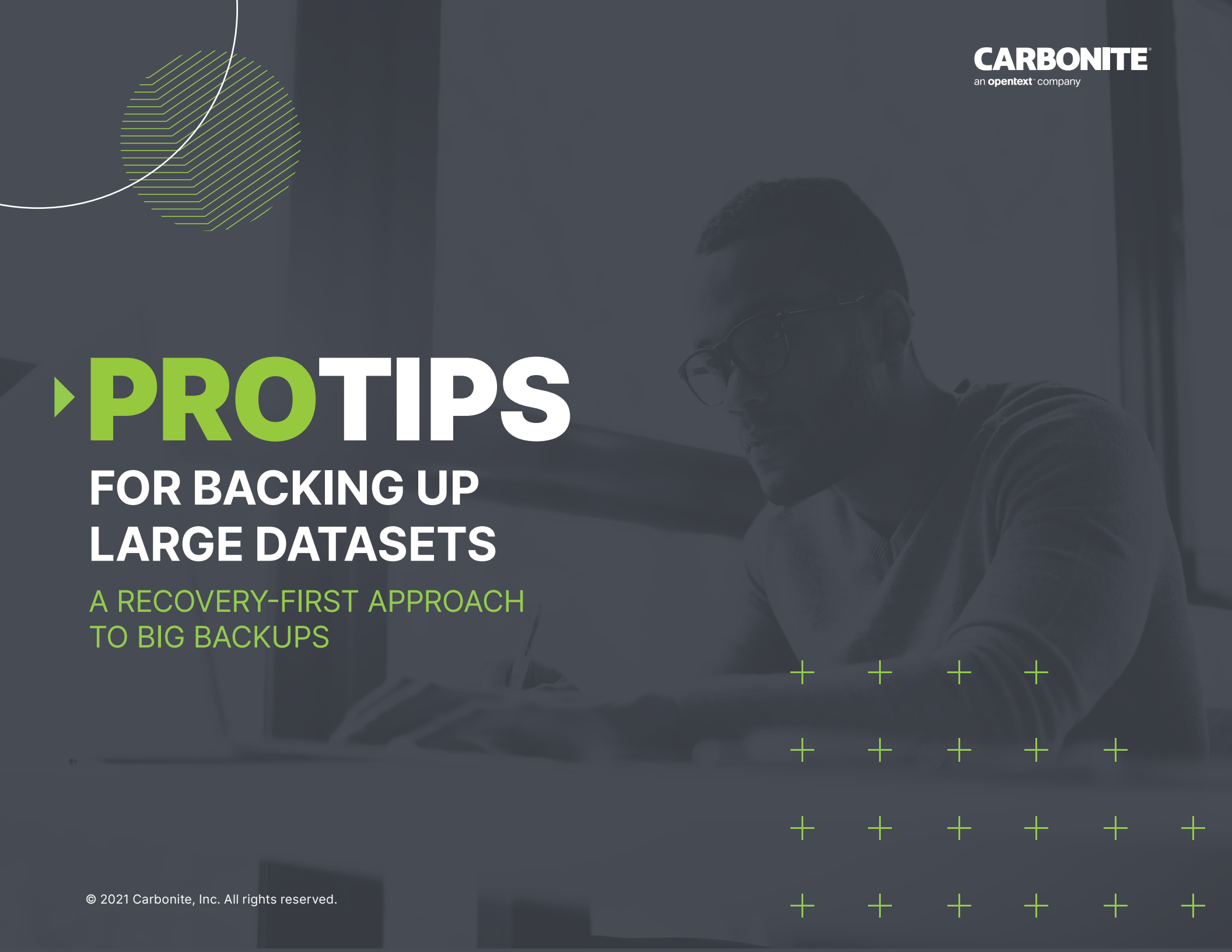
CARBONITE
an opentext company

PROTIPS

FOR BACKING UP LARGE DATASETS

A RECOVERY-FIRST APPROACH TO BIG BACKUPS



IT admins tasked with restoring servers or lost data during a disruption are consumed with a single-minded purpose: successful recovery. But it shouldn't take an adverse event to underscore the importance of recovery as part of an overall backup strategy. This is especially true with large datasets. Before you consider how you're going to back up large datasets, first consider how you may need to recover the data.

Variables abound. Is it critical or non-critical data? A simple file deletion or a system-wide outage? A physical server running onsite or a virtual one hosted offsite? These and a handful of other criteria will determine your backup and disaster recovery (BDR) deployment.

What do we mean by large? A simple question with a not-so-simple answer. If your total data footprint is 5 TB or more, that's considered large. But what kind of data is it? How many actual files are there? How frequently do they change? How much can they be compressed? It's likely that two different 5 TB environments would require different data protection schemes if they were comprised of different file types that changed at different rates.

On the other hand, bandwidth capacity restrictions are a common denominator for all environments. The question boils down to this: How should you back up data so that it can be reliably recovered through a process that doesn't interfere with daily workloads traveling across the network? IT pros on the frontlines have no single tool for determining the impact that backing up large datasets will have on bandwidth. It's a process of trial and error, even for the experts who do it daily. You can only protect as much data as your network will allow. And there's little use backing up data that can't be recovered in a timely fashion.

Before you consider how you're going to back up large datasets, first consider how you may need to recover the data.

Contributing factors

Factors that can determine backup procedures, in addition to size, include:

Type of data
Some data is more compressible than others.

Nature of data
Some data is more critical to protect than others.

Rate of change
How frequently data changes will affect the size of backups.

DATA SLEUTHING

A little preemptive digging can simplify the process of protecting large datasets. The first question: Why do you have so much data to protect? When you're dealing with large datasets, it makes sense to first question the need to protect so much data. If it's critical for the business, if it's required by a federal or industry regulation, or if it aligns with the company's document retention policy, it needs to be protected. Any data that doesn't serve a compelling business purpose and isn't covered under a regulatory framework should be excluded from your normal backup schedule. It pays to be judicious at this step.
Any bandwidth you conserve now will give you added flexibility later when you need to determine backup frequency for achieving predetermined RPO outcomes.

Once you decide on the data you need to protect, you should then decide whether the data is so critical to the business that it needs to be accessible at any hour of the day. This mission-critical data should be protected onsite with a dedicated backup appliance, which allows the fastest possible recovery time and helps you achieve predetermined RTO outcomes. A backup appliance can also replicate data offsite for recovery in the event of a site failure or regional outage.

Any bandwidth you conserve now will give you added flexibility later.

Mission-critical data is anything that can affect operations or the financial health of the business. Identifying mission-critical data is a crucial first step in the process because it allows you to prioritize backup tasks based on desired recovery outcomes. Each business defines “mission critical” differently. For health care organizations, mission critical data could be electronic medical records (EMR) and protected health information (PHI). For a financial institution, it could be balance and transaction information. For a casino, it could be surveillance video. It’s important to establish and agree internally on what’s mission critical before a disruption occurs because these are the applications and data that will require the most stringent RTO and RPO outcomes.

IT pros play detective

Five essential questions for backing up large datasets:

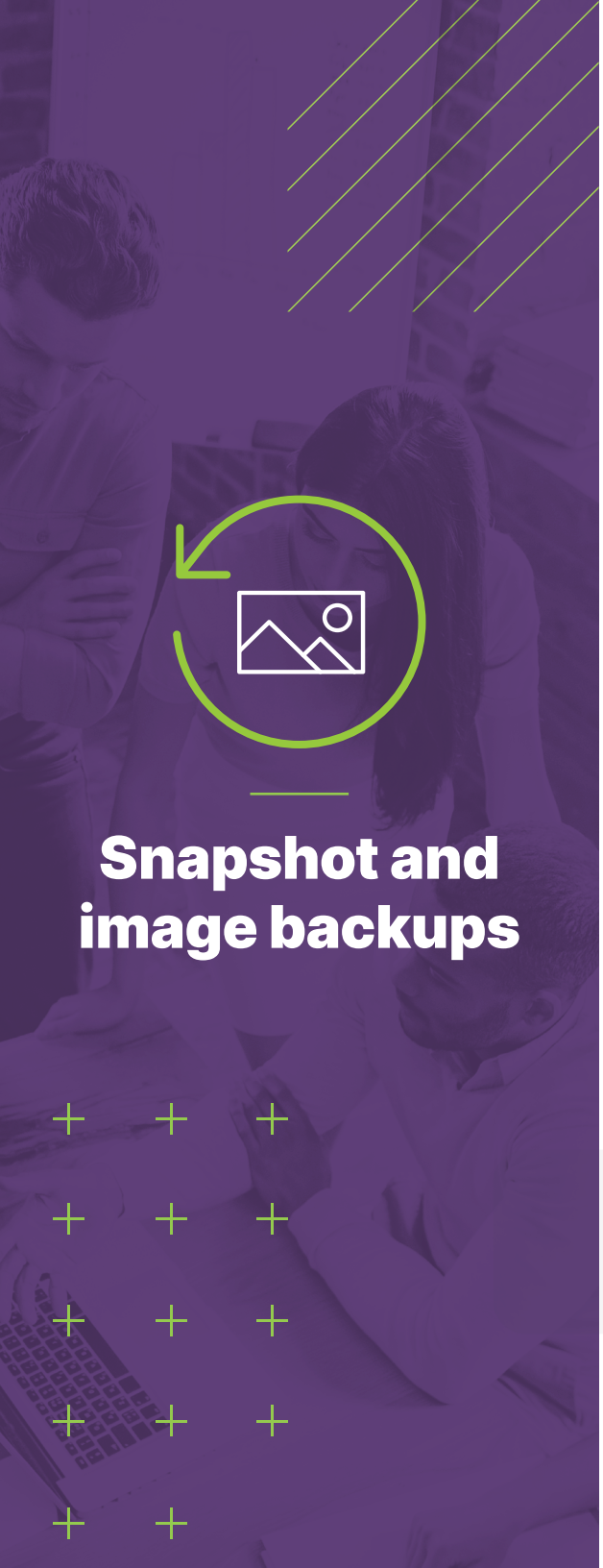
1. What is my company’s document retention policy?
2. Which data is mission critical to the business?
3. What type of data do I have and is it compressible?
4. What’s the typical rate of change for the data?
5. What size backup will my network support?

Changing data

When you're backing up terabytes of data, or even hundreds of terabytes, the size of the data footprint may be less significant than the rate of change for the data. After the initial backup and dedupe process is complete, only changes to the initial backup need to be recorded. In the case of Carbonite backup solutions powered by EVault technology, we use proprietary technology called DeltaPro, where changed data blocks are stacked on top of the initial full backup. If you're backing up once a day and you configure a retention of seven days, then your first backup is the full one, and each subsequent backup is essentially a delta or change. However, Carbonite will make it look like each backup is a full one. Once the eighth daily backup completes, "backup one" expires, and you still have seven copies to recover from. When it's time to recover, Carbonite does the work on the back end to restore a complete backup set. This eliminates the extra step of having to recover the initial backup first and then the differential backups to assemble a complete backup for DR purposes.

We recommend backing up once a day during off-peak hours to allow time for maintenance on the back end. Carbonite software prioritizes backup tasks over maintenance. So, if you send a backup set to the vault before the software has time to complete maintenance, you could wind up with a larger backup size on the back end than if you allow the dedupe process to complete. We have some customers with extremely stringent RPO requirements performing multiple backups a day (intraday backups) to give themselves more frequent recovery points. To keep these backup tasks running smoothly, we recommend breaking down large backup sets (600 GB) into two smaller backup sets (300 GB each). If your business requires multiple intraday backups for large datasets, it might also help to take a deeper look at the type of data, object count and bandwidth. Different businesses use different strategies. Our professional services and support team can offer tailored recommendations and best practices for your organization.

When you're backing up terabytes of data, or even hundreds of terabytes, the size of the data footprint may be less significant than the rate of change for the data.

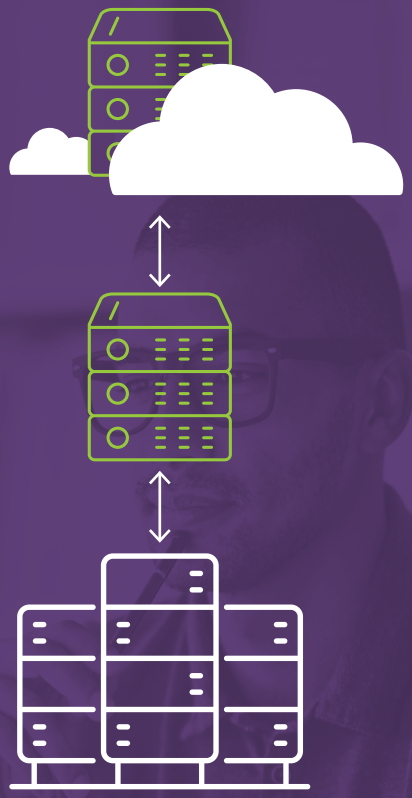


Snapshot and image backups

Data is more than 1s and 0s. Some datasets have more redundancy than others, making them easier to compress. Media files like images, audio and video tend to have less redundancy than application data. Some companies have more of one data type than others. Insurance companies, for example, have a lot of incompressible images. They could easily reach a 5 TB data footprint, but because of the type of data they tend to generate, they would have different BDR requirements than a business with only email and file servers. Insurance companies use an application called SourceOne for email archiving and another called ImageWrite for organizing images. These two applications alone can create millions of files, which could lengthen backup windows and slow recovery efforts due to the vast amounts of data objects. This affects any backup solution.

Businesses with large amounts of incompressible data, with stringent RTO and RPO requirements, typically require snapshot or image backup. A snapshot or image backup allows you to move large datasets over the network more efficiently without interfering with critical workflow. Image backup is appropriate for environments with a data footprint of 1 TB or more or with millions of files or data objects. Carbonite uses image backup to capture a Windows server's operating system and the local data volumes. When it's time to recover, whether you're restoring individual files or a complete system to dissimilar hardware, image backup gives you the flexibility to choose the method that best suits the situation you're trying to mitigate.

A snapshot or image backup allows you to move large datasets over the network more efficiently without interfering with critical workflow.



Cloud backup

With external connections exponentially slower than internal ones, it's even more critical to gain efficiencies when backing up large datasets to the cloud. For large datasets, we recommend disk-to-disk-to-cloud backup (D2D2C), also known as hybrid backup, where mission-critical data is protected onsite for rapid recovery, and the entire footprint is replicated in the cloud. This ensures the speed of onsite recovery with the added protection of cloud backup for disasters. It's also wise to replicate outside of your primary FEMA zone to mitigate against multiple outages in the same geographic region.

For large datasets, we recommend disk-to-disk-to-cloud backup (D2D2C), also known as hybrid backup.

Another option that makes sense for large environments is disaster recovery as a service (DRaaS), where you outsource your entire BDR deployment to a dedicated team of experts. Carbonite Cloud Backup Powered by EVault is a DRaaS solution that offers businesses 1-, 24- and 48-hour SLAs. If you ever experience an onsite interruption, our team will fail over your entire environment to the cloud until your primary site is operational again.

Seeding and courier recovery

Backing up large datasets directly to the cloud has disadvantages—specifically, restricted upload speeds and dropped connections, which could restart the backup job. Depending on the size of the backup and data type, your upload speed may be too slow to configure the initial full backup in the cloud. In this scenario, we recommend seeding the initial backup by replicating to an external drive and shipping the drive to accelerate the process. Subsequent backups only need to capture changes to the initial backup. With a typical rate of change between 3% to 5%, backup jobs shouldn't interfere with normal workflow. You can also throttle bandwidth usage for cloud backups to alleviate network congestion. When it's time to recover, depending on the scale of the data loss, it may be faster to use a courier recovery service, where your backup is shipped to you on an external drive, rather than trying to recover large datasets from the cloud on an overly restrictive connection.

Money Matters

Protecting large datasets has implications beyond the simple mechanics of backing up—it requires purpose-built technology solutions for handling the data.

IT assets represent a significant investment for businesses. As organizations adopt new technologies, like the cloud, virtualization and custom applications, environments are becoming more complex. This makes platform compatibility increasingly important for today's IT ecosystems. With so much invested, there's a need to continually support both legacy systems and new platforms with bolt-on technology (as opposed to rip-and-replace).

Some vendors are slow to adapt to the rapid rate of data growth that businesses are experiencing. The software is up to date but the pricing model is obsolete. You see this with excessive overage charges and "per instance" fees. As businesses scale up their environments, they take a hit with every server, database or application they add to their backup schedule. This is a problem for large environments due to the volume of data and the degree to which it's spread across the organization. As environments grow up and out, the cost of protecting them escalates. Carbonite addresses this by offering unlimited licensing—you can protect as many servers, databases and applications as you need to, no matter where they're located—and what we call "protected footprint" pricing, which is based on the amount of data you need to protect, not how much space you need to protect it. BDR becomes self-defeating once the cost is so prohibitive that businesses skimp on protection.

Large dataset BDR summary

Protecting large datasets has implications beyond the simple mechanics of backing up. It requires purpose-built technology solutions for handling the data. It also requires a carefully considered backup strategy that includes determining the nature of the data (critical vs. non-critical), the type of data and the rate of change. These factors will determine how you deploy a BDR solution to achieve the recovery objectives you've established for your organization.

Get in touch

Phone: 877-542-8637
Email: carb-data_protection_sales@opentext.com