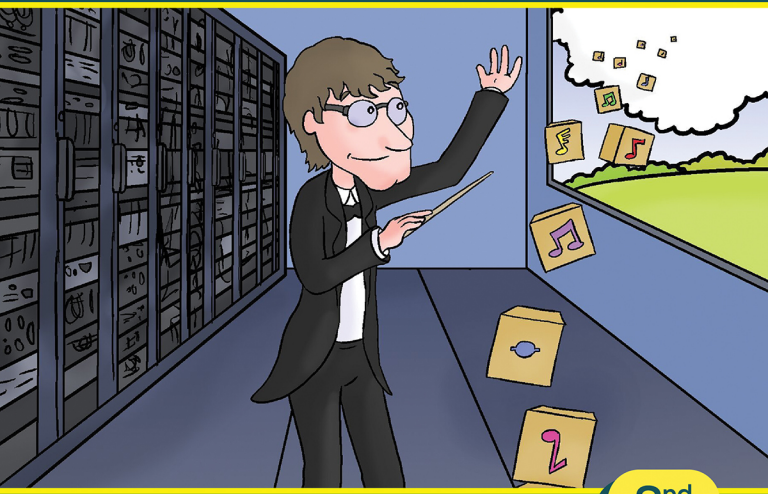
ConversationalGeek®

Conversational
Disaster Recovery and Orchestration for VMWare

Yves Sandfort (Cloud Evangelist and VCDX)
& **Matthias Eisner** (VMware Architect/Designer/Certified VCI Instructor)



Learn about:

- VMware disaster recovery approaches and options
- Data availability solution considerations
- Data recovery reminders

2nd MINI Edition

Sponsored by

COHESITY

Sponsored by Cohesity

Cohesity radically simplifies data management. We make it easy to protect, manage, and derive value from data – across the data center, edge and cloud. We offer a full suite of services consolidated on one multicloud data platform: backup and recovery, disaster recovery, file and object services, dev/test, and data compliance, security, and analytics – reducing complexity and eliminating **mass data fragmentation**. Cohesity can be delivered as a service, self-managed, or provided by a Cohesity-powered partner.

COHESITY

For more information, visit www.cohesity.com

Conversational Disaster Recovery and Orchestration for VMware (Mini 2nd Edition)

by Yves Sandfort and Matthias Eisner

© 2021 Conversational Geek

ConversationalGeek

Conversational Disaster Recovery and Orchestration for VMware (Mini Edition)

Published by Conversational Geek® Inc.

www.ConversationalGeek.com

All rights reserved. No part of this book shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher. No patent liability is assumed with respect to the use of the information contained herein. Although every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions. Nor is any liability assumed for damages resulting from the use of the information contained herein.

Trademarks

Conversational Geek, the Conversational Geek logo and J. the Geek are trademarks of Conversational Geek®. All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. We cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Warning and Disclaimer

Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied. The information provided is on an "as is" basis. The author and the publisher shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or programs accompanying it.

Additional Information

For general information on our other products and services, or how to create a custom Conversational Geek book for your business or organization, please visit our website at www.ConversationalGeek.com.

Publisher Acknowledgments

All of the folks responsible for the creation of this book:

Authors:	Yves Sandfort and Matthias Eisner
Project and Copy Editor:	Pete Roythorne
Content Reviewer:	Douglas Ko

The “Conversational” Method

We have two objectives when we create a “Conversational” book. First, to make sure it’s written in a conversational tone so that it’s fun and easy to read. Second, to make sure you, the reader, can immediately take what you read and include it into your own conversations (personal or business-focused) with confidence.

“Geek in the Mirror” Boxes

We infuse humor and insight into our books through both cartoons and light banter from the author. When you see one of these boxes, it’s the author stepping outside the dialog to speak directly to you. It might be an anecdote; it might be a personal experience.

Within these boxes I can share just about anything on the subject at hand from the author or other SMEs. Read ’em!

Disaster Recover and Orchestration for VMware

What's comes to your mind first when you think of a disaster? Probably, some sort of natural disaster, like a flood, tornado, or an earthquake, right? Maybe even a fire. However, in the modern digital world the bigger disaster risk is being hit by a

cyberattack. Once we consider all the different types of possible disasters, we find it's no longer a question of "if" disaster strikes, but more "when".

Let's start simple, with backup… The basic concept of a backup is that we create a copy of our valuable data. In the event something goes wrong, we can recover and use it. But one of the most common oversights with backups is that we never check that they really work. Like the airbag in our cars, or the insurance policies we have, we pay for something we hope we never need – and typically don't validate that it works. However, in contrast to the airbag in our car we can virtually "clone" our IT and test drive our backup without the need to "crash" the actual production environment.

We're not suggesting you go to your car and test the airbag. But you should test your IT backup and validate it truly works.

What else is important to know about backups? The 3-2-1 basic backup rule is something you might hear

often. We keep three copies of the data on two different media types, and one is (ideally) at a different location. The cloud has facilitated this type of preparedness greatly, supporting varied media and offsite locations; we discuss this later in more detail, as it is quite handy.

But this book's focus is disaster recovery and orchestration. *Why would I need that? I have backups… right?* Unfortunately, backups alone do not ensure recovery after a significant incident. While a disaster recovery process requires some version of data copy (most likely a backup or replication), it is more about the process, including how to get back online and how long this will take. As we all know, recovery isn't just powering on a few workloads; as modern data center infrastructure is complex.

Now imagine a few hundred workloads and VMs to recover (or even thousands). All of them have specific dependencies and complexities, which we usually cover in some manual runbooks or documentation. While virtualization and automation made the infrastructure "disappear,"

the workloads often have more dependencies than there were in the past.

So, a disaster hits, what's next? Let's hope you have a backup to recover from. Even then, you will likely experience chaos. Someone needs to decide which workloads take priority, what their dependencies are, etc. Bringing infrastructure back online can be difficult if all of this isn't documented. At a conservative estimate, this can take days, weeks, or longer.

And depending on the type of disaster it might be necessary to take an older version or a mixture of versions back into operation. This is especially true with ransomware and cyberattacks which might have been in our systems for more than 30 days, so the newest update is not always the first (or safest) choice.

This is where an automated disaster recovery solution is hugely beneficial. It's going to combine the data copy you have with contextual knowledge of dependencies, in a location suitable for your workload after a disaster. Depending on the disaster, this could be a different platform

altogether. In many cases, your workload needs not only to failover to the DR site, but also failback once your primary site is back up and running.

VMware Disaster Recovery Approaches/Options

There are a variety of approaches and options for disaster recovery in VMware shops (including the built-in VMware options and features like vSphere Replication with Site Recovery Manager). VMware also offers well-documented APIs, enabling third-party vendor solutions to provide greater value to the whole infrastructure.

While this primarily impacts the IT operations team, we should never forget about the business itself, and the end users, who expect a similar or identical experience after a recovery as they had before it.

With VMware, virtual machines are just files, and that provides the amazing capability to replicate those files to a different location. In this case, replication means copying those files after creating a consistent state.

Having a consistent state of virtual machine files at a different location enables various possibilities for disaster recovery; most importantly, a tight integration into the vSphere stack for smooth operations or the ability to use any cloud even if it is only temporary.

What is important for Disaster Recovery?

First, it's important to classify what a disaster is and its impact on the business. Based on that, you need to decide if it's necessary to fail over to the disaster site. Initiating a failover has its consequences. If you decide to fail over, it's important that the scenario has been practiced before to be sure what needs to be done and to have a solution in place that enables those involved to fulfill their jobs as quickly and consistently as possible.

Think about it like an aircraft pilot; they practice for disasters constantly in a simulator (similar to your DR test run). You should do the same for your DR scenario.

For disaster recovery, it makes perfect sense to use a solution that provides testing capabilities and orchestration features. If a failover is necessary, the

solution should need just a few manual interactions to bring the data center back online (ideally only to trigger the process).

Isn't replicated storage enough?

Honestly, this question can be answered with a simple "no." Replication comes in two flavors: synchronous and asynchronous. Synchronous is mainly used within a single data center if one of the storage systems fails to ensure continuous data accessibility. Synchronous replicated storage solutions are great inside a data center if one of the systems fails. Asynchronous is mainly used to ship data to a different location with a defined time frame. Sometimes this is called a "trailing data center."

Nevertheless, even if asynchronous replication is used, it doesn't mean that you're able to do an orchestrated failover; the only thing you have is your data at a different location. That brings up questions like "*how do I get everything back up and running*?" In that scenario, orchestrated disaster recovery solutions come into play. Perhaps they integrate with your replication solution, or they may be shipped with their own mechanism.

But my storage vendor does snapshots…

Snapshots are great for some specific use cases, enabling a very fast and easy recovery to a certain point in time. Still, neither storage nor VMware snapshots are a backup or disaster recovery solution; they're suitable to restore to a certain point in time. Of course, snapshots might be used to replicate a defined and consistent state of a virtual machine to an offsite location, but the magic word is still replication. Keep in mind that a snapshot might also be corrupted because of a software or hardware issue.

What happens if disaster strikes between snapshots?

This is where the best Disaster Recovery solutions differentiate themselves from the rest. The best disaster recovery solutions understand that mission-critical applications need to be *continuously replicated*, and only a few vendors offer this in the form of continuous data protection. They leverage the vSphere API for I/O filtering (VAIO), and maintain a journal of each change in a VMDK on the primary, which is replicated faithfully

on the secondary with almost zero lag. For mission-critical applications, this is the way to go.

How do I modernize my DR solution, if storage arrays are not enough?

The key differences between modern and legacy DR is software-definition, cloud integration, and a platform approach. Modern DR plans are fully software-defined – protection, replication, and orchestration are all software-driven, and preferably run from the cloud. Moreover, modern DR solutions always offer cloud as a failover destination, with the inherent capability to spin VMs to a cloud-native format. A small selection of vendors offer the entire DR solution as SaaS (first-party DRaaS).

Finally, the best DR solutions are platforms. Rather than forcing the customer to write custom scripts across vendors and infrastructure, the best modern vendors can do backup, replication, and DR orchestration, as well as target storage all on the same platform. The benefits of this cannot be overstated – a single policy engine underneath a

single UI to protect all applications is critical and massively helpful when you need it most.

Data Availability Solution Considerations

First, there is no ideal solution. While every vendor tells you theirs is perfect for everything, the truth is far more complex. Nearly every solution has its advantages and disadvantages. Let's look at some of them in more detail.

Choosing a disaster recovery solution

No matter whether we discuss backup or disaster recovery solutions, each vendor has their own way to store data, most likely with their own secret sauce involved. Although some allow you to access data over something like NFS, others will label their method of storing data the "industry standard." It's clearly an advantage if I can use just one solution to satisfy my backup and disaster recovery needs; I might even be able to use the same procedures for DR testing as I can for recovery tests of individual workloads. Regardless, the solution should not add additional complexity, preferably making day-to-day operation more straightforward. As mentioned

above, software-defined is key, not only for DR but the complete data-lifecycle. The more operations that are done in software, the more resilient the system is, and the more nimble the recovery process will be.

Consider the risk of vendor lock-in, both with backup and DR and for the virtualization platform, which can be difficult if you want to recover into the cloud. Some vendors have solutions to allow workloads to run in the cloud on a different solution, but even though this may be technically feasible, be sure to check your applications and OS.

Platform dependency

While you back-up from one platform, you may recover to a different one. Imagine you have a data center on premises and – while you may plan to recover locally from individual outages – your disaster recovery plan may be to run the workload in the cloud. This might involve migrating the workload into a different VM form factor (i.e. from VMware to Amazon EC2 or other clouds). In such a scenario, your backup solution needs to be recovered into the cloud. *And what about your Disaster Recovery solution?* Your cloud might

require workload startup, network definitions, and other changes as well. The more software-defined your solution is, the easier it becomes to pre-define these DR components ahead of a failure scenario. Software-defined solutions also allow for easier mobility between on-premises, the cloud, or even between clouds.

Automation

Let's face it, you most likely designed your backup solution a while ago, it's mostly automated, and probably runs fine. Recovery and Disaster Recovery is mostly an afterthought and not automated, which is like running your car with an airbag which will only inflate in an accident if you press a button.

At least the recovery (and we mostly mean disaster recovery) needs to be pre-scripted and automated, as the disaster will already have a massive impact on operations. This is not the time for manual interaction, especially when most enterprises need to fail over hundreds or thousands of workloads in a disaster, often with very complex dependencies.

Additionally, maybe some backup files are unreadable, perhaps certain services don't start as

expected, databases need to be recovered etc. So, ensure that you automate whatever can be automated. Consider also that certain workloads may require recovering to a different point in time.

Remember: *the airbag is also not manually inflated.*

Disaster recovery (source of data)

When a disaster strikes, we need a reliable data source. *Why not leverage our backup?* Yes, that is the perfect starting point! Depending on your RPO/RTO definitions, some workloads might start off as active copies, others will be started from a recovery process.

Modern backup solutions are far more flexible than the old legacy tape solutions, due to compression, deduplication, and defragmentation capabilities. *Why is that important?* We can more easily jump to different points in time and start workloads from our backups. Legacy solutions often required us first

to do a full recovery to see if the data was usable at all. The newest iterations allow us even to start natively out of the backup solution.

Compression and de-duplication

As more and more backups run on either hard disks or flash storage (like SSD), compression and de-duplication becomes more valuable. With legacy tape-based solutions, backups were often time-consuming. Now, we can have very short backup intervals for our critical data, leveraging compression and de-duplication to reduce actual space consumption on the target platform.

Policy-based protection, rules and agents

When we talk about backups and disaster recovery, we often think about when, and how often, a backup needs to be done; *but there is more to it*. Ideally, there is a policy associated with the workload which defines not only how it is protected but also what is required to bring it back in case of a recovery. Think about it as a contract which describes how it is protected and how it should be recovered. This contract or policy is linked to the workload and maintained with the workload. Such a

policy might also contain vital information, like dependency on other workloads, special information about recovery, and more.

When choosing a solution, it is important to understand the policy system. While many vendors switched to a policy-based approach, some take a very complicated approach to this. Remember, these policies are important for your runbooks and if they are complex to maintain, there is a high chance they won't get updated on business rule changes.

Recovery

When the day comes that your workload needs to be restored, you need to know where it should go or – even better – where it goes doesn't really matter. A perfect solution allows not only a simple recovery in the spot where the original workload was operating, but also gives you the option to recover on different environments, like the cloud. On top of this, as we prepare to deal with cyber-attacks it is no longer enough to be able to get a complete machine backup, we also need to be able to validate if individual files have been tampered with and when, allowing you to get not only a

known good status back, while also ensuring consistency in data and protection from bringing back an already contaminated version.

Imagine you have lost your complete data center; *wouldn't it be handy if you could just press a button and bring it back online in the cloud?* That's what modern recovery solutions offer, no matter if we're talking full disaster recovery or individual workload recovery. However, you should also check that your solution is capable of mass recovery in parallel. This allows you to bring your business back online faster, reducing the negative business impact the disaster might otherwise have.

It shouldn't matter whether you lost a few workloads or a complete data center, your DR solution should allow you to adapt easily to the actual need at the point of a disaster or a recovery. At the same time, if this is policy-based you know all dependencies immediately, so the recovery can be instantaneous and fully operational. Think about a critical database holding all your customer data; a recovery does not only recover the system, but also ensures the web and application servers giving you access are operational.

Data Recovery Walk-through

Let's start with a simple story I experienced a few years back. I was on-call support for some mission-critical customer when the phone rang at 3:00 AM. The customer on the other side was very calm, but asked if I could guess where he was. I said "*no*" and that I, honestly, wouldn't really care at that time… *It was all silent on the other side…* Then he said that he was in the middle of the data center, and I immediately understood the issue. It shouldn't be silent and he was *not* really prepared.

Why, you might ask? It's simple; he had a data center with redundant power, redundant cooling, fire suppression – everything you could think of. Still, he was in a massive room of silence. This is why you need to plan for disasters, as there is this little devil sitting in the details and identifying that one little weak point. Trust me, it's not a question of *if* it will happen, it's just a question of *when*.

We tend to create a high-level flow chart of what to do first, to help people in case of such a disaster.

The first thing to do in a disaster scenario is *assess the situation*. What is affected by the outage? Is it individual workloads we could recover, or maybe just some virtualization host? A single workload, perhaps? Maybe a complete datacenter?

Don't get confused with the *why* question… This is something you can worry about later. Think like a pilot flying a plane. You need to quickly assess the impact and what options you have – it doesn't matter if the engine failed because of a bird strike, a technical fault, or just faulty usage; it's more important if you have a backup engine or other fallbacks you can use.

So, just individual recovery or disaster failover? This is not always simple to answer. Sometimes we tend to spend time trying to find the cause and fix it, but what if that is not feasible? Would recovery of a

workload help to ease the situation, or do we need to declare a disaster? The latter usually means a different recovery procedure and is most likely not so easy to revoke. This is why it normally requires a manual startup process (although the recovery itself should be automated). Often, I talk to people who expect a disaster recovery to be automated, like a simple system failover. While modern technology allows us to fail over from a single workload failure (usually a large segment or complete data center outage) this requires a more senior decision. The recovery will force us to operate out of a backup location for a while, making operations more complex.

What comes first in a disaster recovery? Once we decide to declare a disaster, we depend on something we call runbooks. They provide us with the necessary structure and procedure to get back into "normal" operations. Ideally these runbooks are automated, so we don't have to do all of this manually. Trust me, there will usually be enough to do while the recovery procedure runs.

RPOs and RTOs

This is where it gets interesting, as these define how quick we get back into operations. Better yet, they define how quickly we should be back in operation and how much data we are allowed to lose. Trust me, once the big day comes, people will be trying to force you to change priorities to make miracles happen. Stick to your plans, that's why you designed a disaster recovery plan in the first place. If it doesn't match your actual demand, then this is something to work on, but don't get hung up on changes, *they will make it worse*.

On premises or into the cloud?

A few years back companies were planning either for in-place recovery or with a secondary data center; lately, this has changed. The cloud has given us a new choice. While we operate in a local virtual environment, we can recover into the cloud, enabling immediate recovery at much lower costs.

Why would we failover into a secondary site?

When we are hit by a disaster, we might not be able to recover in the primary site. In the case of a natural or other physical disaster the site might just

be unusable. In case of a cyberattack the forensics might still need to take place and people might try to recover unchanged, untampered, or unprotected data (which hopefully you shouldn't have in the first place). All of the above are reasons why companies in the past had pre-equipped secondary – even tertiary – data centers up and running to recover into.

In the cloud generation we live in, it no longer makes sense to take this approach. Maintaining an additional site and keeping it up and running is not very cost effective. Organizations of all sizes and across all industries can move their DR sites into the cloud, either fully reserved and ready or consumed as a service.

Warning: Be aware that cloud resources are not endless. While people tend just to enact the minimum resources, disasters can incite a high demand for resources, and you might not be the first in line. *For example*: Imagine you just booked storage capacity at your favorite cloud platform, a natural disaster wipes out data centers for hundreds of companies (earthquakes, for example), and – all of the sudden – you are not the only one

in need for resources. Look for a solution that enables you to choose different cloud regions, so you don't get stuck in a recovery traffic jam.

How to plan for cloud recovery

Have your most pressing resource demands reserved all year, ensuring you have them if you need them. Test drive dependencies; it doesn't help if your CRM and ERP are back in operation while your access technology is missing. And ensure, if you decide to recover to the cloud, that the format conversion – both for failover and later failback – are validated and disaster-ready.

Returning to our earlier disaster example, since the impact assessment was simple (somehow, you lost power and nothing worked anymore) you might believe a simple "power on" is the best solution. *In this case*, the customer decided to bring the data center back vs. a full disaster recovery into their failover data center. *Why?* They were scared that the data in the other location was not complete and the runbooks were not current enough.

So they tried to bring the data center back up and running. This was not the best decision, as it took

them several days before some of their most critical workloads were back. Why? They were not fully tested and trained so had to try out a lot.

What could have prevented a lot of the unforeseen? Nothing? Just accept that you will not be able to build a completely available solution. Accept your risks and be prepared. Test what could happen multiple times and train your team how to react and handle the situation. The cloud makes this a lot easier, with physical resources to test against a failover instance and test drive a disaster while still being fully protected at a fraction of the cost.

The Big Takeaways

Be prepared: It's not a matter of "if", but "when" you need your disaster recovery in place

Reduce complexity: The cloud makes it easier to consume a disaster recovery solution to your needs. Whether you consume it as a service or hold a complete site up and running, it's easier than an additional on-premises site and most often more cost-effective.

Plan early: Trust me, if you plan to create and update your disaster recovery solution in three months, you're most likely to be hit in two months. Why? Because that's this little devil I mentioned.

Automate everything: A disaster is stressful enough. Ensure you have most of the process automated, making it more reliable and reducing the pressure.

Be realistic: No system is perfect, and there will be challenges to overcome.

Plan for the future: How do you get back into normal operations? Is there a failback option?

Tests can be fun: Make it a fun event, press that red button, and do a test failover. The more you test, the more relaxed you can be. This is not just an IT exercise (users need to work after a disaster). So, ensure they are included in the test; you don't want to have a disaster and find that all your plans for access (and more) don't work.

COHESITY

Disasters are Unpredictable — Recovery Doesn't Have to Be

Disasters, including cyberattacks can devastate you. While complex hybrid and multicloud environments make it more difficult to recover your applications. If left unchecked, you face a higher risk of data loss and downtime.

To achieve data resilience and simplified operations, you need a solution that automates complex disaster recovery tasks and provides failover to another site or the cloud to prevent data loss and downtime.

Achieve near-zero downtime, avoid data loss, simplify operations, and eliminate costly infrastructure with Cohesity SiteContinuity disaster recovery and disaster recovery as a service solutions.

Learn more about Cohesity SiteContinuity at Cohesity.com/products/sitecontinuity

© 2021 Cohesity, Inc. All rights reserved.Cohesity, the Cohesity logo, SnapTree, SpanFS, DataPlatform, DataProtect, Helios, and other Cohesity marks are trademarks or registered trademarks of Cohesity, Inc. in the US and/or internationally. Other company and product names may be trademarks of the respective companies with which they are associated. This material (a) is intended to provide you information about Cohesity and our business and products; (b) was believed to be true and accurate at the time it was written, but is subject to change without notice; and (c) is provided on an "AS IS" basis. Cohesity disclaims all express or implied conditions, representations, warranties of any kind.

Recovering from a disaster is about more than simply having backups; it's about the recovery process in place to get your VMware environment back up and running. This book will walk through key aspects of that process with a focus on automation for a smoother recovery should disaster strike.

About Yves Sandfort

Yves Sandfort has 20+ years of experience in the IT industry and is a VMware Certified Design Expert (VCDX) and a VMware Certified Instructor L2.

About Matthias Eisner

Matthias Eisner also has 20+ years of experience in the IT industry and is a VMware architect, designer and trainer with multiple VMware certifications including VMware Certified Instructor (VCI) L2.

ConversationalGeek®

For more books on topics geeks love visit

conversationalgeek.com