Ontrack®
by KLDiscovery

Data Recovery from Virtual Machine Data Loss

October 2023

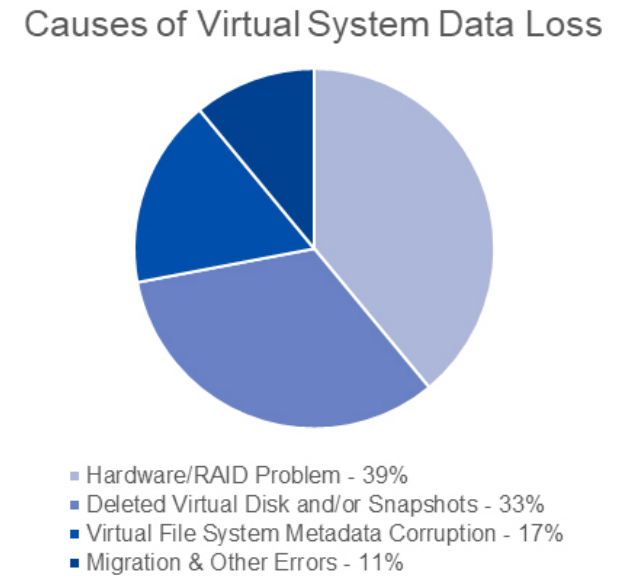
Introduction: Virtual Machine Data Loss

Virtualization technology dominates the enterprise landscape. According to Gartner, most firms report 75% or higher virtualization. Improvements in hypervisors have reduced the complexity of setting up and maintaining physical servers, greatly improved server utilization, and increased IT flexibility and responsiveness to the needs of the business. It's no wonder that the bulk of modern IT systems are virtualized.

But, whether you use VMware, Hyper-V, Citrix, Oracle or any of the other hypervisors, there is a potential downside to virtualization. In order to transform a physical server into many virtual machines (VMs), an additional software layer is added. While simplifying the admin user experience, virtualization raises the overall complexity of the IT environment as the underlying hardware is obfuscated, making it more difficult for admins to know which physical system their VMs are running on or which storage is used for a particular machine in the event of data loss. With fewer people to maintain and monitor a larger number of virtual machines (compared to physical servers), there are greater chances for problems and data loss.

Data gathered across the globe by Ontrack Data Recovery reveals that several causes of data loss incidents for virtualized environments. The leading reasons for virtual machine data loss are user error, ransomware, hardware failures and RAID corruption.

The purpose of this paper is to lay out what leads to virtual data loss and explain how global data recovery providers are able to resolve a high percentage of even the most challenging data loss situations in virtualized environments.



Source: Ontrack Data Recovery Lab Results 2009-2019

Primary Causes of Virtual Machine Data Loss

The preceding chart gives a summary of the primary causes of data loss in virtual systems.

Hardware/RAID Issues

To help prevent against data loss, modern systems will often use some form of replication of data across multiple physical drives (HDD or SSD) that is consolidated into a single logical unit. This data protection can be a hardware or software based solution. RAID combines multiple hard drives or data stripes to improve redundancy, increase data reliability and boost I/O (input/output) performance. RAID effectively fragments data across many disks and reassembles it when requested by the user or needed by the system. It takes a robust RAID system to keep track of everything and manage the data.

The hardware problems facing virtual systems are basically the same as in physical systems, such as failing drives, failing controllers, failing server components and power issues. But RAID corruption is a far greater challenge with VMs due to the nature of virtualization.

The **#1 cause** of virtual machine data loss is a hardware/RAID problem.

Unfortunately, data loss is not uncommon with RAID storage. The complexity of modern hardware and software RAID, is added to by the presence of deduplication and compression. Now factor in an additional virtualization layer and the likelihood of a fault increases. RAID controllers are responsible for mapping where all information resides across the many disks at their disposal. But if a RAID configuration becomes corrupted, files can't be rebuilt. When that happens, the interconnectivity of multiple systems can potentially cause significant data loss and downtime.

Formatting/Software Issues

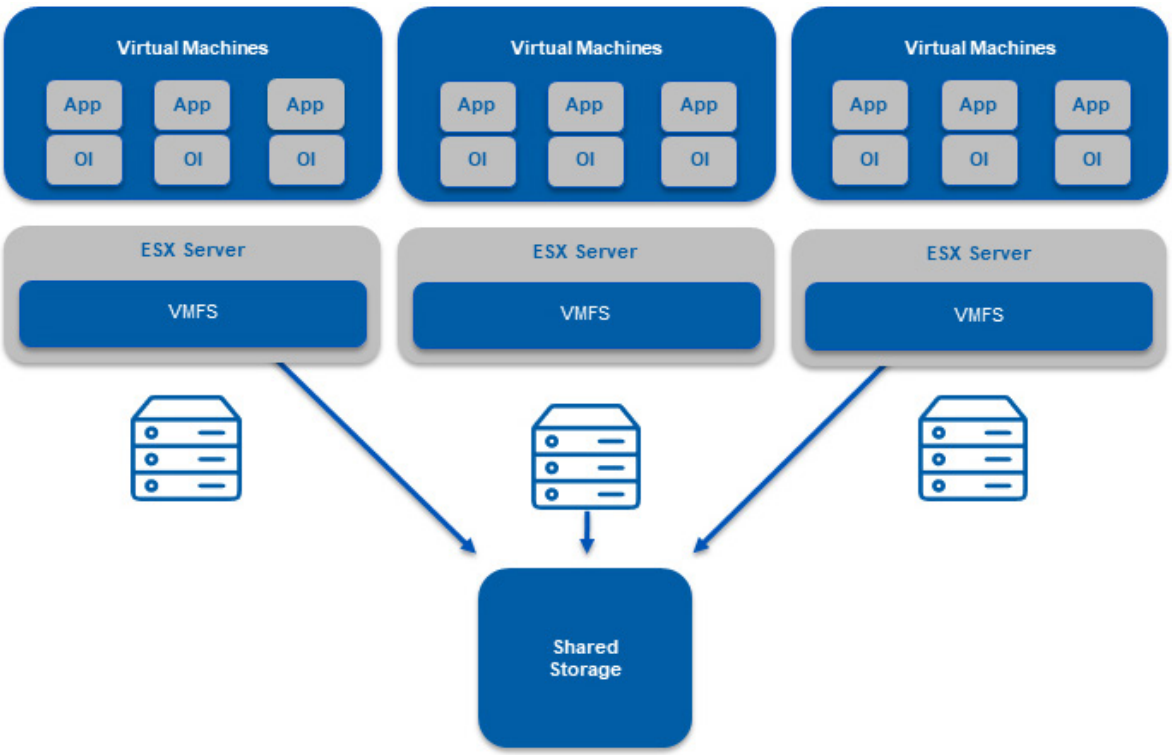
Reformatting a disk, virtual disk, array, LUN, vDisk, volume, etc (or other storage media) and re-installing software are additional causes of data loss in virtualized environments. Specific to VMs, for example, there can be reformatting at the Guest or Host level.

Corruption can also come about due to buggy patches and updates without an offline backup, poorly planned implementation of new software, integration issues and database corruption. These issues can also cause host file corruption and guest file system damage.

Thin provisioning data loss, too, should be considered. Instead of allocating all the data the VM will need and positioning the file system structures at their specified physical offsets, thin provisioning only provisions the amount of space immediately needed and adds additional blocks to the virtual disk as it grows. This can result in a more complex and fragmented virtual environment on disk. If the metadata pointers to the data are missing or damaged, it is challenging to locate the various fragments and rebuild the virtual disk. Alternatively, the mapping layer within the virtual disk may be damaged or overwritten, making reassembly extremely difficult.

Virtual File System Metadata Corruption

Yet another source of data loss is metadata corruption. Metadata is even more important in virtualized environments due to the number of layers and VMs that exist. A small problem with VMFS metadata can have serious repercussions to data availability.



User Error

Many of these sources of data loss can be categorized as user error by administrators. Access privileges allow admins the capability to delete VMs by mistake. But even if access rights are correctly managed, errors remain commonplace.

A surprisingly large amount of failures are due to virtual disks deleted by mistake, VMs being overwritten or their space reassigned. There can also be snapshot chain corruption, i.e. one of a series of snapshots is either corrupted, gets deleted or becomes unavailable for some other reason. This can foul up backups and make it difficult to recover data.

Ironically, the ease of use of modern hypervisors is causing organizations to invest in less training. Inexperienced staff are being handed responsibility for managing large and ever-growing virtualized environments. Small and mid-size managed service providers (MSPs) may not have sufficient number of experienced staff to monitor virtual environments frequently enough to catch issues as they develop. In some cases, IT admins may not initiate adequate security measures on a database or omit the documentation of changes. If encryption is enabled and a volume is deleted, for example, data becomes difficult to recover.

Employee turnover is another source of problems. The new incumbent can't figure out the intricacies of the virtualized architectures. He or she inadvertently deletes VMs or introduces changes that result in data loss. In other cases, the original flat file may be stored but nobody can find it when data loss occurs. Neglect of backups, too, is a common reason for virtual data loss.

And how about different storage, hypervisor and guest teams working in silos? One team might create a volume, another might attach the hypervisor and a guest admin then sets up the virtual machine. This type of organizational structure provides an opportunity for gaps and mistakes. Reformatting, overwriting or deletions can more easily occur.

What can enterprises do, then, when they experience data loss from a virtualized environment? There is no back or undo button. A deleted VM is gone. Backups? They are often incomplete or corrupted. Fortunately, data recovery is often possible through global data recovery service providers.

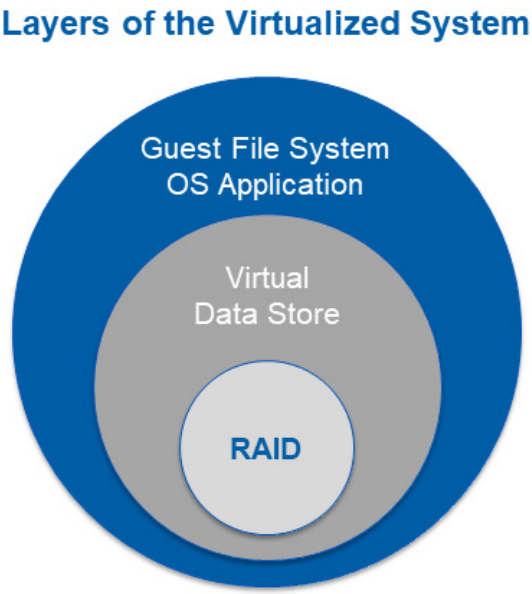
Recovery Options

The good news is that there are a great many ways to recover some and, in many cases, all of the lost virtual data. The first point of entry is at the storage level. It can be possible in some cases to directly recover data from physical drives by taking an image of the drives and reading whatever raw data might be available on the disk.

The next option is to attempt to recover data from the logical volumes (LUNs) or RAID. If the RAID controller is available, it can be used to track down the many slices of data spread across virtual disks. By determining what the configuration should be, engineers can virtually rebuild the array and gain access to the storage. If the RAID controller is corrupted, it may be necessary to emulate the RAID controller and rebuild what is missing.

The next level up, with each representing a higher degree of recovery difficulty, is the host file system level. In VMware this would be VMFS and in Hyper-V, NTFS or ReFS. In many cases, data isn't available directly at the storage level. But if the right tools are used, recovery experts can trace data from the basic storage data blocks, map it to the host level and recompile it.

If that process doesn't provide an adequate recovery, additional tools can be employed to extend further into the guest file system level. By investigating the virtual file system, data recovery specialists can sometimes find data that would otherwise be lost. Finally, it is possible to reach into the guest file level and access data lurking in application files such as SQL, Exchange, SharePoint, Oracle, Office files, ZIP files and more.



What it takes is an understanding of each level and knowing what might be available where. Those well-versed in storage architectures can track down data that seemed lost by finding pieces of it in one level and other parts in another level.

This is perhaps best understood by looking at a RAID example. It is a fact of life that drives will eventually fail. If RAID 1 or greater is being used, a new drive can be installed and the data storage map rebuilt without data loss. But what if the drive failure exceeds the redundancy capacity of RAID? To recover data in this case, it is usually necessary to bypass any physical failures that may have occurred, reconstruct the RAID file system, and assess the various layers and complexities of any virtualized architecture that may exist. This often makes a recovery extremely challenging and time-consuming. However, with the right provider, in many cases, recovery efforts can be successful. Make sure the provider has the tools and expertise, as well as direct partnerships with storage vendors.

Virtual Machine Data Recovery by Ontrack

With over 30 years of global experience in data management, data recovery, secure data erase, ediscovery and computer forensics, Ontrack has recovered virtualized data for thousands of enterprises.

Engineers image drives and read raw data on the disk, determine what the configuration should be and then virtually rebuild the array and gain access to the storage. To do so, Ontrack has developed tools such as those that emulate the RAID controller and rebuild what is missing. The company has also developed a wealth of additional tools to accomplish such things as to prevent further writing of data onto volumes, address the complexities of virtualized files systems and more.

Ontrack'sdevelopment team continually updates its tools for the latest virtualization platforms and storage environments. Thanks to its knowledge of the various storage media, operating systems and underlying storage architectures, Ontrack offers comprehensive services for data recovery, as well as follow up services for intelligent backup and data management.

Let's take a look at a few examples:

Accidental wipe of a NetApp System

A Korean Managed Service Provider (MSP) had a client with a NetApp FAS8060 system containing 161 x 900GB SAS HDDs. They were arranged in two separate aggregates (68 drives + 93 drives). The client presented three 468GB Fibre Channel LUNs from each aggregate to a production Sybase server. Six LUNs in total were combined into a single disk pool with three logical volumes carved out of the pool.

An engineer at the MSP attempted to make configuration changes to a NetApp filer. However, he inadvertently started a wipe command on some LUNs, effectively wiping 45 GB of data from the Sybase server. The MSP potentially faced loss of a client contract and possible liability costs.

Ontrack was brought in via phone consultation. The MSP was instructed to bring the aggregates offline to avoid any further overwrite damage. This was accomplished 12 hours after the time of the original data loss event. The client was instructed to present all 161 HDDs from both aggregates to a single Windows machine. This was then connected remotely to Ontrack's Remote Data Recovery server. As both aggregates had the same name, it was not possible to easily rebuild them. As a result, the drives had to be sorted into aggregate groups and manually rebuilt to a point in time as close as possible to the time of the incorrect wipe command.

At that stage, another problem arose. The logical volumes were used as raw storage by the Sybase server. This made it impossible to extract the internal data directly. The workaround for this was to extra all six LUNs as flat files and coordinate with NetApp support to present these LUNs back to the Sybase server. The recovered logical volumes passed integrity checks and were made operational with no loss of data.

Data Loss due to Reformatting of VMware

The IT team for a food production company based in Singapore mistakenly removed a VMFS datastore LUN from the VMware ESXi host and attached it to a Windows server. That led to the LUN being reformatted to the NTFS file system. This action corrupted front-end VMFS metadata, which brought about the loss of all virtual machines in the datastore. The company called Ontrack for help. Our engineers were able to rebuild the VMFS structures to regain access to the VMs stored on the system. Several VMs were recovered intact, while others required additional repairs to the internal guest file systems and extraction of the resulting data.

Deletion of VMs

A health services provider in Australia mistakenly deleted seven thin provisioned VMs from a production datastore. Due to the sensitive nature of the lost data, the company immediately called Ontrack and requested that our engineers come directly onsite. Once they arrived at the data center, they were able to recover all the VMs, though some damage was apparent. At that point, additional repairs were performed to the guest file systems of each VM using Ontrack's proprietary tools. This process enabled more critical internal data to be extracted to external storage. Although there was some data loss, the bulk of the data contained in the VMs was recovered.

Conclusion

Virtualization may save time and eliminate complexity from the user view. But it comes with a unique set of challenges, one of which is a rising incidence of corruption and data loss. Whether through volume corruption, deleted volumes, deleted or corrupted virtual backups, RAID and hardware failures and deleted or corrupt files within virtualized storage systems, data loss is a reality for anyone managing virtual systems. Backup is necessary to safeguard enterprise data, can help but it is far from foolproof.

As the world's leading data recovery supplier, Ontrack stands ready to provide comprehensive data recovery services for virtual servers and storage. Ontrack Data Recovery Services can:

- Recover data from virtually any type of data storage device – from hard drives, flash and SSD drives to servers, NAS, SAN and virtual systems.
- Minimize downtime through fast turn-around times, emergency service options and the industry's only lab-quality remote data recovery service.
- Report all recoverable files and the condition of each file as part of the evaluation before you pay recovery fees.

BRO ONT US_DR Virtualization_Oct 2023

Ontrack®
9023 Columbine Road
Eden Prairie, MN 55347

800.872.2599
www.ontrack.com