Wishful Thinking vs. Reality in Regards to Virtual Backup and Restore Environments

Best practices for backing up virtual environments

Published by NovaStor

www.novastor.com
# Table of Contents

Why choose virtualization? .................................................................................................................. 4  
The virtualization challenges ........................................................................................................... 4

What is virtualization? .......................................................................................................................... 4  
Physical server ..................................................................................................................................... 4  
Virtual (or logical) server .................................................................................................................. 5

Virtualization systems and backup interfaces .................................................................................... 5

Backup concepts: definition of terms .................................................................................................. 6  
Backup ............................................................................................................................................... 6  
Image/Imaging/Image backup ........................................................................................................... 6  
Disaster recovery .............................................................................................................................. 6  
Snapshot ............................................................................................................................................. 6  
Incremental backup / differential backup ......................................................................................... 7  
Agent/Backup agent .......................................................................................................................... 7  
Host .................................................................................................................................................... 7  
Client/Client operating system .......................................................................................................... 7

Backup methods for virtual machines (VM) ....................................................................................... 7

Advantages and disadvantages of incremental or differential backups and the utilization thereof ........................................................................................................... 8

Inconsistent snapshots cause errors during a recovery ..................................................................... 8

Identifying a successful backup strategy ........................................................................................... 9

Objectives of a recovery ..................................................................................................................... 9  
Recovery of an application/file ......................................................................................................... 10

Disaster recovery .................................................................................................................................. 10

Best practices ....................................................................................................................................... 10

Summary ............................................................................................................................................. 12

About NovaStor ................................................................................................................................... 12
Introduction

NovaStor is a leading provider of software for data protection and availability, including world-class online backup solutions. Many well-known international service providers all over the world run data protection services based on NovaStor's technology to serve millions of home and business users.
Why choose virtualization?

Virtualization represents one of the most important developments in the IT sector over the past ten years. Virtualization allows users to install numerous instances of different operating systems on one hardware system, thereby enabling them to save money while increasing the productivity of the hardware. In a broader context, virtualization is viewed as a precursor for cloud computing and offers a broad range of innovative concepts and methods for turning IT into an on-demand product.

The virtualization challenges

One side effect of virtualization is a rapid growth in the amount of data to be managed that is stored and processed on a single server. The load placed on hardware components running virtual machines increases, and the data volumes on the individual components also rise. As a result of this, the loss of hardware components or their virtual machines corresponds to a loss of multiple application servers. The protection of virtual machines therefore plays an extremely important role in all virtualization concepts.

The protection of virtual machines requires a different approach to that of physical machines. Each physical unit can support any number of virtual machines, which can run different operating systems in turn. In larger installations – those with hundreds or even thousands of virtual machines distributed across multiple physical servers – load distribution represents a very significant challenge. Based on the respective load, virtual machines can move from one hardware unit to another. However, it must still be possible to back up and restore them on a consistent basis. The larger the virtual environment, and the more complex the mapped business processes, the more challenging it is to implement an appropriate backup strategy.

This whitepaper provides an overview of the principles and daily challenges of protecting virtual systems. The document describes the difference between virtual and physical systems and discusses the question of the right backup strategy under consideration of a key criterion for success – the fastest possible recovery of the desired virtual system state and its applications. Following this, the reader will be provided with some best practice suggestions for backing up virtual environments which have proven themselves over time.

What is virtualization?

Virtualization relates to the abstraction of software from the hardware. In order to be able to understand this concept and the effects thereof, the following terms must be differentiated from one another.

Physical server

A physical server is a hardware component on which software (usually an operating system) is operated. Physical servers can have as many CPUs, amounts of memory, etc. as desired.
Virtual (or logical) server
A virtual (or logical) server is a software server which is run as a self-contained and independent component on a physical server.

In the past, physical and logical servers were generally viewed in a 1-to-1 relationship, even though computer systems which could operate multiple “process areas” completely separated from one another (similar to a virtual server) were already available at the time. In order to optimize load distribution and increase fault tolerance, these processes could be shifted between two physical hardware units. The concept of virtualization therefore only partially represents a recent innovation.

Virtualization refers to the introduction of a hardware abstraction layer which separates a virtual server from the hardware of the physical server. Modern virtual servers can be moved freely between different hardware components nowadays, and multiple virtual machines can run in parallel on a single, powerful hardware system.

The benefits of virtualization are clear:

- Multiple virtual machines can dynamically access the same resources. The number of required hardware components can be reduced, thanks to more efficient utilization.
- Failover architectures can be implemented more easily and cheaply.
- New servers can be rolled out much more quickly and simply.

What does virtualization change in relation to the applications, however? The answer is simple: nothing at all!

A Windows operating system remains a Windows operating system, an Oracle database remains an Oracle database – no matter whether the installation is virtualized or carried out on a physical server. There are no differences whatsoever. A point-in-time recovery of a database remains a point-in-time recovery, and data corruption through human error still results in data corruption.

Virtualization systems and backup interfaces
Three virtualization systems currently dominate the market:

- VMware
- Microsoft Hyper-V
- Citrix XEN Server

The purpose of this document is not to review the advantages and disadvantages of the individual solutions; instead, it is primarily to take a look at their backup options and methods for protecting the virtual technologies.

All three solutions provide a backup interface, which can be used to take a snapshot of a virtual machine as well as an image of this snapshot. These interfaces do differ
from one another; the basic principle is however the same. In order to ensure the representation does not become too complex, the following explanation of backup concepts will focus only on the market leader: VMware.

**Backup concepts: definition of terms**

In order to be able to define some of the fundamental terms, a brief introduction and explanation of the core methods of data backups will be provided first of all.

**Backup**

A backup refers to the creation of one or more copies of a dataset (e.g. files or a database) with the objective of restoring a computer system or application to a previous working version or state in the event of data loss or data corruption. Backups are not simply copies. In contrast to a simple copy, a backup is characterized by the following properties:

- Backups are not dependent on the original data. No matter what happens to the original data, it will have no effect on the backup. (This is why data mirroring and data synchronization do not represent backups.)
- A backup provides a consistent copy of an application, which is required to restore the state of the application as it was at a specific time in the past.

**Image/Imaging/Image backup**

An image represents an entire hard disk or hard disk partition at the block level. An image contains the raw data of the hard disk. It contains all of the data, however, the interpretation of the file system structure is required in order to be able to view and access the individual objects (files). Imaging or image backup refer to the creation of an image.

**Disaster recovery**

Disaster recovery comprises all measures for restoring the IT infrastructure, the organization and the applications following a disaster or emergency situation. Data restores form a part of the disaster recovery measures. In this whitepaper, we define the term disaster recovery as the recovery of a server after a complete crash, which is different to a file or individual database restore on an otherwise functioning server.

**Snapshot**

A snapshot is a “frozen” image which contains the data on a hard disk at a specific point in time (similar to taking a photo or “snapshot” of a specific moment). Unlike a backup, which provides a consistent representation of an application or database, a snapshot does not provide a consistent representation of the hard disk content. In contrast to a backup, a snapshot is often not sufficient for restoring applications or a database. For example, important data or transactions could still be located in the cache of the file system or the application, which means they would not yet have been transferred to the hard disk. Data files in a relational database could have different time stamps within a single snapshot if the snapshot was created between
two file updates. This can result in a loss of database integrity, which means recovery and restart of the database would no longer be possible.

**Incremental backup / differential backup**
With an incremental backup, only data that has been changed since the last backup is backed up, irrespective of whether the last backup was full or incremental. With a differential backup, all data that has been changed since the last full backup, is backed up.

**Agent/Backup agent**
An agent is a software component that is installed on a physical or virtual server in order to carry out a specific activity. A backup agent takes over various activities related to a backup. An agentless backup is a backup that can be carried out on a server without any agents, and which can only access existing mechanisms.

**Host**
In the framework of virtualization, a host is a computer system on which a virtual machine can run (the system simply serves as a host). In VMware the ESX Server represents the host on which the virtual machines are operated. With Hyper-V, the virtual machines run on a Windows system.

**Client/Client operating system**
The term client is used to refer to a virtual machine that is running on a host. The client operating system is therefore the operating system of the virtual machine.

**Backup methods for virtual machines (VM)**
There are two different ways to create a backup for a virtual machine.

- Backing up of the data with the assistance of an agent within the virtual machine
- Creation of an image of a virtual machine on the host

The creation of a backup copy using an agent within the virtual machine (VM) is the equivalent of a standard backup. The VM is simply viewed as another server within a network. In contrast to the creation of an image of the VM, a backup with an agent enables all of the functions and convenience provided by a database agent to be utilized for incremental backups and point-in-time recoveries, thus ensuring a consistent backup. The approach of protecting virtual machines with an agent is simple and reliable, and has proven to be an extremely reliable technology over the years.

The creation of an image of a virtual machine on the host, which is also referred to as an agentless backup, only utilizes the functions provided by the virtualization technology, however. The backup is created via the backup interface of the virtualization system. No agent is therefore installed on the virtual machine. The backup software can only communicate with the virtualization software via this
interface. During this process, the virtualization software creates a snapshot of the virtual machine, which is then backed up by the backup software.

It is often suggested that no agent is required within a virtual machine, and fans of agentless backups praise the low administrative requirements. However, what may initially appear to be an advantage can lead to consistency problems later on. Deciding to forgo the installation of an agent can prove to be fatal, especially if an application needs to be restored. Inconsistent snapshots, which will be commented on later, are a perfect example here. However, we will first of all take a look at the utilization of incremental backups for backing up virtual machines before going any further.

**Advantages and disadvantages of incremental or differential backups and the utilization thereof**

Virtual machines are complete servers, which means a lot of data is involved when performing a backup. One of the most important objectives of the backup is, therefore, to reduce the amount of data to be backed up. This can be achieved by using incremental or differential backups. VMware supports the incremental and differential backup of virtual machine images. Although this enables a reduction in the amount of data to be backed up, it also makes the restore process more complicated. A number of points should therefore be considered before choosing an incremental backup.

1. In order to activate incremental backups, the Change Block Tracking option of VMware must be activated for the VM. As this function is then active at all times – and not just during a backup – the operating performance of the virtual machine can suffer.

2. If a backup only contains the changed blocks, the data blocks of one file, which was changed between backups, may be split across multiple backup sets. Multiple backup sets would then be needed to restore this one file, which would obviously slow down the entire process.

**Inconsistent snapshots cause errors during a recovery**

The backup interfaces of the virtualization system are used to back up the virtual machines. However, these snapshots cannot guarantee the successful and consistent recovery of an application.

Snapshots represent an image of a virtual machine taken at a specific point in time. They correspond to the contents of the hard disk that would exist when, for example, the system is switched off directly using the main switch. Snapshots are therefore often termed “crash consistent”. However, any changes that are stored in memory when the snapshot is created, but which have not yet been saved to the hard disk, will be lost.
When developing a backup strategy, users should therefore ask themselves the question “will it be possible to restore an application from such a state”? Even if the backup process runs smoothly, it will not be possible to verify whether an application can be reconstructed successfully or not until after a restore has been completed. With regard to databases, the probability that a complete restore will be possible is low.

Even if it was possible to restore a snapshot successfully within the scope of a test, there is no guarantee that this will work again in the event of a problem, as the application may have been in a different state during the backup.

If Windows is the host operating system, the creation of VSS snapshots may help. However, VSS snapshots are also not always sufficient for backing up databases. In order to resolve this problem, VMware allows users to, for example, invoke self-made scripts within a VM before and after the creation of the snapshot. The task of these scripts is to ensure a consistent state when restoring an application. The utilization of scripts, however, also increases complexity and therefore the potential for errors within the process significantly, as the scripts must be created by the backup software or installed manually by the user.

This process also no longer corresponds with the definition of an agentless backup. Due to the complexity of the process described above, a standard backup procedure is preferable in such a case. In the above-mentioned scenario, it would be simpler and more reliable to implement the data backup in the form of an agent while utilizing the backup interface of an application or the point-in-time recovery option of a database.

**Identifying a successful backup strategy**

The problem mentioned above merely represents a single example; there are many different configurations in which snapshots will not enable the successful recovery of data. Although snapshots provide a convenient option, they do not always fulfill the objective of ensuring the availability of important data. Yet how can users identify a promising method? In order to be able to assess the probability of success of a method, it is necessary to roll out the entire process starting at the end – i.e. from the perspective of a data restore. The first step in determining a backup strategy should therefore include the definition of objectives related to a restore.

**Objectives of a recovery**

No two recoveries are alike. There are different scenarios with different objectives, requirements and also challenges. Two principle scenarios are:

1. Recovery of an application/file
2. Disaster recovery
Recovery of an application/file
If an application or file needs restoring, the reason for this is rarely a hard disk crash or other type of hardware error. Data loss is generally the result of software problems or database inconsistencies – which, in turn, are often the result of user errors. In such cases, the application must be restored using the last intact state, which means a point-in-time recovery is required. This of course means that more than just a snapshot needs to be restored. Application-specific procedures for a recovery must be run through, with regard to databases, a combination of backups and log files are often required. Snapshot backups therefore only represent the state of an application at the specific time of creation (of the snapshot), which means they are only suitable for performing point-in-time recoveries.

Disaster recovery
If a virtual server is lost, the complete server will still need to be restored. With physical servers, however, such a recovery often represents a problem, especially if the server hardware is replaced. Virtual servers are based on abstracted “virtual” hardware. This is defined using software and does not change when a virtual server is reinstalled. In such a case it is often enough to simply restore an image backup of the virtual machine. The simple disaster recovery process (also referred to as duplication) for a virtual machine is one of the key strengths of virtualization.

Best practices
In the previous sections, we have presented various methods for protecting virtual machines and, using a number of examples, illustrated that theory and practice often differ significantly. This brings us to the decisive question regarding the best method for backing up virtual machines.

The answer: the objective must be taken into consideration while choosing a method. As previously mentioned, it is important to review the objectives of a recovery before implementing a backup strategy. And as we have previously illustrated, there are in principle two different objectives and two different backup methods. The following matrix highlights the objectives and methods as well as the advantages and disadvantages of the alternatives.
It is advisable to examine the activities “recovery of an application/file” and “disaster recovery” separately from one another and to choose the method that can fulfill the objective in the simplest and most reliable manner. The backup software should support all methods without generating additional costs.

In order to reduce the amount of data to be backed up, the dual backup concept should be supplemented with the following measures:

- Utilization of incremental backups for restoring applications or files
- Reduction in the backup frequency for disaster recovery purposes – the VMs should only be backed up if something has been changed
- Removal of inactive VM’s from the backup (as no changes will have been made to these VM’s). This function is not supported as standard by all backup software products. The availability of these functions should therefore be checked when selecting the software.
Summary
Virtual machines need to be protected. Initially, many users may be tempted to use the backup methods integrated within the solution. However, those who rely solely on these options will not be protected against problems and nasty surprises. Such problems may include inconsistencies in data or databases, which may no longer be recoverable as a result.

Those who wish to develop a reliable backup concept for virtual machines should take the following principles into consideration in order to ensure the technological pitfalls can be avoided:

1. Just because something worked during a test, does not guarantee it will work in the event of an emergency
2. Backup strategies should be based on the requirements related to restores and recoveries, and not on the simplicity of the actual backup process.

The defined best practices for simple and successful restores of data, applications and servers can provide assistance in selecting the right data protection strategy. Those who plan with foresight can avoid nasty surprises!

About NovaStor
NovaStor (www.novastor.com) is a leading international provider of software solutions for data protection and availability. NovaStor provides software, SaaS solutions and services for local and online backup, restore and retention of business-critical data. Clients include home, mobile, and SMB users, service providers as well as international corporations. NovaStor's cost-effective solutions are platform- and hardware-independent and ensure that optimal technological and economical use is gained from the customer's existing and future IT environment. NovaStor is headquartered in Switzerland (Zug), has offices in Germany (Hamburg) and the USA (CA, Agoura Hills), and is represented in numerous other countries through partnerships.