WHITE PAPER

Drowning in Data – How to Back up More Than You Can Handle

By Nick Cavalancia

UNITRENDS

TABLE OF CONTENTS

UNITRENDS

You're already charged with backing up just about everything in your environment.

You have an insurmountable volume of data with backup requirements that just aren't realistic. So how can you stay afloat but also get to a point where backing up everything that's requires is not only possible, but is actually feasible?

You're already charged with backing up just about everything in your environment. And it's unlikely your company will ever decrease the amount of data, applications, and systems you need to back up – when you consider the sheer number of data sets, applications, services, servers, operating systems that need to be recovered, there simply isn't enough expertise to properly provide recovery coverage. Backups are easy enough, but when it comes time to recover, you need to know more than just which button in your backup solution to press – there may be sync issues, database inconsistency issues, etc.

So, how are you supposed to protect this never-ending, and constantly growing, sea of information while still being able to recover it at a moment's notice?

You might try to figure out where to place the blame – there are a few suspects:

- **Backup Methods** – There are lots to choose from: file, image, p2v, v2v, etc. And the method you choose can put constraints on you, like doing a file-level backup but needing to recover an entire system.

- **Backup Granularity** – Your backup solution may provide application-aware selections, but if you're backing up just the related files on a server, the granularity may be doing you a disservice.

- **Backup Frequency/Speed** – A lot goes into this - hardware, use of the cloud, available bandwidth, and beyond. All of these restrict the backup (and recovery) window possible.

- **Backup Capacity** – If you don't have enough storage in the first place, you're really feeling the pressure.

In reality, none of these issues are truly to blame. It's likely the issue is more about organizing your backup and recovery strategy than anything.

Since drowning is not an option – let's skip the overwhelming entirety of all that needs to be backed up and recovered, and instead focus on three basic blocking and tackling steps that will bring some order to your backup chaos, beginning with the simplest of steps – dividing and conquering.

Divide and Conquer

If you were to take a poll of the various executives within your organization as to which applications and systems were critical to the business, you'd probably end up with a list that included just about everything. But the reality is, it all can't – because when everything's critical, nothing is.

To get anywhere with your backup and recovery strategy, it will be necessary to break up recovery into smaller, more tangible pieces.

Think about it, if everything needs to be up and running in, say, 15 minutes, then you simply elevate the "norm" to "critical, where no systems take higher priority than any other. Even so, in the face of a disaster where each and every system in that exec list is unavailable, you simply can't be back up and running in 15 minutes. And it's this exact thinking that brings you to a state of being overwhelmed, unable to deliver anything.

Since you can't address it all at once, don't. To get anywhere with your backup and recovery strategy, it will be necessary to break up recovery into smaller, more tangible pieces where you can begin to address each one individually and determine what course of backup action is necessary.

Start with One

The very first step in addressing all that you need to backup is to pick one thing that needs to be recovered. It can be a set of files, an application, one system, a single VM – something that represents a subset of the entirety of what you're supposed to back up. Without taking this first step, you'll remain in the overwhelming wave of everything that needs backing up.

Once you have that single data set, break it down using a few considerations. These considerations should be used to help you determine what kinds of backup method and frequency is necessary to accommodate your selection.

There are a few considerations when it comes to your data that will impact when and how you back it up.

- **Location** – Where is the data set currently located? Is it on-premises or in the cloud? And if in the cloud, is that public or private? Knowing where you'll need to back up from will play a big difference when you begin to think about how to recover the selection.

- **Criticality** – While this is somewhat subjective, you should determine how is the business affected if the selection is unavailable or goes down. You may want to think in terms of your SLA, the business cost of loss, and is there zero tolerance for downtime for the data set in question.

- **Data Characteristics** – There are a few considerations when it comes to your data that will impact when and how you back it up. Size of the data set impacts backup window sizes (which may or may not be possible) and how often the data changes impacts whether changes are reflected in a given backup (e.g. a backup with a retention of 30 days will never include data that changes every 35 days). The result here may be the changing of backup type (e.g. from a full image to a file-based differential backup) or choosing a smaller data set.

- **Recovery Objectives** – Your recovery point objective (RPO) defines how far back you need to recover. It could be 5 hours, 5 days, or 5 minutes – all impacting the frequency of backup. Your recovery time objective (RTO) defines how much time it will take to recover. These usually align with criticality pretty neatly.

The goal is to take all these considerations and determine what's possible from a backup perspective. An RTO of, say, 5 minutes for a complete recovery of an entire server located across a WAN is likely impossible, requiring you to rethink either the RPO, or opting for a smaller data set to recover. Once you have properly identified a reasonable and backup-capable data set, start the process over for the next data set in the list.

When this process is over, you should be left with the definitions of what critical data needs to be backed up and in what order of priority. But that doesn't necessarily mean you can recover.

Recovery... then Backup

Defining what needs to be backed up doesn't guarantee recoverability. To do this, the next step is to start with what needs to be recovered and work backwards to the data set you defined to make sure that critical data you backed up can actually be usable within the RTO defined.

Take for example an Exchange server. You define the backup data set as the Exchange databases. You walk through the considerations and determine you can back up those databases daily. But when it comes time to recovery, you need the rest of the server. This simple check and balance is there to ensure you don't just focus on having data to recover, but so you can actually recover.

The first step is to identify the disaster.

So what's the right way to plan for recovery and align with the backup data sets defined?

The first step is to identify the disaster. Remember a "disaster" can be anything including corrupt files, hardware failure, fires, or even a hurricane. In essence, it's the scenario you want to be able to recover from. You'll need to determine which are most likely – each may require a different backup approach. For example, the corrupt Exchange databases only require a recovery of the databases, but the failure of a server requires a bare metal recovery and a database recovery. Each of these is still protecting the data set you started with, but defines additional data sets necessary to reach the recovery goal.

A simple way to think about identifying the disaster is to consider what will be lost (that needs to be recovered). Looking at **Figure 1**, you can see a disaster severity index of sorts that lists everything from a loss of data, all the way up to a loss in operations will facilitate the need for a specific type of recovery.

You can use this index as a guideline, but it's not perfect, as every recovery can be unique. Like selecting data sets, there are a few considerations for choosing the right recovery method as well.

- **Local or Cloud Recovery Source** – You need to know where your backups reside – on-premises, at a secondary CoLo site, or up in the private or public cloud.

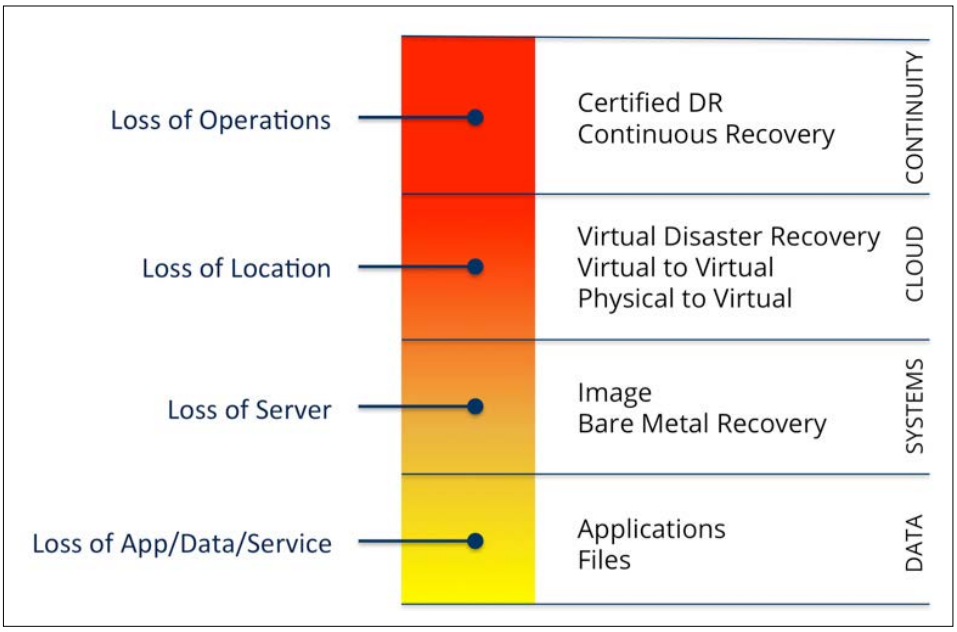

Figure 1: *A Disaster Severity Index*

Many backup solutions also employ technologies to compress and optimize the transmission of backup data.

- **Local or Cloud Recovery Target** – Pairing your source with your target is critical; as it will define whether the disaster you're preparing for is even possible. For example, if you employ a cloud backup source and only plan for a local target, and you have a flood in the office, you're probably not going to recover quickly.

- **Service Availability** – This one ties in the previous two when you're considering a scenario, for example, this time dependent on being able to recover from a local disk-based backup and that same flood happens.

- **Speed of Recovery** – More than just bandwidth (although very important), many backup solutions also employ technologies to compress and optimize the transmission of backup data. Understanding how fast the type of recovery you want (e.g.: a full server image) on a GB/min basis will give you a realistic expectation of how feasible the assumed recovery type is.

Putting it all together

Think of this like a grid – data sets (including the new ones defined by your recovery goals) on the left, disasters on the top, and you place the methods of recovery where each intersects, as shown in **Figure 2**.

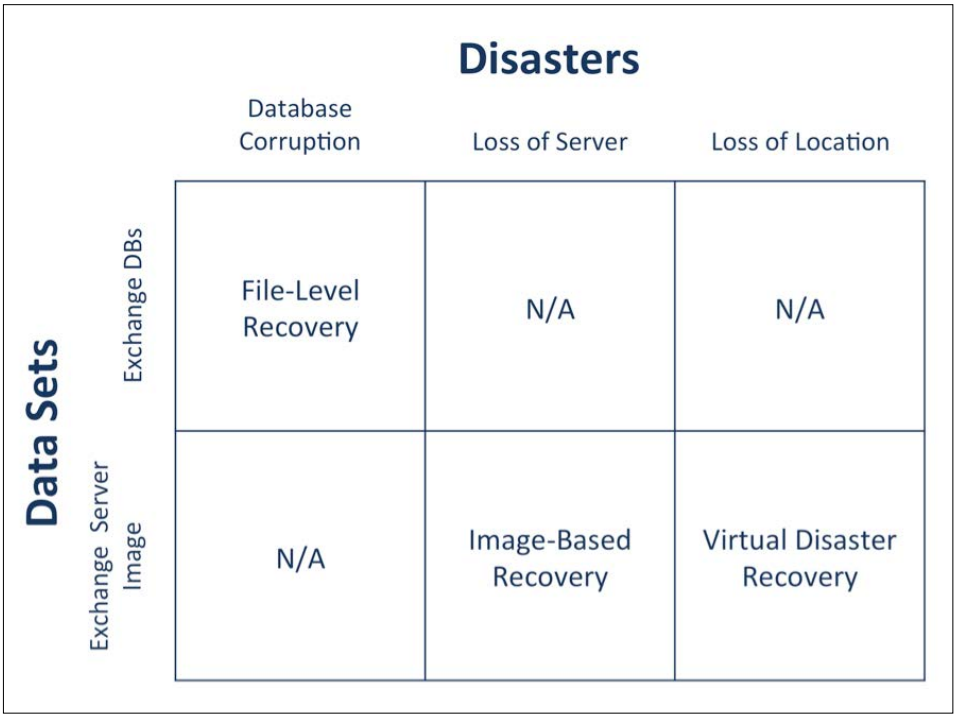

Figure 2: *Building a Disaster Recovery Grid*

The last part of the plan to keep you from drowning in data involves limiting what needs to be backed up.

When combined, using this grid, along with all the considerations listed within this paper, will help determine the reality of whether you can actually accomplish the intended recovery. For example, if you are planning for the loss of your Exchange server using a cloud source and a local target – pulling down an entire VM may not meet the initially defined RTO.

Don't Back Up … Archive!

The last part of the plan to keep you from drowning in data involves limiting what needs to be backed up (and, therefore, recovered) in the first place by utilizing archiving. By implementing a backup archive, you take a portion of your data set (for example, files on a file server that are over 3 years old) and eliminate it from the regular process of backup and recovery, increasing the likelihood you can meet recovery objectives by needing to recovering less data.

Many companies don't purposefully create an archive – it's more like the monthly backup becomes the archive, as opposed to architecting a plan for what needs to be archived and than creating a purpose-driven backup specifically to address the archive need. And given that most

backups these days are only retained for 30 days, having an archive becomes even more important.

So, how can you create an archive strategy that fits in with your backup and recovery strategy?

You need to first recognize you're backing up more than you need… from two possible perspectives: First, you may not be archiving the correct data sets and are, instead, backing up everything all the time – so your backups are too large. Secondly, you may be actually archiving the wrong things – making your archive too large. You need to find the balance.

The general rule of thumb of the Grandfather-Father-Son methodology still stands true.

Determining what to archive will depend on internal requirements, generally for compliance, legal / HR, or audit reasons. Additionally, backing up critical files that don't change often should be considered, as well as the need for a malware infection response plan should be part of the conversation.

The general rule of thumb of the Grandfather-Father-Son methodology still stands true, allowing a new archive to be generated weekly so that you meet the standard of 7 daily archives, 4 weekly archives, 12 monthly archives, and 1 yearly or multiple yearly archives.

By carefully deciding what should be a part of the archive (along with the plan to ensure you always have one), can safely eliminate some portions of data sets from your regular backups, leaving recovery to the incremental backups of your non-archived data.

Summary

If you focus on the sea of data, you'll simply become overwhelmed at the enormity of the problem. Like focusing on the simple act of treading water instead of trying to build a master plan of how to get back to shore 5 miles away, you need to address the problem one step at a time.

By focusing on individual data sets that need to be recovered, defining recovery objectives, offloading any data that should simply be archived, and testing the resulting data sets against the disasters you want to recover from, you'll end up with a reasonable plan of attack to recover from just about anything.

With nearly 20 years of enterprise IT experience, Nick Cavalancia is an accomplished consultant, speaker, trainer, writer, and columnist and has achieved certifications including MCSE, MCT, MCNE and MCNI. He has authored, co-authored and contributed to over a dozen books on Windows, Active Directory, Exchange and other Microsoft technologies. He has spoken at conferences such as the Microsoft Exchange Conference, TechEd, Exchange Connections, and on countless webinars and at tradeshows around the world.

UNITRENDS