THE BACKUP BIBLE

The Complete Guide to Protecting your Data

PART 1:
CREATING A BACKUP & DISASTER RECOVERY STRATEGY

by Eric Siron

ALTARO

CONTENTS

INTRODUCTION

Humans tend to think optimistically. We plan for the best outcomes because we strive to make them happen. As a result, many organizations implicitly design their computing and data storage systems around the idea that they will operate as expected. They employ front-line fault-tolerance technologies such as RAID and multiple network adapters that will carry the systems through common, simple failures. However, few design plans include comprehensive coverage of catastrophic failures.

Without a carefully crafted approach to backup, and a strategic plan to work through and recover from disasters, an organization runs substantial risks. They could experience data destruction or losses that cost them excessive amounts of time and money. Business principals and managers might even find themselves facing personal liability consequences for failing to take proper preparatory steps. At the worst, an emergency could permanently end the enterprise.

This book seeks to guide you through all stages of preparing for, responding to, and recovering from a substantial data loss event. In this first part, you will learn how to assess your situation and plan out a strategy that uniquely fits your needs.

WHO SHOULD READ THIS BOOK

This book was written for anyone with an interest in protecting organizational data, from system administrators to business owners. It explains the terms and technologies that it covers in simple, approachable language. As much as possible, it focuses on the business needs first. However, a reader with little experience in server and storage technologies may struggle with applying the content. To put it into action, use this material in conjunction with trained technical staff.

COMMONS TERMS USED IN THIS BOOK

Where appropriate, the book will define terms close to their first usage. However, some terms will feature right from the beginning and recur regularly throughout the entire work. Their definitions appear here so that you have a convenient reference point.

1. **Backup:** A “cold” instance of duplicated data.
2. **Business continuity:** The ability of an organization to continue performing its desired activities and functions throughout an emergency situation.

3. **Disaster recovery:** The process of returning to full functionality after an emergency.

4. **Replica:** A “warm” or “hot” instance of duplicated data.

The book contains a full discussion on the “cold”, “warm”, and “hot” distinctions between backup and replica. For now, understand that they have different meanings and that the technologies have different uses.

ALTARO BACKUP SOLUTIONS

Altaro is an award-winning developer of a range of backup solutions including Hyper-V & VMware Backup and Replication, Office 365 Backup, Physical Server Backup and Endpoint Backup (coming soon). Altaro also offers an attractive Partner Program for VARs, resellers and IT consultants as well as dedicated programs for Managed Service Providers (MSPs). With 50,000+ customers in 121+ countries, 10,000 partners and 2,000+ MSPs, Altaro provides affordable enterprise-class functionality coupled with outstanding 24/7 support.

This eBook has been designed to cover the theory and practical exercise of backup and disaster recovery planning. It contains references to Altaro Backup solutions where relevant, but it is not specifically about Altaro or a guide to using Altaro products. As such the information contained here is relevant to whichever backup software you choose to deploy.

GETTING STARTED WITH DISASTER RECOVERY PLANNING

A solid disaster recovery (DR) plan needs time and attention to properly form. Usually, the total investment closely coincides with the size and scope of the organization. Due to the level of effort, many businesses need help improving their process beyond regular backups. Some also struggle with finding a logical starting point.

To get started, you need to build a checklist. It should include clear goals and the activities that will achieve them. You will need to create a custom list that fits your particular needs. You will likely need to refine the list items as you work through it.

You can use the following example checklist as a starting point.
You may not recognize all of the terms used in this list just yet, but you will find definitions later in the book. We will explore each of these topics in further detail but for now here are the essential items you will need to include in your backup and disaster recovery checklist.

DISASTER RECOVERY CHECKLIST

- ☑ Make the business case for a disaster recovery plan
- ☑ Identify risks
- ☑ Determine key stakeholders
- ☑ Define a data prioritization strategy
- ☑ Discover data protection scope
- ☑ Define recovery objectives and tolerances (RTOs and RPOs)
- ☑ Determine solutions
- ☑ Define capital and operating budgets
- ☑ Create implementation plan
- ☑ Create business continuity plan
- ☑ Create disaster recovery plan
- ☑ Create test plans
- ☑ Create review plan
- ☑ Follow the implementation plan
- ☑ Schedule and follow the review plan

Your list will grow beyond this one, usually with several sub-items specific to your particular requirements. Out of all these items, the last, “Schedule and follow the review plan”, may be the most important. Disaster recovery planning is an ongoing process, not a one-time event. You will do the bulk of the work during the initial planning phase, but your organization cannot simply abandon the plan after implementation.

Your first item, making the business case, usually spans a few of the other items. You can easily gain acknowledgement of the importance of backup, but organizational commitment is essential to a realizing a thorough plan and then carrying it out.

IDENTIFYING DATA RISKS AND PRIORITIES

The lack of understanding around data protection presents a serious barrier to proper planning. Some organizations fail to adequately plan simply because they do not realize its importance. Others do not feel that the danger justifies the effort. The lack of a plan presents the greatest danger of all. This chapter helps you to paint a fuller picture of the risks that a disaster recovery strategy can mitigate.

NEGATIVE ATTITUDES TOWARD DISASTER RECOVERY PLANNING

No one has conducted in-depth studies into the behaviors and attitudes around disaster planning. We do not know what percentage of organizations minimize or even skip this critical component. Most importantly, we do not conclusively know why system designers tend to reduce the importance of disaster recovery. We do have common anecdotes from individuals that have worked with companies to plan for or recover from disasters. Some reasons frequently cited:

REASONS WHY DISASTER RECOVERY FAILS

SUCCESS BREEDS A SUCCESS MENTALITY

EXPENSE

TIME

SCOPE

MISUNDER-STANDING

SHORT-TERM THINKING

- **Success breeds a success mentality** - The longer an organization survives without experiencing a catastrophe, the less its members believe in the possibility of it occurring. As is true for humanity in general, few people tend to strongly consider emergencies until it's too late.
- **Expense** - Generally, businesses perform infrastructure computing and storage operations in bulk. They purchase and deploy several components all at once. In the case of clusters and replicated storage devices, they may have no other options. Usually, planners design the functional portions first, then add in the protection schemes. As the capital expenditure sum climbs, the willpower to spend tends to decline.
- **Time** - Building and implementing a proper disaster recovery strategy requires time. There's no way around that. Much of it requires the involvement of principals and senior staff. They may feel that they have better ways to allot their times than sitting in meetings and filling out questionnaires to prepare for an event that might never occur. They may also feel that their technology teams should focus on other endeavors.
- **Scope** - Sometimes, a backup plan does exist, but falls short of organizational needs. Taking a nightly backup certainly grants better protection than doing nothing at all, but that cannot represent the entire strategy.
- **Misunderstanding** - Even in today's world of ubiquitous technology, few people understand the differences between the datacenter and the desktop. Consumers rarely back up their personal computers or devices. They simply do not comprehend the risks. Without an experienced guide, they tend to underestimate the hazards.
- **Short-term thinking** - In most cases, improper planning results from innocent ignorance and naivete. Unfortunately, not everyone will have the organization's best interests in mind. A trusted consultant might try to win a contract by providing a cheap solution with little or no backup functionality. A less-than-scrupulous business manager might decide to check that important "under-budget" box by skimping on backup. Or, a well-meaning principal might adopt a "let's deal with that later when we have a little more money" stance – but later never comes.

As you work on your disaster recovery plan, keep all of these things in mind. Because backup and disaster recovery have no immediate benefit, you will almost certainly face resistance. You need to remain prepared to answer “why” at any time. The next section can help.

ASSESSING THE RISKS THAT NECESSITATE A DISASTER RECOVERY STRATEGY

If you study computer security, you will have heard of “threat modeling”. Essentially, it means that security experts first identify potential threats. They can use that list to predict the extent of possible damage from an attack. That helps them to design a clear strategy for defense and mitigation. You can use a similar approach to building backup and disaster recovery systems.

In the case of disaster recovery, the risks consist of a superset of the security threat model. Malicious actors pose one kind of threat out of many. You also have to worry about hardware failures, natural disasters, and human error.

With each risk, you must consider its possible impact. Attackers might steal your data. A failed storage system could cause complete data loss. A flood might make your premises entirely unusable. Someone could delete a critical e-mail that places your organizations in a legally vulnerable position. Each danger type presents a unique challenge for every organization.

At this point, you may only be able to draft a cursory idea of your risks. A proper assessment includes a detailed analysis. However, in all but the smallest companies, these investigations need more than one person. At this stage, you only need enough to make a solid case for spending time and capital on designing and creating a comprehensive backup and disaster recovery solution.

A LIST OF COMMON RISKS

To help you start your list, consider some of the major risks that all organizations face:

- Data theft
- Physical theft
- Malicious digital attacks (ransomware, viruses, etc.)
- Rogue insiders
- Social instability
- Power failures
- Arson
- Sabotage

- Natural disaster
- Departure of critical staff

Take some time to research risks particular to your industry. You may not add anything to the list, but you might need to adjust its priorities. For instance, if your organization creates software, then “intellectual property theft” will feature prominently. If you transport commodities, then physical threats will rank higher.

This might be the point at which you create and present the business case for undertaking disaster recovery planning. If you need more material, then perform preliminary work on some or all of the next three items in the checklist.

DETERMINING KEY STAKEHOLDERS

Depending on your organization’s size and your position within it, you may not have the authority or knowledge to conduct a deeper investigation on your own. Whatever your role, start with what you consider important. If you’re a systems administrator, you may think of your e-mails or files. If you have a more general operational position, you might think of your book of business or inventory. To define a fuller plan, you need a wider view.

To gain that perspective, you will likely need the approval of your organization’s executives. Properly analyzing risk requires time and attention. In the absence of an obvious threat or recent catastrophe, you will likely struggle to move this phase of the plan along.
Even people that understand the risks tend to consider it a low-priority task. Set a goal of getting the appropriate people involved in the conversation and ensuring that they have sufficient motivation and opportunity to participate.

To start the conversation, use an informal approach. Start asking things like, “Which people would know the most about our risk profile?” and, “Who has the best knowledge of what we need to protect?”. Expect to need input from:

- Executives or principals
- Head and leads of IT
- Key stakeholders – these vary greatly between organizations. It might mean department heads or product owners or individuals in major roles.
- Intellectual property creators and proprietors

With a starting list of names, you have options: individual interviews, forms, or group meetings. You may eventually use all of these things, but you will likely find that brainstorming meetings will get you the farthest in the beginning. However, the risk discovery task neatly connects with several of the following activities. Therefore, you will likely want to read ahead before scheduling anything.

DISCOVERING YOUR UNIQUE RISKS

You will not need to spend much time on this particular part of the process. Business continuity and disaster recovery both mean working through and after major problems, regardless of how they occur. Smaller events need different responses. For instance, you might need to restore a single database after an accidental deletion. So, you need to know how an accidental (or malicious) deletion might happen.

As you and your colleagues work through this discovery phase, you might find mitigation strategies that allow you to reduce exposure to your unique risks. Where possible, choose prevention over response. You will probably not completely remove many items from your list of concerns but take every advantage that you can.

Be mindful of course-altering events. For instance, if your organization centers around physical product in a warehouse, and a disaster completely destroys the facility and all of its contents, then you probably won't concern yourself as much with a pick-up scheduling application.

DATA PRIORITIZATION

Meetings and discussions about risk will inevitably cover the vital portions of your organization's systems. As you outline your exposure, you can take the opportunity to rank your assets. Most disaster recovery plans will encompass everything, but even in the best cases, restoration takes time. For now, do concern yourself with the rebuild order. Focus on mission-critical applications – what does the organization need for minimal operation?

At this phase, organize your priority list at its highest level. For example, instead of making line items that make sense to administrators, such as "customer database", use business-oriented labels such as "ERP system". You can work out the technical details later. Things will necessarily look different once you translate this list into an implementation document.

As you build up this list, ensure that everyone involved remembers that top priority belongs to the systems that your organization requires for operational performance. Try to avoid using terms like "critical", as not everyone will agree on the definition, and sometimes, you can function for a while without a crucial system. As an example, consider a company that transports freight. No one can dispute the importance of keeping the electronic customer record system available, but can the operation continue without that longer than it can continue without the system that maintains contact with delivery and pickup drivers? The question to ask of every system: "What is the business impact of an outage?" For now, you may need to keep those answers short.

WIDEN THE SEARCH FOR ESSENTIAL DATA

Meetings alone will not uncover everything that you need to protect. They serve as a starting point for the attendees. They will need to look within their departments. To complete the data protection model, key staff in each department must create a thorough inventory.

The search should not restrict itself to digital assets. Your organization may predate the advent of digital record keeping, or it may fall under the purview of regulations that require physical copies. Business continuity and disaster recovery will mean protecting those items as well.

WHAT'S NEXT

As your risk and priority models take shape, you will naturally build up an idea for the tolerances and expectations that you have in your disaster and data recovery planning. You might be able to clearly define all of those in the same meetings. However, they often require a more detailed examination of the supporting systems. Department managers may need to break to gather input from daily operators. To help you through this portion, the next chapter defines the terms and processes related to recovery objectives and tolerances.

RECOVERY OBJECTIVES AND LOSS TOLERANCES

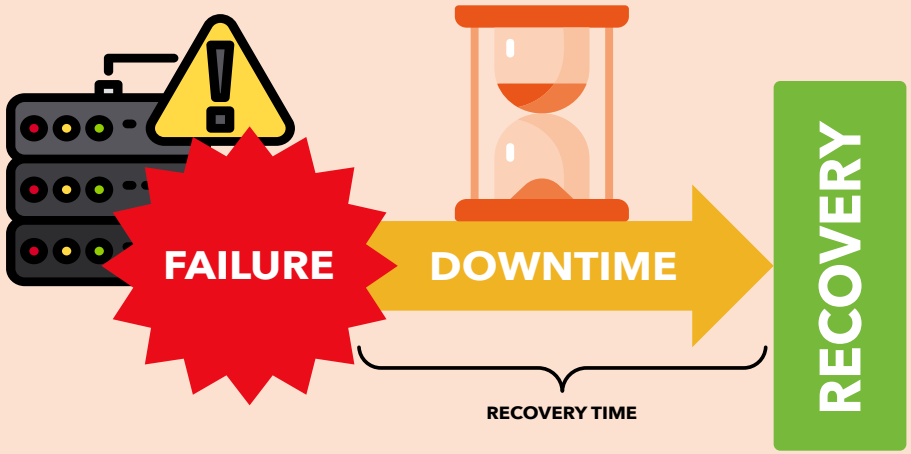
With sufficient funding and infrastructure, any system could theoretically achieve near-constant uptime through any situation. Reality dictates a more conservative outlook. To establish a workable budget and a practical plan, you will need to determine your organizational tolerances for outages and loss. This chapter explores the related terminology and processes.

In the previous chapter, you were instructed to ask about the business impact of a system outage. Now, you will need to have the key employees of each system explore the question more deeply.
If the term “business impact” does not convey the desired level of urgency, ask questions such as “How much does this system cost us per hour when offline?” and “How many hours of work would we need to recover after losing one hour’s worth of data?” You need to build up a set of objectives: recovery time and recovery points.

ESTABLISHING RECOVERY TIME OBJECTIVES

The simple question, “How long can we operate without this system?” can get your teams started. The term “recovery time objective” (RTO) applies to the goals set by this enquiry. RTOs establish the desired maximum amount of time before a system returns to a defined usable state.

RECOVERY TIME OBJECTIVE (RTO)



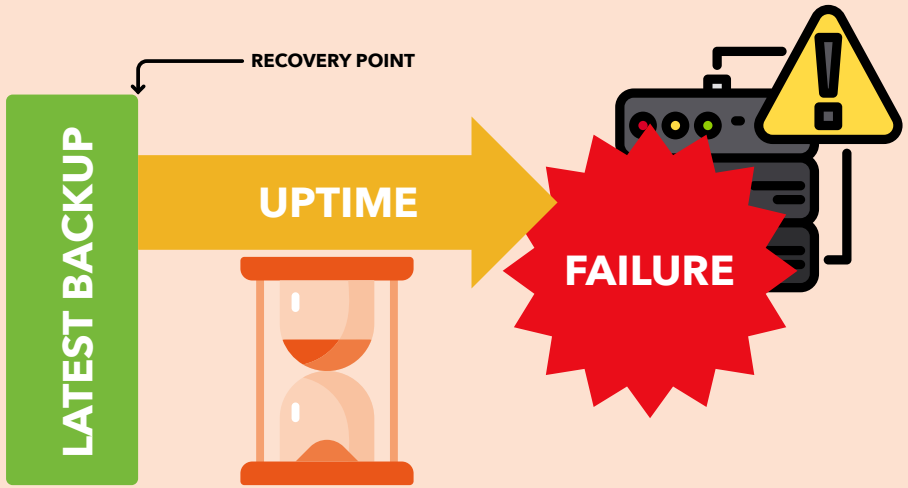
Complex systems can have different RTOs. For instance, you might set an RTO of four hours to restore a core electronic records system after a major failure but set a separate objective of one hour after a minor glitch. Your objectives might also set differing levels for acceptable functionality. Your organization might consider a functioning receipt printer in the customer service area as a meaningful success metric; it can have its own RTO as a part of a larger recovery objective.

RTOs should feature prominently on your long-term disaster recovery planning. Rely on the managers and operators of the individual systems to provide guidance. Use executives to resolve conflicting priorities. You may also need them to grant you the ability to override decisions in order to ensure proper restoration of functionality.

ESTABLISHING RECOVERY POINT OBJECTIVES

RTOs apply mainly to functionality. Events that trigger recovery actions also tend to cause data loss. Your organization will need to establish tolerance. Of course, no one wants to lose anything, which will make these discussions difficult.

RECOVERY POINT OBJECTIVE (RPO)



Because most backups occur at specific time intervals, you use them as the basis for "recovery point objectives" (RPO). An RPO sets the maximum acceptable time duration between the latest backup and the data loss event. This determination coincides with the work to determine the business impact of an outage. A system's downtime not only prevents its users from retrieving or utilizing its contents, but also represents matching post-recovery work: they will need to recreate any data that was not in the backup and they will need to complete postponed operations.

You will need to establish multiple RPOs for most systems.
Not all events will have the same impact, so you must set expectations accordingly. For instance, you have options for continuously created replicas and backups. Those work well as buffers against physical hardware failure. They work poorly against malicious attacks, especially encrypting ransomware. You can establish a tiered recovery approach to address the various risks. As an example:

1. First-line hardware failure or malfunction: RPO of zero hours, using continuous replication
2. Corrupted data: RPO of 1 hour after corruption detection, using on-site hourly backups
3. Site destruction: RPO of 24 hours, using off-site daily backups and cloud hosting providers

Consider the possible outcomes of each risk category as you work out RPOs. You don't just the data to restore; you need something to restore it on. If you need to acquire replacement hardware or bring in third parties for assistance, that might add time. If you have a secondary site available, add an RPO item for recovery to that location. Also include an item that addresses cross-facility challenges. Do not forget to account for availability of critical staff.

You can read more about RTO and RPO on the Altaro Dojo.

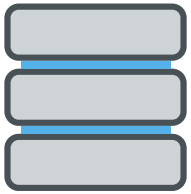
DEFINING RETENTION POLICIES

Your teams have a final major decision to make: how long to keep data. These decisions are highly dependent on the nature of the business and the data. For European-based operations you may also have to consider GDPR requirements. If you don't have immediate answers, use two major guidelines:

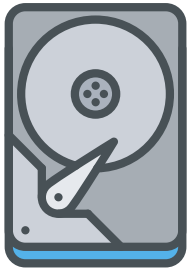
1. Legal requirements. As an example, you may need to keep records of taxable events for a number of years.
2. How long will the data have value?

Use the answers from these questions to create “retention policies”. A retention policy dictates how long data must be retrievable. You will likely need more than a single company-wide policy. “Forever” may seem like an obvious answer for some things but ensure that everyone understands that data storage has an associated cost.

Data retention has two tiers: live storage and backup storage. In disaster recovery planning, IT often only considers the backup tier. However, remember that a backup captures existing data, regardless of its age. So, if a live database has records that go back a decade, then the most recent backup contains ten-year-old information. Therefore, both your current live data and yesterday’s backup satisfy a ten-year retention policy.



RETAIN IN ONLINE STORAGE



RETAIN IN OFFLINE STORAGE

To accommodate both the live and the backup tiers, retention policies must consider two things:

- Purge policies for active data
- Probability of unnoticed undesired deletion

Some electronic records systems prevent true deletion from databases without a purge action. It might move “deleted” records into a historical table or it may have a flag that removes them from visibility in client applications. Such safeguards reduce the probability of accidents. They can help against malicious deletion as well. Remember that individuals with administrative access can usually override application-level security. For the greatest safety, assume that you will not achieve your retention policies for live data. You can relax that expectation for non-critical data. Factor in the results of impact analysis from the earlier exercises. Ask, “If we lost this data forever, how would it impact the organization?”

ADJUSTING RTOS, RPOS, AND RETENTION POLICIES TO MATCH PRACTICAL RESTRAINTS

Shorter RTOs and RPOs almost always require greater financial and technical resources. Short backup intervals consume more media space and network bandwidth. Lengthy retention policies increase storage and administrative costs. Layered approaches to cover the various risk profiles can multiply those needs.

2 hour RPOs = 500 GB per day

4 hour RPOs = 250 GB per day

Backup operations place a load on the production system, which might add more strain than your current equipment allows. Replication and continuous backup technologies need more technical expertise than typical nightly backups. Staff must periodically test the validity of backup data, adding effort and overhead.

Make all of these constraints clear during early planning meetings. As executives and department heads express their wishes for speedy and RTOs and short RPOs, ensure that they understand that costs will rise accordingly. They may need to adjust their expectations to match.

Your plans will also need to factor in time and expense to re-establish infrastructure after a failure. You may need to replace physical systems. Vital foundational infrastructure, such as domain controllers, take automatic precedence over anything that depends on them. Adjust RTOs and RPOs for other systems accordingly.

The backup software that you choose will play a role in your RTO and RPO restrictions. Altaro VM Backup provides highly customizable backup scheduling options as well as Continuous Data Protection (CDP). You need fine-grained flexibility such as this to balance your backup needs against your available resources.

WHAT'S NEXT

The major activities of this chapter include input from all sectors of the business. Through interviews, questionnaires, and meetings, you can assemble an organizational view of what you need to protect. Next, you need to determine how you will implement that protection. You have not completely finished working with the non-technical departments, but you have a different phase of the project to work on now. In the next chapter, you will explore the ways that you can use technology to achieve the desired disaster recovery strategy.

TRANSLATING YOUR BUSINESS PLAN INTO A TECHNICALLY ORIENTED OUTLOOK

Now that the business-oriented personnel have given their input on the design, you need to determine how IT can deliver it. In order to accomplish that, you need to discover the capabilities of the technologies available to you. Once you know that, you can predict the costs. You can take that analysis back to the business groups to build a final plan that balances what your organization wants for disaster recovery against its willingness to pay for it.

Mapping out your backup requirements will then help you plan software subscriptions fulfill your needs. Altaro recognizes the need for multiple backup solutions and as such currently provides backup for Hyper-V & VMware virtual machines, Office 365 (including Sharepoint and OneDrive for Business), and physical servers with further solutions on the way.

DISCOVERING THE TECHNOLOGICAL CAPABILITIES OF DATA PROTECTION SYSTEMS

At this point, you have an abstract list of high-level business items. Few backup solutions target line-of-business applications. So, you need to break that list down into items that backup and replication programs understand. To attract the widest range of customers, their manufacturers specify services and products that most organizations use.

Common protections include:

- Windows Server and Windows desktop
- UNIX/Linux systems
- Database servers
- Mail servers
- Virtual machines
- Cloud-based resources
- Physical hardware configurations

You'll need to create a map from the prioritized business-level items to their underlying technologies. Bring in technical experts to ensure that you don't miss anything. Gather input on what needs to happen in order to recover the various systems in use at your organization. Many require more effort than a simple restore-from-backup procedure.

Some examples:

- Active Directory
- Log-based SQL recovery
- Mail servers
- Multi-tier systems
- Cluster nodes

Take input from line-of-business application experts as well as server and infrastructure experts. Seek out the experience of those that have faced a recovery situation with the systems that you rely on most. You might find exceptions or special procedures that would surprise generalists.

FIRST LINE OF DEFENSE: FAULT-TOLERANT SYSTEMS

Ideally, you would never need to enact a recovery plan. While you can never truly eliminate that possibility, you can reduce its likelihood with fault-tolerant systems. “Fault-tolerance” refers to the ability to continue functioning with a failed component. A truly fault-tolerant system should allow an operator to replace the deficient part and return to full operational status without any service interruption.

Most fault-tolerant systems mostly function at low level, usually on the internal components of computer systems. To provide protection, they usually employ some method of hardware-level data duplication. In the event of a failure, they use the redundant copy to continue providing expected functionality. However, until someone replaces the defective part, the system does not provide redundancy. Further failures will result in an outage and possibly data loss.

COMMON FAULT TOLERANT SYSTEMS

Storage technologies make up the bulk of fault tolerant systems. Not coincidentally, they also have the highest failure rate. You can protect short-term storage (main system memory) and long-term storage (spinning and solid-state disks).

> “ A TRULY FAULT-TOLERANT SYSTEM SHOULD ALLOW AN OPERATOR TO REPLACE THE DEFICIENT PART AND RETURN TO FULL OPERATIONAL STATUS WITHOUT ANY SERVICE INTERRUPTION. “

SYSTEM MEMORY FAULT TOLERANCE

To provide full fault tolerance, memory controllers allow you to pair memory modules. Every write to one module makes an identical copy to the other. If one fails, then the other continues to function by itself. If the computer also supports memory hot-swapping and technicians have a way to access the inside without unplugging anything, then a replacement can be installed without halting the system.

Of course, system memory continues to be one of the more expensive components, and each system has a limited number of slots. So, to use fault-tolerant memory, you must cut your overall density in half. Doubling the number of hosts presents more of a cost than most organizations want to undertake.

Fortunately, memory modules have a low rate of total failure. It is much more likely that one will experience transient problems, which can be addressed with cheaper solutions. Server-class computer systems usually support error-correcting code (ECC) memory modules. ECC modules incorporate technologies that allow for detection and correction of memory errors. Some vendors provide proprietary technologies to defend against problems.

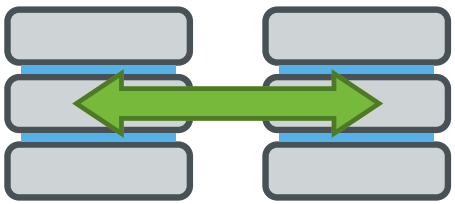
In most cases, you will choose ECC memory over fully fault-tolerant schemes. ECC cannot defend against module failure, but such faults occur rarely enough to make the risk worthwhile. ECC costs more than non-ECC memory, but it still has a substantially lower price tag than doubling your host purchase.

HARD DRIVE FAULT TOLERANCE

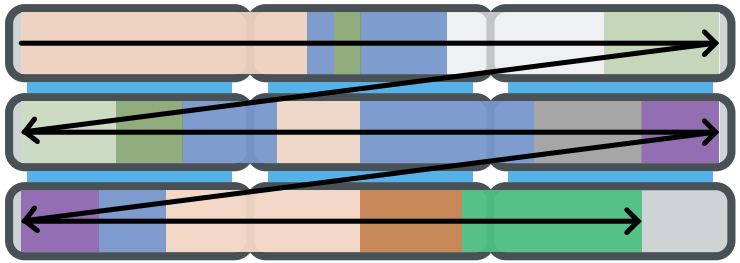
Hard drives, especially the traditional spinning variety, have a high failure rate. Since they hold virtually all of an organization's live data, they require the most protection. Due to the pervasiveness of the problem, the industry has produced an enormous number of fault-tolerant solutions for hard drives.

RAID (redundant array of independent disk) systems make up the bulk of hard drive fault tolerance designs. These industry-standard designs use a combination of the following technologies to protect data:

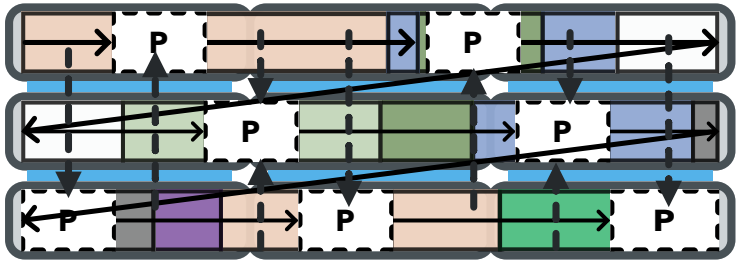
- **Mirroring** - Every bit written to one disk is written to the same location on at least one other disk. If a disk fails, the array uses the mirror(s).



- **Striping** - A block size is set for the array. When data written to a disk fills a block, writes continue on the next disk in the array. After filling a block on the last disk, writes wrap around to a new block in the first disk. Striping alone does not provide any data redundancy.



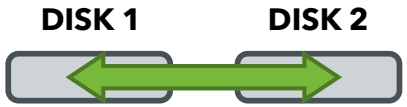
- **Parity** - Parity also uses a striping pattern, with a major difference. One or more blocks in each stripe holds parity data instead of live data. The operating system or array controller calculates parity data from the live data as it writes the stripe. If any disk in the array fails, it can use the parity data in place of the live data. A parity array can continue to function with the loss of one disk per parity block per stripe.



If you wish to use RAID, you can choose from a number of "levels". Each level of RAID provides its own balance of redundancy, speed, and capacity. With the exception of RAID-0 (pure striping for performance, no redundancy), all RAID levels require you to sacrifice space for protection. Disks present a relatively low expense when compared to system memory, and you have many expansion options beyond the base capacity of a system chassis. So, while RAID presents a higher cost per stored bit than single disk systems, it is usually not prohibitive.

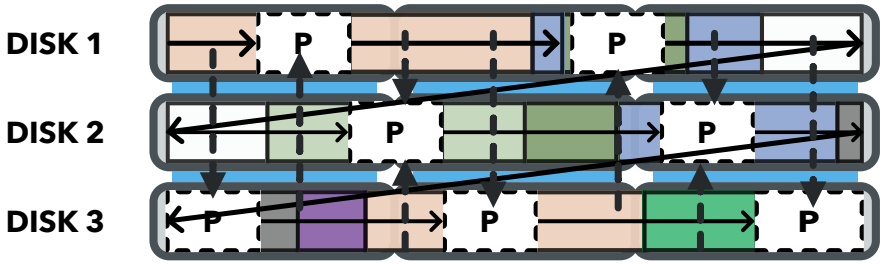
You have several choices when it comes to RAID. Many levels have fallen out of favor due to insufficient protection in comparison to others, and some simply consume too much space for cost efficiency. You will typically encounter these types:

- **RAID-1** - A simple mirror of two disks. Provides adequate protection, slightly lower than normal write speeds, higher than normal read speeds, and a 50% loss of capacity.



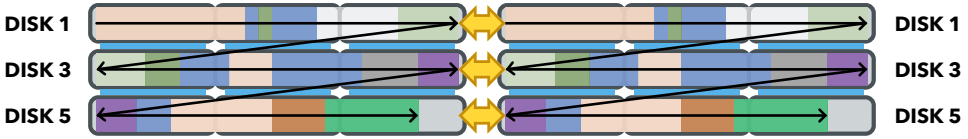
- **RAID-5** - A stripe with a single parity block. Requires at least three disks. Each stripe alternates which disk holds the parity data so that in a failure scenario, parity calculations only need to occur for 1/n stripes. Can withstand the loss of a maximum of one disk.

- Provides adequate protection, above normal write speeds, above normal read speeds, and a loss of 1/nth capacity. Not recommended for arrays that use very large disks due to the higher probability of additional disk failure during rebuilds and the higher odds of a failure occurring between patrol reads (scheduled reads that look for bit failures).



- **RAID-6** - Like RAID-5, but with two parity blocks per stripe. Requires at least four disks. Safer than RAID-5, but with similar concerns on large disks. Slower than RAID-5 and a capacity loss of 2/n.

- **RAID-10** - Disks are first paired into mirrors, then a non-parity stripe is written on one side of the mirror set, which is then duplicated to the corresponding mirror disk. Can function with the loss of one disk in each mirror but cannot lose two disks in the same mirror. Provides better performance and a higher safety rate than parity schemes, but at a loss of 50% of total drive capacity.



Due to the preponderance of drive failures and reduced performance of standardized redundancy schemes, many vendors have introduced proprietary solutions that seek to address particular shortcomings in RAID. Whereas RAID works at the bit and block levels, most vendor-specific systems add on some type of metadata-level techniques to provide protection or performance enhancements.

You have an overwhelming number of choices when it comes to fault-tolerant disk storage, so keep a few anchor points in mind:

FAULT-TOLERANT DISK STORAGE CONSIDERATIONS

Storage vendors naturally want you to buy their highest cost equipment. Use planning tools to predict your capacity and performance needs before you start the purchasing process. Businesses frequently overestimate their space and performance requirements.

You can almost always expand your storage after initial implementation. You do not need to limit yourself to the capacity of a single chassis as you do with system memory.

Solid-state disks have a substantially lower failure rate than spinning disks. You can leverage hybrid systems that incorporate both as a way to achieve an acceptable balance of performance, redundancy, and cost.

The most important point: downtime costs money. Storage redundancy directly reduces the odds of an unplanned outage.

ADVANCED STORAGE FAULT TOLERANCE

The advent of affordable, truly high-speed networking (ten gigabit and above) has brought exciting new options in storage protection. Today's networking speeds exceed even high-end storage equipment.
Once the sole purview of high-end (and very high-cost) storage area network (SAN) devices, you can now acquire chassis-level, and even datacenter-level, storage redundancy at commodity prices.

These technologies depend on real-time, or synchronous, replication of data. In the simplest design, two storage units mirror each other. Systems that depend on them can either connect to a virtual endpoint that can fail over as needed, or they connect to one unit at a time in an active/passive configuration. In more complex designs, control systems distribute data across multiple storage units and broker access dynamically.

The most advanced examples of these technologies appear in relatively new hyper-converged solutions. These use software to combine the compute layer with the storage layer on standard server-class computing hardware. In most cases, they involve a hypervisor to control the software layer and proprietary software to control storage.

While costs for distributed storage and hyper-converged systems have declined dramatically, they remain on the higher end of the expense spectrum. Unlike traditional discrete systems, you will need significant infrastructure and technical expertise to properly support them.

You can consider the duplicated data in this fashion as a "hot" copy. It's updated instantaneously and you can fail over to it quickly. Some synchronous replication systems even allow for transparent failover or active/active use.

APPLICATION AND OPERATING SYSTEM FAULT TOLERANCE

At the highest layer, you have the ability to mirror an operating system instance to another physical system. To make that work, you must run the instance under a hypervisor capable of mirroring active processes. It's a complex configuration with many restrictions. Few hypervisors offer it, it won't work universally, it won't survive every problem, and the performance hit might make it unworkable for the applications that you want to protect most.

At a more achievable level, some applications allow a measure of fault tolerance through tiering. For instance, you can often run a web front-end for a database. You can use load-balancers that instantly move client connections from one web server to another in the event of failure. Some database servers also allow for multiple simultaneous instances that can instantly redirect connections to a functioning node. These technologies have greater functionality and feasibility than operating system fault tolerance.

CAVEATS OF FAULT-TOLERANCE

As you explore options for fault-tolerance, you'll quickly notice that it comes at substantial cost. Almost all of the technologies will require you to purchase at least two of everything. Most of them will necessitate additional infrastructure. All of them depend on expertise to install, configure, and maintain. Those costs always need to be scoped against the cost of equivalent downtime.

The primary purpose of fault tolerance is to rely on duplicates to continue functioning during a failure. That has a negative side effect: your fault-tolerant solution might duplicate something that you don't want. For example, if ransomware attacks your storage system, having RAID or a geographically redundant SAN will not help you in any way. Even in the absence of a malicious actor, redundant systems will

happily copy accidental data corruption or delete all instances of a vital e-mail on command.

While fault tolerance will serve your organization positively, it cannot stand alone. You will always need to employ a backup solution for asynchronous data duplication. However, you have options between fault-tolerance and backup. Those technologies reside in the high availability category.

SECOND LINE OF DEFENSE: HIGH AVAILABILITY

You can't use fault tolerance for everything. Some systems have no way to implement it. Some have a prohibitively high price tag. Instead, you can deploy high availability solutions. High availability has a more nebulous definition than fault tolerant. It applies less to actual technologies and more to outcomes. Where fault-tolerance means working through a failure without interruption, high availability measures uptime against an expected metric.

As an example, your organization sets a target of 99.99% annual availability for a system that they want working at all times. To achieve that, you would need to ensure that the system does not experience more than a few minutes of total downtime in the course of a year. 365 days times 99.99% equals 364.9635 days of uptime, which allows a little less than 48 minutes. That's an aggressive goal, but not necessarily unachievable.

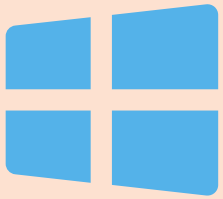
From the technology angle, any tool that specifically helps to improve uptime falls under the high availability umbrella. All fault-tolerant technologies qualify. However, you also have some that allow a bit of downtime in exchange for reduced cost, wider application, and simpler operation. Among these, clustering is generally the most common.

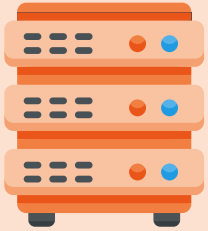
HIGH AVAILABILITY WITH CLUSTERING

Clustering involves using multiple computer or appliance nodes, usually in an active/passive configuration, to host a single-instance resource. Some examples that depend on Microsoft's failover clustering technology:

- **Microsoft SQL** - A clustered Microsoft SQL database runs on one of many nodes. In a planned failover, the database becomes unavailable for a few seconds while its active node stops and one of the passive nodes start. In the event of active node failure, the database is offline for a few seconds while a passive node starts it. Active transactions might drop in an unplanned failover.

HIGH AVAILABILITY WITH CLUSTERING MICROSOFT EXAMPLES





- **Hyper-V** - A clustered virtual machine can quickly move online (Live Migration) or offline (Quick Migration) to another node in a planned failover. If its active node fails, the virtual machine crashes but another node can quickly restart it.
- **File server** - The standard clustered Microsoft file server hosts through an active node, with planned and unplanned failovers occurring quickly. Microsoft also provides a scale-out file server, which operates in a more fault-tolerant mode.
- **Storage Spaces Direct** - Commonly called “S2D”, Storage Spaces Direct is Microsoft’s distributed file system offering. It can also operate on Hyper-V nodes, providing a complete hyper-converged infrastructure.

You will find clustering technologies in other operating systems, hypervisors, and physical appliances. Remember that these differ from fault-tolerance in that they allow some downtime. However, they should greatly reduce downtime risks when compared to standalone systems.

CAVEATS OF CLUSTERING

Clustering provides a duplicate of the compute layer. It ensures that a clustered workload has somewhere to operate. It does not make any copies of data. Without additional technology, a critical storage failure can cause the entire cluster to fail.

Because of the necessity of hardware duplication, clustering costs at least twice as much as operating without a cluster. You might also need to purchase additional software features in order to enable a clustered configuration. Clustering requires staff that know how to install, configure, and maintain it.

You must also take care that the backup solution that you choose can properly protect your clustered resources. Solutions such as Altaro VM Backup protect Hyper-V and VMware clusters. You can sometimes successfully employ a backup solution that doesn't interoperate with your high availability solution, but it will require significantly more administrative effort.

HIGH AVAILABILITY WITH ASYNCHRONOUS REPLICATION

You can employ technologies that periodically copy data from one storage unit to another. Asynchronous replication can use a snapshotting technique to maintain complete file system consistency. Some replication applications use a simple file-copy mechanism, which works well enough for basic file shares but not for applications.

Some applications have their own asynchronous replication built in. Microsoft's Automatic Directory will automatically send updates between domain controllers. Most SQL servers have a set of replication options. Microsoft Hyper-V can create, maintain, and control virtual machine replicas.

You can consider data created by asynchronous replication as a "warm" copy. It does require some sort of process to bring online after a failure, but you can place it in service quickly.

CAVEATS OF ASYNCHRONOUS REPLICATION

Unlike clustering, asynchronous replication requires some human interaction to switch over to a copy after a failure. Clustering technologies use some sort of control technique to prevent split-brain situations in which two copies run actively and simultaneously.

Most replication systems have no built-in way to do that. So, if you choose to implement replication, ensure that you plan accordingly.

Replication shares the main drawbacks of clustering: it requires duplicated hardware, special software, and expertise. It also does not protect against data corruption, including ransomware.

THE UNIVERSAL FAIL-SAFE - BACKUP

Out of all available disaster recovery and business continuity technologies, only backup is both sufficient on its own and necessary in all cases.
You can safely operate an organization without any fault tolerant or high availability technologies, but you cannot responsibly skip backup.

Before you start shopping, ensure that you understand common backup terms:

- **Full backup** - A complete, independent duplication of data that you can use to recover all data without any dependency on any other data.
- **Differential backup** - An abbreviated backup that only captures data that changed since the most recent full backup. Usually operates at the file level.
- **Incremental backup** - An abbreviated backup that only captures data that changed since the most recent backup of any kind. Usually operates at the file level.
- **Media** - Storage for backups. Intended as a catch-all word whether you save to disks, tapes, optical discs, or anything else.
- **Delta** - In backup parlance, delta essentially means “difference”. Most backup vendors use it to mean a measurement of how a file or a block has changed since the last backup. Most vendors use the term “delta” in some way to designate technology that operates below the file level.
- **Crash-consistent** - A crash-consistent backup captures a system’s data at a precise point in time. It is referred to as “crash-consistent” because if you restore to such a backup, the system will act exactly as though it had crashed when the backup was taken. A crash-consistent backup does not protect any running processes, nor does it give them any opportunity to save active data. However, it captures all files exactly as they were at that moment.

- **Application-consistent** - An application-consistent backup interacts with applications to give them an opportunity to save active data for the backup. If the backup tool does not have a way to notify a particular application, it will save that data in a crash-consistent state.

- **Restore** - The act of retrieving data from a backup. Restoration can return data to a live system or to a test system. Most tools allow you to choose between complete and partial restores.

- **Rotation** - Re-using backup media, usually by overwriting older backups. Some backup software has intricate rotation options.

Not everyone agrees on the definitions of “crash-consistent” and “application-consistent”, and some vendors have introduced their own labels. Ensure that you understand how any given vendor uses these terms when you study their products and talk to their representatives.

As you explore backup solution choices, you need to use the plan created by your business teams as a guideline. You want to try to satisfy all requirements for data protection and retention. Consider these critical components of backup technologies:

BACKUP COMMANDMENTS

1. Backups must create a complete, standalone duplicate of data
2. Backups must maintain multiple unique, non-interdependent copies of data
3. Backups should complete within your allotted time frame
4. Backups should provide application-consistent options
5. Backups should work with the type of backup media that you want to use
6. Backups should work with your cloud providers, both to protect your cloud resources and to back up to your cloud storage account(s), as desired

The above list only constitutes a bare minimum. Realistically, all backup vendors know that they need to hit these targets, so only a few will miss. Usually, those are the built-in free options or small hobbyist-style projects. You will find the greatest variances among the last two items.

Products will distinguish themselves greatly in operation and in optional features. You should avail yourself of trial software to experience these for yourself. Altaro VM Backup provides several outstanding and noteworthy features that you can look for in all of the products that you try. Some things to look for:

- **Ease of operation (especially restores)** - In a disaster, you cannot guarantee the availability of your most technically proficient staff, so your backup tool should not require them.
- **Speed of operations** - Backup and restore operations need to complete in a reasonable amount of time. However, they cannot sacrifice vital functionality to achieve that. Most backup vendors utilize some sort of deduplication technology to reduce time and capacity needs, but you absolutely must have a sufficient number of non-interdependent copies of your data.
- **Retention lengths** - Most backup applications allow an infinite number of backups – except in their free editions. If your organization won't allow you to spend money on backup software, that might prevent you from achieving their requirements.
- **Support for the products that you use** - As mentioned earlier in this book, very few backup applications know anything about line-of-business software. However, they should handle the operating systems and hypervisors that you use. Some will have advanced capabilities that target common programs, such as mail and database servers. If you choose a solution that does not natively handle your software, ensure that you know how to use it to perform a proper backup and restore.
- **Offsite support** - Because you will use backup to protect against the loss of your primary business location, your backup tool needs to have some method that allows you to take backup data offsite. Traditionally, that meant some sort of portable media. Today, that also means transmitting to an alternative location or a cloud provider.

- **Support for alternative hardware** - After a disaster, you probably won't have the luxury to restore data to the same physical hardware that it protected. Make sure that your backup application can target replacement equipment.
- **Technical support options** - Hopefully, you'll never need to call support for your backup product. However, you don't know who might need to perform a restore. That task might fall to a person that will need help. You also need to consider future product updates and the possibility of bugs that need attention. Ensure that you understand your backup provider's support stance and process. If possible, try to talk to them before purchase.

Consider data created by backup as a "cold" copy. You must take some action to transition the data from its backup location before you can use it in production. It usually has a much higher time distance from the failure point than replication.

WHAT'S NEXT

You have now seen all of the planning concepts. You have enough knowledge to tackle the planning phase of your disaster recovery strategy. We will close out this book by wrapping everything together with an template action plan for developing your backup and disaster recovery strategy.

CHECKLISTS AND QUESTIONNAIRE

No two companies will build the same disaster recovery plan. Your final plan will apply only to your organization. However, much of the initial work looks similar regardless of your size or industry. These checklists will serve as templates to help you build your own solution. Do not consider them as the only approach. Feel free to adjust them as necessary.

MEETING CHECKLIST

Disaster recovery touches every part of your enterprise, so it needs input and participation from all corners. You will need to organize and conduct several meetings.

- ☐ **Business case meeting** – explain the need for a disaster recovery plan
  - ○ Invite executives and department heads
  - ○ Explain the consequences of inaction
  - ○ Secure a willingness to commit to funding
  - ○ Secure a commitment of personnel time
  - ○ Identify key stakeholders
- ☐ **Scope meeting** – begin discovering the scope of your disaster recovery plan
  - ○ Invite key stakeholders
  - ○ Explain the desired data points
  - ○ Provide a checklist or questionnaire
  - ○ Allot time for completion with a deadline
- ☐ **Scope follow-up meeting**
  - ○ Invite key stakeholders or designees
  - ○ Collate data points
  - ○ Discover overlaps
  - ○ Schedule additional time and meetings as necessary
- ☐ **Technology presentation meeting**
  - ○ Invite executives and stakeholders
  - ○ Show options and possibilities
  - ○ Begin pricing discussions
- ☐ **Additional planning meetings** – further topics
  - ○ Roles and responsibilities
  - ○ Implementation planning
  - ○ Procedure design

Remember not to make these discussions solely about data and technology. You also need to prepare for losses of physical assets and personnel. More business continuity procedures will deal with employees and activities than with disks and tapes. Use these meetings as opportunities to explore topics such as work-from-home policies during a disaster.

DATA PROTECTION QUESTIONNAIRE

As a technology professional, you know how to protect data.
As company experts, your key stakeholders know what data to protect.
You need to combine your knowledge into an actionable plan.
You can use questionnaires as a way to gather the necessary information. Use the following questionnaire as a starting point to design your own.

- Which employees have the most knowledge of the computer and data systems that your department relies upon?
- Where does your critical data reside?
  - Network servers
  - Centralized desktops
  - Employee laptops
  - Cloud providers
- What computer and software systems do you require for full operational functionality?
  - Line-of-business applications
    - Workstation-based
    - Server-based
    - Cloud-based
  - Commodity applications
  - Internally developed applications
- What technologies do you require for full operational productivity?
  - Communications devices
  - Specialty hardware
  - Commodity hardware (desktops, laptops, printers)
- What personnel does your department rely on? What if those individuals are not available?
- Do you currently protect your data? Technology? Assets?
- Who supports your technologies?
- What technologies do you require for minimal functionality?
- How long can you operate at minimal functionality?

- What technologies do you require for acceptable functionality?
- How long can you operate at acceptable functionality?
- Estimate the departmental cost of an hour of complete downtime.
- In the event of a system failure, how far back could your department recreate data?
- How long must you keep copies of your various data points?
- Prioritize your technology

Make certain that everyone with knowledge receives a copy of the questionnaire. Even if an individual cannot provide an answer to every question, gather as much information as possible.

INFORMATION TECHNOLOGY DEPARTMENT CHECKLIST

As the other departments work on their questionnaires, you need to gather your own information.

- ☐ List of key stakeholders
- ☐ List of application experts
- ☐ List of data experts
- ☐ Organizational contacts
- ☐ Support contacts
- ☐ Mission-critical categorized systems and data
- ☐ Important categorized systems and data
- ☐ Low priority categorized systems and data
- ☐ Available technology solutions, capabilities, and costs

It will also be your responsibility to process all of the information as it comes from the departments. Create a usable plan, share it with the stakeholders, and acquire funding for the project.

WHAT'S NEXT

This concludes the first part of The Backup Bible. Once you have completed everything above, you can move on to the implementation phase of your project – best practices. The next book in the series will walk you through what you can expect to do and encounter on a day-to-day basis. If you have never worked with this sort of technology before, I recommend that you read ahead before finalizing the planning phase.

ALTARO BACKUP

HYPER-V | VMWARE | PHYSICAL | OFFICE 365

Hyper-V & VMware Backup & Replication

For Companies

Award-winning virtual machine (VM) backup and replication solution for Hyper-V and VMware environments

Learn more

For MSPs

Monthly subscription program enabling Managed Service Providers to offer Hyper-V, VMware and physical server backup services

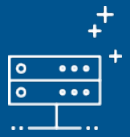
Learn more

Physical Server Backup

Back up the physical servers on your network through this P2V solution and benefit from fast and easy recovery should they be impacted by a disaster

Learn more

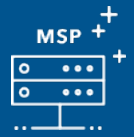
Office 365 Backup



For Companies

Office 365 mailbox backup and recovery solution with centralized backup management and storage to Altaro's Microsoft Azure infrastructure

Learn more



For MSPs

Monthly subscription program enabling Managed Service Providers to provide Office 365 mailbox backup, recovery and backup management services

Learn more

50,000+ Customers | **10,000+** Partners | **2,000+** MSPs | **121** Countries

MORE GREAT IT CONTENT

Continue your learning on the Altaro Dojo, our dedicated learning platform for IT professionals:

FOLLOW ALTARO AT:

SHARE THIS RESOURCE!

Liked the eBook? Share it now on:

ALTARO BACKUP

PUBLISHED BY ALTARO SOFTWARE

https://www.altaro.com/

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 authors 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.

FEEDBACK INFORMATION

We’d like to hear from you! If you have any comments about how we could improve the quality of this book, please don’t hesitate to contact us by visiting www.altaro.com or sending an email to our Customer Service representative Sam Perry: sam@altarosoftware.com