

Backup of Virtual Environments

Challenges And Pitfalls

The virtualisation of server infrastructures is common practice these days and delivers many advantages. Of course these new technologies also require some changes to the management of IT environments. One of the new challenges is the backup of virtual servers. At first sight, this is no extra challenge, as the backup of servers is a common task in every IT department. But with the introduction of virtualisation and clustered virtualised environments there are additional opportunities, and pitfalls.

The basis of all backup solutions in a virtual infrastructure is the concept of initiating the backup from the physical host system. With this idea, there is no need for additional backup agents to be run in each virtual guest and you have only to look at the physical host.

In addition, the market leaders, who are VMware and Citrix XenServer provide Backup API's which allow backup utilities to directly access the Management layer and give a transparent view to the physical hosts. This is fairly straight forward and the easiest way. As always, the devil is in the details.

The Backup System directly contacts the API of the hypervisor to start a backup of a virtual guest. The hypervisor generates a snapshot of the running guest and delivers the snapshot to the Backup Software.

The API delivers the data of the guest together with the Meta information of the hypervisor. The snapshot is executed on the running guest and the virtualised systems are not interrupted.

Also the physical Hardware is transparent – the Backup System does not have to know on which physical host the guest is running during the backup.



Figure 1: Backup APIs

Limitations to the snapshot backup

This snapshot backup concept itself is the limitation of the Backup API's. Consider a server, it will contain different types of data:

- The Operating system with all its drivers, DLL's, configurations and Hardware related information.
- Binaries of Applications, Configuration files and very static files, like in a web server and files within a file server.
- Data managed in Databases and Groupware Systems.

The first two types of data are perfect for a snapshot backup. This type of data is very static, so during the snapshot a consistent saveset is generated by the hypervisor. Databases and Groupware Systems contain dynamic data, which is regularly changing.

The Backup API from the hypervisor is not able to inform the database about the ongoing snapshot. Consequently there will be data from the database stored during the backup, however the data will not be consistent and may be corrupted. It is likely that if a restore was required the database data would be useless.

The only way to avoid this problem is to use the traditional agent backup of the database. Install a hot backup module of the backup system into the virtual guest and perform the backup of the database via the LAN and not via the hypervisor.

The second typical problem appears when restoring. The backup via the API is a complete snapshot of the guest, Generally it's not possible to restore individual files out of this snapshot. To restore a single file the administrator has to restore the complete snapshot and extract the single file out of the snapshot. Bear in mind that to restore a file of 2 MB you may well have to restore a complete snapshot of 200 GB.

Backup of Virtual Environments

Challenges And Pitfalls

Differences with MS Windows Guests

These two scenarios are less black and white with MS Windows guests. First the hypervisor API's are able to use the Microsoft Volume Shadow Copy Service (VSS) for the operating system in the guest as well as for the Databases / Groupware systems in the guest. Before the snapshot is executed the Database gets the VSS command and is set to a consistent state. In this way the provided snapshot of the Database contains a consistent set of data. This works for systems which support VSS from Microsoft.

But – it's always the complete MS SQL or MS Exchange Server. A restore of a single MS SQL Database or a single MS Exchange mail is not possible out of this VSS snapshot. Also a PIT (Point in Time) restore is not possible.

Secondly, during a snapshot of a MS Windows guest the backup API's of the vendors provide a file listing of the files backed up. Several backup tools, like SEP sesam, interpret this file list and store the information. In this scenario a restore of individual files directly from the snapshot is possible, without having to restore the complete snapshot. This is only possible for MS Windows guests, and is not yet available for NetWare, OES or Linux guest systems.

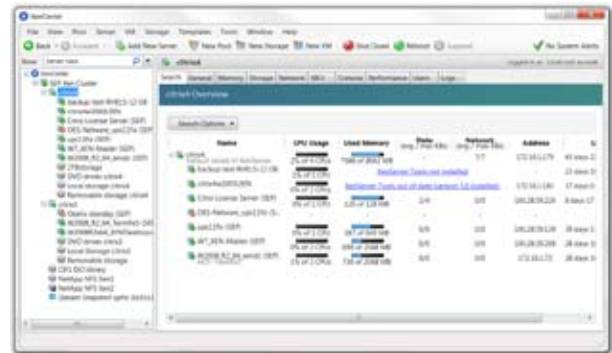


Figure 2: XEN backup administration

Another problem – distributed systems

In a MS Windows environment an additional issue occurs. More and more application environments do not run their databases on a single server. For instance a SharePoint Server may be distributed to a farm with three servers: The Domain Controller, the Database and the Web Service.

With a backup via the hypervisor API and VSS a consistent snapshot of a single server is provided, for example the MS SQL Database. But this snapshot is not consistent with the corresponding data in the second server, for example the domain controller. A similar problem appears if Active Directory is distributed to several servers.

Again the only way for a consistent and useful backup is the utilisation of hot backup agents of a backup solution in the virtual guests.

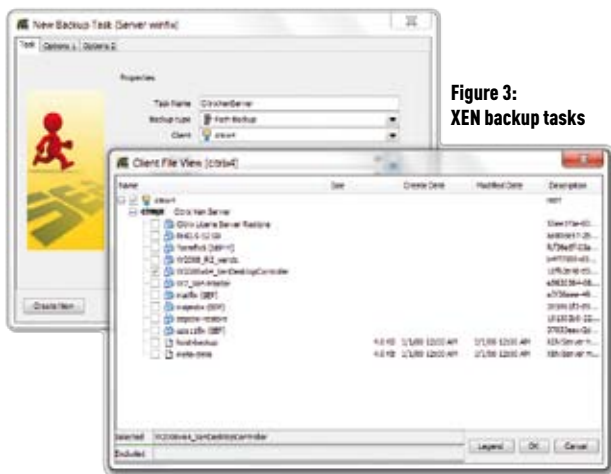


Figure 3: XEN backup tasks

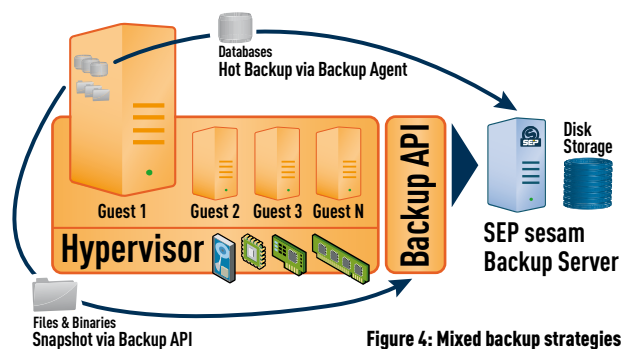


Figure 4: Mixed backup strategies

Summary

Some virtual environments provide dedicated Backup API's:

- VMware – VADP (vStorage APIs for Data Protection)
Citrix XenServer – Backup API
- Hyper V – VSS Writer (Volume Shadow Copy Service)

All provide very similar functionality to perform backups of the virtual guests directly via the hypervisor. Here is a summary of the pros and cons of this backup strategy:

Pros:

- Easy configuration – no agents need to be installed
- Fast backup of complete virtual servers
- Perfect for Disaster Recovery of complete servers
- Usable for static data, like Application server, Infrastructure server and file servers (limited)

Cons:

- Restore of single files only for Windows guests
- Limited use for Databases and Groupware Systems with VSS Interface
- Restore only for the complete Database / Groupware
- Unsuitable for Databases and Groupware Systems without VSS
- No consistent backup of distributed Applications / Databases

Recommendations

The utilisation of the hypervisor APIs in an organisation's backup strategy is important to ensure the availability of server systems and data.

However it has to be integrated into the complete backup and restore strategy. In addition to a backup via the virtualisation API's you may require a traditional agent backup for dedicated systems.

To ensure an integrated backup concept it is best to use the same solution for the backup of the virtual environments as well as for the physical and agent based backups.

SEP sesam provides agentless backup for VMware, Citrix XenServer and Hyper V environments. As well as hot backup agents for all types of databases and groupware systems, including NetWare, OES and Groupware backup. With SEP sesam the administrator gets one backup solution for the complete physical and virtual infrastructure. The result is an effective and easy to manage backup environment.



Author: Hubert Schweinesbein studied computer science in Nürnberg. Starting 1999 he worked for more than 6 years in the product management of SUSE and led the EMEA pre sales Team of SUSE and Novell. Since 2008 he has been building up the international sales channel for SEP. SEP is the leading European backup vendor. SEP sesam provides a single backup solution for Windows, Linux, Novell, Unix and for virtualised environments like VMware, Citrix Xen and Hyper V. You can reach Hubert via hs@sep.de

About SEP AG

SEP has delivered highly reliable software solutions to multi-national industry partners for over 15 years. Working closely with open source and proprietary computer manufacturers SEP developed industry-leading expertise developing network-wide, enterprise class data backup and recovery solutions. SEP sesam is installed and securing mission critical data in over 2000 enterprises and organisations in over 40 countries world-wide.

Headquarter:

SEP AG

Ziegelstrasse 1

83629 Weyarn

Germany

Phone: +49 (0)8020 180 0

Fax: +49 (0)8020 180 666

Web: <http://www.sep.de>

US Office:

SEP Software Corp.

94 Benthaven Place

Suite 100

Boulder, CO 80305

Phone: +1 303 417 6316

Web: <http://www.sepsoftware.com>

All brand names and product names are registered trademarks and trademarks of their respective owners.