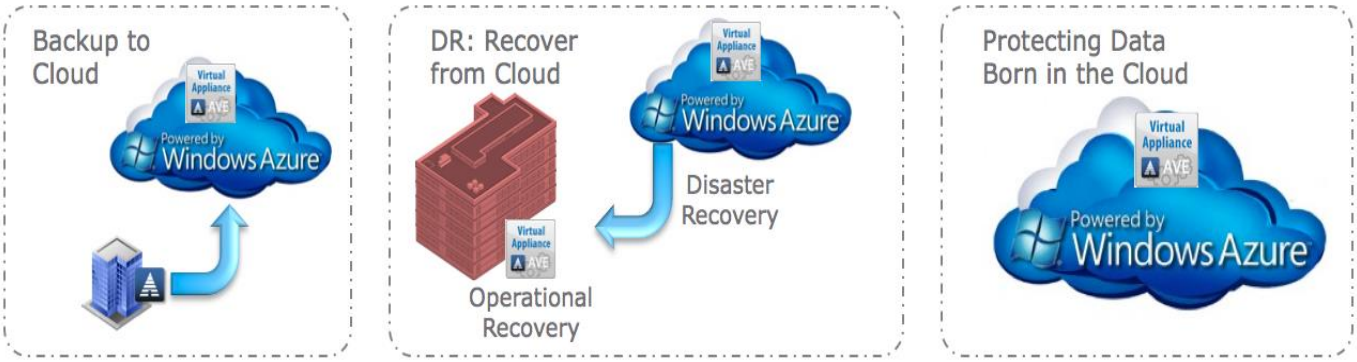
- Microsoft Exchange VSS
- Microsoft SQL Server
- Microsoft SharePoint VSS
- Oracle
- VMware image backup and restore

How recovery works

Data recovery is also a process that is completely transparent to the Avamar client, regardless of whether the data is being retrieved from the Avamar Data Store or from the Data Domain system. Backup administrators can recover data using native Avamar recovery processes and recovering from the Avamar front-end is also seamless, as Avamar will automatically retrieve required datasets from the Avamar Data Store or Data Domain system.

Avamar for Microsoft Azure cloud

EMC Avamar provides a complete data protection solution for Azure cloud environments. The Avamar Virtual Edition (AVE) is designed to support Microsoft Azure cloud infrastructure. More and more companies are expanding their cloud initiatives to include public clouds. Following are the design model can be implemented for Azure.



Backup to Azure

A small office can deploy Avamar software agents in their Hyper-V environment or on physical hosts. The Avamar agents will reduplicate, compress, and encrypt backup data, then send it to an AVE hosted in the Azure cloud.

Replicate Backup Data to Azure for disaster recovery

A mid-sized branch office can choose to deploy an AVE for their Hyper-V private cloud for local backup and recovery. They can also replicate their daily changed data to an AVE hosted in Azure to meet their disaster recovery compliance.

Protecting VMs in Azure cloud

Virtual machines deployed in the cloud can now be protected with AVE for Azure, which provides cloud-based backup and recovery.

Avamar replication solution for virtual infrastructure

Avamar solves the challenges associated with remote replication, enabling fast, robust replication of single or multisite recoveries in virtual environments. Avamar utilizes global data deduplication technology to identify redundant data segments at the source – before transfer across the wide area network (WAN), considerably reducing required daily WAN bandwidth.

This efficiency allows existing bandwidth to be used for backup and disaster recovery (DR), despite slow or congested networks and infrastructure.

Root-to-replicate

This is a scheduled replication and is the most common implementation of Avamar. “Root-to-replicate" can be thought of as an offsite copy of backup data. A copy of the backup data is now located at the Avamar target for a directed restore. The primary gain with this mode is the ability to have an offsite copy of selected Avamar domains, subdomains, and their client data. This mode enables fan-in replication as well, supporting unidirectional, bi-directional, and multi-hop replication combinations.

Root-to-root

Root-to-root (R2R) migration is a replication mode that creates a logical copy of an entire source Avamar server to a designated destination server. This means that data and clients in the root ("/") on the source Grid are copied directly to the root domain of the destination server (replication target). The destination will look exactly like its source. The result of this replication type is that source server data is fully modifiable on the destination server as though the source clients were registered with this destination server.

Use cases in virtual environment

A common scenario is where the customer has either 1 TB nodes or an AVE in which they’d like to migrate to 2 TB capacity nodes. The need to move to a higher-capacity AVE (.5 TB to 1 or 2 TB, or 1 TB to 2 TB) is also a typical driver. The upgrade path of each of these scenarios is to conduct an R2R replication to the new Avamar server or AVE, use it as a swing server, and upgrade the source. The primary purpose of the R2R mode, though, is to simply create a logical copy of the Avamar Grid for DR purposes.

AVE-to-AVE replication

AVE can replicate backup data to a remote site where destination can also be an AVE. Deduplication technology is used to reduce the duplicate data in replication. In cases of larger infrastructure, multiple AVEs can replicate to a single Avamar grid server (multi-node) at remote site.

AVE-to-AVE replication with Data Domain in backed is a supported replication combination.

Calculate ROI & TCO

While offering Avamar data protection solution for virtual infrastructure, consider the following factors to assess the Return on Investment (ROI) to reduce Total Cost of Ownership (TCO)

- New application workload or migration?
- Growth rate of VMs?
- Infrastructure management cost.
- Data center data growth rate.
- Rate of out-bound data transfer to public cloud.

Glossary

AVE – Avamar Virtual Edition

DD – Data Domain

VM – Virtual Machine

Referred links

http://storageswiss.com/2013/12/06/avamars-virtualization-journey-continues/

https://nobackupblues.wordpress.com/2013/07/10/avamar-7/

http://searchitchannel.techtarget.com/Microsoft-Hyper-V-vs-VMware-ESX-cheat-sheet

http://www.computerweekly.com/feature/Data-protection-The-next-generation

http://www.thomasmaurer.ch/2014/07/hyper-v-is-eating-vmwares-lunch/