rubrik

Buyer's Guide to Backup and Recovery

TABLE OF CONTENTS

INTRODUCTION

Backup and recovery needs a radical rethink. When today's incumbent solutions were designed over a decade ago, IT environments were exploding, heterogeneity was increasing, and backup was the protection of last resort. The goal was to provide a low cost insurance policy for data, and to support this increasingly complex multi-tier, heterogeneous environment. The answer was to patch together backup and recovery solutions under a common vendor management framework and to minimize costs by moving data across the infrastructure or media.

The underlying constants have been—but also continue to be—the need for backups to be reliable and restores to be fast and dependable.

What has changed? For one, IT departments have moved toward private cloud models with virtualization and converged architectures replacing multi-tier architectures. Secondly, since the amount of data under management has exploded, IT is challenged to do more with less. IT teams are now composed of fewer specialized roles and more broad roles. Finally, public and hybrid clouds have opened up new data use cases, such as analytics and test/dev, that create challenges for managing that data.

Any IT professional considering a backup investment should ask which old assumptions are still relevant and whether a new approach is better. In this guide, we'll discuss backup and recovery and the emergence of Cloud Data Management, which provides opportunities for protecting data, capturing new value, and making data available whenever, wherever it is needed.

Special thanks to those from the IT community who provided their perspective on these topics.

We are in a new era where business is driving change within IT. Solutions must first match business needs and then technical needs.

Andrea Mauro, @Andrea_Mauro

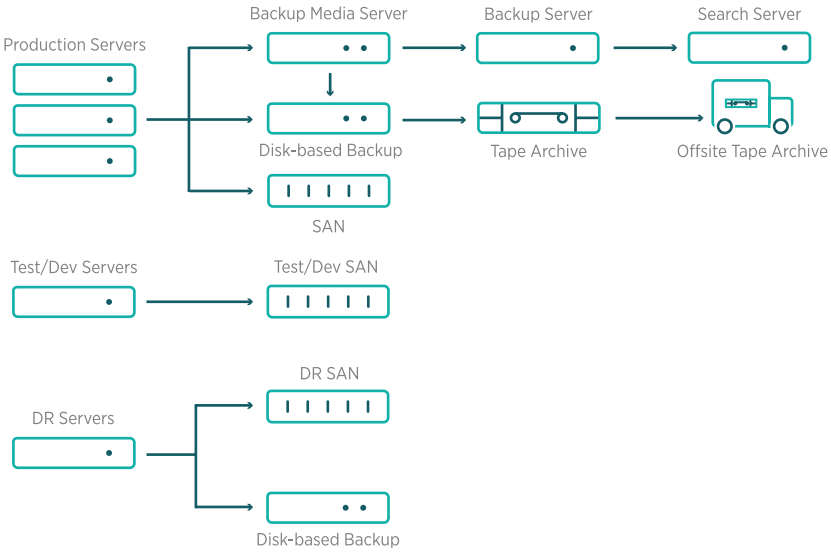
vinfrastructure.it

WHAT MATTERED YESTERDAY

TRADITIONAL BACKUP REQUIREMENTS AND CHALLENGES

The first batch of backup and recovery solutions were built to address the challenges of application tiers powered by heterogeneous infrastructure. As the platform of last resort, backup and recovery solutions became the point of logical consolidation. Backup systems needed to move large amounts of data across sprawling environments and manage it across multiple media types like disk and tape to control costs. Traditional backup systems also needed to satisfy long-term data retention requirements, generally using offsite tape archives.

Your Data Management Today



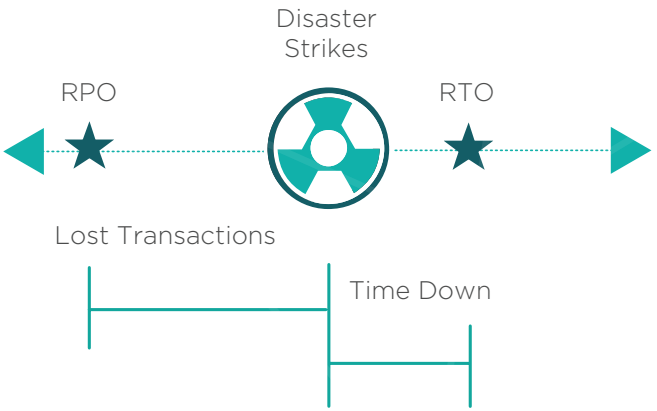
Complexity of today's environments -
multiple systems, vendors, and architectures

JOB AND SCHEDULING FLEXIBILITY

Historically, the focus was often on backup windows and job schedules. These areas were necessary to achieve the desired RPO & RTO for a business. Unfortunately, this often turned backup engineers into “glorified job schedulers” - an unintended complication of complex architectures.

Recovery Point Objective (RPO) defines the point in time used for restoration and is determined by the frequency of backups. In the event of a primary system failure, a lower RPO means less data loss. Backup and recovery systems achieve low RPOs by taking more frequent backups at the expense of more traffic traveling across the network and more copies of data stored. In the case of mission-critical applications, RPOs need to be available as points in time measured in minutes, as opposed to hours or days.

Recovery Time Objective (RTO) defines how long it takes to recover an object such as a file, server, or data center. A lower RTO means less downtime in the event of a primary system failure, but at the expense of using more expensive, faster access media like a disk, as well as costly network switches to move data back to where it can be accessed.



Visual Representation of RPO and RTO

RPO and RTO are the hidden heroes when it comes to recovering from a business disruption in IT. Being able to be confident that you can meet or exceed the SLA you've got with the business is what makes the difference between hope and a real proven strategy.

Eric Wright, @discoposse

discoposse.com

REPLICATION

Replication is the capability to copy data from a primary location to a secondary location. This is often referred to as disaster recovery because it is protecting against a site-wide failure at the primary location. Replication only indirectly relates to RPO and RTO; the assumption is that most failures occur at the subsystem level rather than the site level. However, replication is a common requirement for critical applications.

BUSINESS REQUIREMENT TRANSLATION

With all this in mind, the key requirement for any backup and recovery solution is to take business level requirements for recovery time and data recovery, otherwise known as Service Level Agreements (SLA), and to translate them into a set of instructions for placing, retaining, and expiring data on different storage media.

The main problem with traditional backup and recovery systems is that the "translation" from business requirements to platform executable instructions requires professional services. In other words, traditional solutions have imperative vs. declarative operating models. Moreover, once this translation is complete, these traditional solutions lack intelligence to optimize resources to avoid failed backups. This in turn leads to ongoing tuning and sometimes re-architecting.

The best way to evaluate RPO and RTO in your current system is to ask an executive to pick some data (use different granular types) from random points in time. Quantify how close to the RPO you can achieve and the RTO of the recovery. Compare this to the cost of downtime while recovery is in place. This has always been one of the most challenging areas of backup and recovery systems.

Data protection is a critical part of every Business Continuity Plan because data loss or corruption may have a huge impact. While there are different solutions to solve different aspects of recovery and data redundancy, most important is to have a clear approach to achieve good RPO and RTO.

Andrea Mauro, @Andrea_Mauro

vinfrastructure.it

If you don't understand your RPO and RTO requirements, you don't understand your workloads. And if you don't understand your workloads, you don't value the business data, and you don't value the business.

Dan Frith, @penguinpunk

penguinpunk.net

TAPE BACKUP AND LONG-TERM RETENTION

Backup and recovery is typically designed for short-term data retention up to one month. Archive is used for long-term data retention with 1-7 years being common timeframes. Long-term data retention is especially important in businesses that require regulatory compliance, such as healthcare or financial services.

Until recently, the only economically viable choice for archival has been tape. For all but the largest enterprises, tape involves manual handling, offsiting, logging, and rotation of tapes. Restoring from tape is time consuming, manual, and complicated because tapes are typically stored offsite, and a single file restore requires a broader system or volume restore. In addition, tapes degrade over time and must be refreshed.

Tape also lowers the value of data by sequestering it. Tape-archived data is typically poorly indexed and limited in accessibility. By placing your most valuable strategic asset in a vault, your data locks in and reduces its value to the business.

In some countries, tape archive was explicitly mandated to meet data retention, legal, and regulatory requirements. Yet, an increasing number of agencies and jurisdictions are adapting data retention policies to specify functional requirements rather than media.

If it's complex, it means it can be done incorrectly in more ways. The simpler the process, the less likely it will be misconfigured.

Eric Wright, @discoposse

discoposse.com

COMPLEX INSTALLATION AND CONFIGURATION

Configuration and installation of enterprise backup has always been a challenge. Almost all vendors require professional services to install and configure a backup system to the point at which all promised functionality is available. In order to use the system, administrators often must attend a week-long training.

Backup and recovery needs to be intuitive for the average administrator to use. In addition to being easy to setup and administer, backup should be automated as much as possible. This ensures that new systems are automatically protected when added.

> "Any solution with a sufficient level of complexity is doomed to suffocate on its own technical debt. Focus on solutions that are easy to use and easy to manage to gain the trust of IT and the business."
>
> Eric Shanks, @eric_shanks, theithollow.com

COST

The cost of backup and recovery has always been a significant part of the IT budget. As data has grown exponentially, so has the cost of backing up and storing data. Sometimes, data protection even costs more than primary storage. Experienced IT organizations would often budget for 2-3x the primary data cost to cover data protection and backup costs.

WHAT MATTERS TODAY

MODERN BACKUP REQUIREMENTS AND CHALLENGES

If you built a present-day backup and recovery solution, what would that look like? How would it be similar to a traditional backup solution? The requirement to support customizable SLAs based upon RPO and RTO, disaster recovery, and archival capability would stay the same.

What would be different? Since IT departments are increasingly adopting hybrid cloud models, they need (hyper)converged infrastructure with modular scalability and increasing levels of virtualization. The solution should also reflect that enterprise IT teams are adopting technologies like IoT, big data, and DevOps to harness the value of data. And lastly, it would need to prioritize security, as ransomware attacks and data leakage are constantly growing threats.

Virtualization makes it possible to protect not only files and application data but also the entire VM - effectively a simpler "Bare Metal Restore."

Andrea Mauro, @Andrea_Mauro

vinfrastructure.it

VIRTUALIZATION

Most backup systems were originally designed to support physical hosts. Virtualization was the last major wave of computing innovation. Before virtualization, systems' RAM and CPU were underutilized, which provided resources during off hours for backup processes. Through virtualization, overall RAM and CPU usage were driven much higher, and storage moved onto a central array. Without careful planning or newer technology, backups can push virtualized systems past their resource maximums.

> Virtualization has accelerated environment sprawl and the assumption that it should "just work" because of how easy it has become to provision a server. We automatically assume that we should be able to protect and recover those resources.
>
> Eric Wright, @discoposse, discoposse.com

SIMPLICITY AND AUTOMATION

IT departments are getting leaner with specialized roles such as dedicated Backup Admins being replaced by more generalized roles such as IT Admins. Backup and recovery solutions need to be simple to use with well-designed user interfaces - not requiring extended training for day-to-day use.

They also need to embrace declarative operations, with low-level instructions handled by heuristics and intelligence, requiring minimal administrator intervention. Solutions also need to be easily extensible and automated with RESTful APIs to embrace popular scripting tools such as Chef, Puppet, and Ansible, among others.

Automation is no longer just for cutting edge IT shops. The every day IT admin will soon be executing API calls to provision and manage the infrastructure that supports their applications. A robust API must be part of these data center solutions if they plan to be relevant.

Eric Shanks, @eric_shanks

theithollow.com

SHORTER BACKUP WINDOWS VS. LARGER ENVIRONMENTS

Data continues to grow and modern IT departments are managing more data than ever while backup windows continue to shrink. Backups need to protect more data, more reliably, in a shorter period of time. Newer approaches such as snapshot-based backups remove the need to stop applications for backup, reduce the backup window, and eliminate resource loads (placed by backup agents) on hosts.

CLOUD USAGE AND APPLICATION AGILITY

Another current trend is the increasing use or planned use of cloud to run agile workloads or cloud-native apps. Cloud storage and archive is also rapidly declining in cost. So, enterprise IT departments cannot ignore the nearing demand for data to be either located in the cloud or easily orchestrated to and from the cloud or private cloud. New features such as cloud archive can be an elegant method for migrating data to the cloud. Typically, production data will reside on a local backup system.

Many customers use tape because of the requirement that multiple copies of backups be stored separately from the primary data location. Replication to the cloud satisfies this requirement and often allows the elimination of tape backups, unless obligated by regulation.

SECURITY AND ACCESS CONTROLS

Data theft and other cyber crime continues to proliferate in sophistication and frequency. Security threats can now originate inside a firewall or in the cloud. Data that is in the hybrid cloud environment needs to be secured with encryption at rest and in flight, and it needs to have proper access controls to prevent theft and damage.

BACKUP SECURITY AND RANSOMWARE

Backup theft is a favorite tool of hackers and identity thieves. Special measures are required to prevent tampering and to ensure regulatory compliance of sensitive data. Among other items, examine your backup system to see if it is vulnerable to underlying OS security issues. Are your backups immutable? Can you audit access logs?

Ransomware is malicious software that blocks access to your systems until money is paid to remove the code. While estimates vary, over 40 percent of organizations will likely experience a ransomware attack each year. A backup and recovery system should enable you to recover quickly and with confidence. Lacking a system that efficiently recovers from a ransomware attack can have major financial and reputational impacts on your company.

WHAT SHOULD MATTER TO YOU

SELECTING A BACKUP AND RECOVERY SOLUTION FOR TODAY AND TOMORROW

Changing backup vendors always requires some level of technical or organizational effort. So, if you're considering a change, here are some questions to ask: Is the backup solution something that has been around for years with little change? Is it something created from multiple company acquisitions that have been cobbled together? Does it require multiple systems and interfaces for managing backup and recovery, replication, archive, and compliance?

Backup innovation starts with recognizing that a new approach needs to align with the dramatic technology changes and data growth in centers over the last decade. Today, there are new approaches from industry visionaries who understand current business needs and how quickly the industry changes.

Cloud can be used as a low cost and long-term target for backup data or archive data. In this case, the lock-in risk could be minimal because you can just move your data to another cloud provider.

Andrea Mauro, @Andrea_Mauro

vinfrastructure.it

CLOUD DATA MANAGEMENT

Cloud Data Management is designed to orchestrate mission-critical application data across private and public clouds while unifying backup, instant recovery, replication, search, analytics, archival, compliance, and copy data management into one scale-out fabric. Solutions built for the cloud generation remove the complexity of legacy systems via an automated policy engine that manages data throughout its lifecycle across all data management functions. The vision is to deliver a software fabric rooted in a cloud vendor-agnostic platform, preventing vendor lock-in to any particular cloud.

ECOSYSTEM SUPPORT

100% virtualization is where many enterprises are headed. A modern day backup and recovery solution should be designed to support, optimize for, and integrate with virtualized environments. Methods like flash can handle large volumes of data in virtual environments without impacting production. Look for features such as global indexed search that enable quick data access and granular instant recovery.

However, many enterprises are not fully virtualized. A strong backup and recovery solution delivers full capabilities to physical environments and allows you to manage all environments from a single interface.

Backup of NAS devices should be done in a vendor-agnostic method without vendor-specific plugins or proprietary storage formats. Additionally, this should be in a native format unlike NDMP that does not require “unpacking” before a restore. An ideal solution requires periodic full backups but is incremental forever. Finally, all capabilities of the backup platform should be available for NAS backups.

Lock-in is always a challenge but we have to accept that it is part of the deal. As long as you have processes that give you as much flexibility as possible, the lock-in risk can be reduced. Cloud is a huge asset for backup and recovery for both cold data and potentially sandbox-style recovery. For good reason, public cloud will be a big thing on the mind of every CIO for active workloads, as well as recovery.

Eric Wright, @discoposse

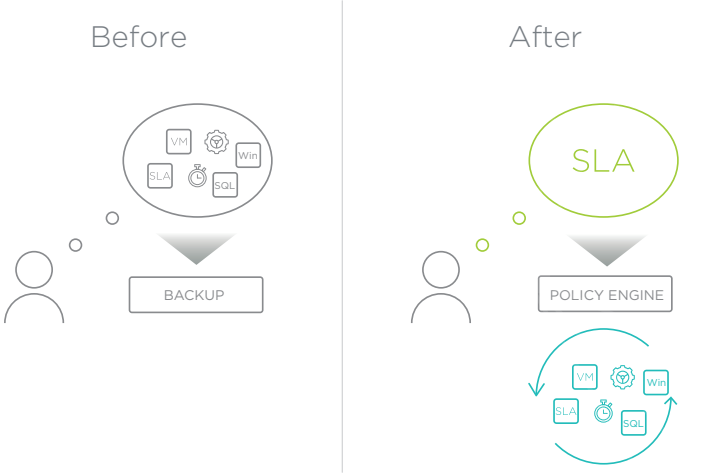
discoposse.com

A DECLARATIVE POLICY ENGINE AND AUTOMATION

With less specialization among IT admins, simple backup has become a non-negotiable item for many organizations. Backup and recovery solutions must be usable by all team members.

While traditional architecture has relied on the imperative model, the declarative model is simple to understand and can easily align with business goals. With a declarative model, an admin inputs his or her desired state for a workload into a policy engine. Once a policy is set, the intelligent system executes. Ironically, a strong policy engine even removes some of the traditional need for automation as the system can handle requirements with less manual steps.

Let Your Policy Engine Do the Thinking



SLA Policies allow you to collapse multiple manually implemented settings into a single easy-to-configure and zero-maintenance policy

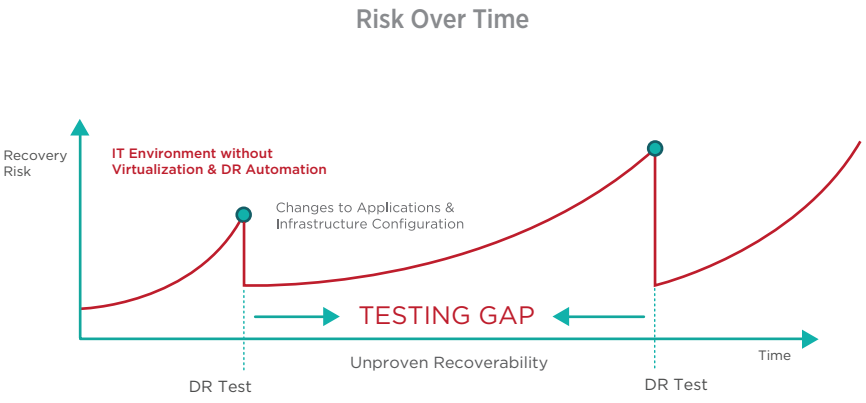
A solution with API-first architecture allows for further automation. Connect with third-party services to automate data protection, recovery, and other custom data management workflows.

Automation is critical. Automation means consistency and reduced need to spend valuable engineering time doing day-to-day processes – everything from the menial tasks up to more complex provisioning and protection processes. An hour spent automating a process can net back literal weeks.

Eric Wright, @discoposse

discoposse.com

Additionally, automation enables regular backup validation - a requirement to mitigate the "testing gap" risk as shown below. If backups aren't tested regularly, IT cannot guarantee their validity to the business.



Without regular testing, guaranteeing reliable restores is impossible

SECURITY AND COMPLIANCE

Management includes ability to configure role access to data usage, compliance reporting, and ability to monitor system events, operational tasks, capacity, logs, and user events.

Secure data involves encryption both at-rest or in-flight, key management, and the ability to instantly recover from breaches such as ransomware. If you're in an industry with strict compliance policies like HIPAA, your backup solution should be designed to align with these requirements.

EASY SCALABILITY

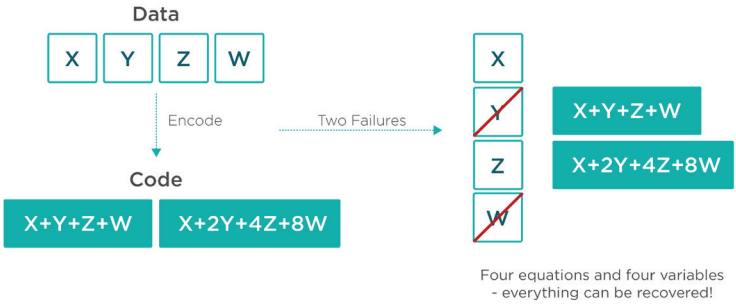
Like the modern primary environments they support, backup and recovery solutions need to quickly and easily scale, or they risk being the bottleneck to growth. Thus, a modern solution should be built using commodity hardware and scale-out software with simple cluster management. This solution should scale from TB of data to PB of data with consistent performance and usability.

When selecting a backup vendor, it's important to know how easily the solution scales and the maximum size to which it can scale. Metadata and data should be distributed across all nodes within the cluster and support global deduplication. No single management node should be a point of failure, and the system should have self-healing capabilities. When the system has a node failure, find out if system restores are as efficient as when the system if fully functional.

Typically, backup solutions also have a specified amount of data that can be backed up. Once this limit is reached, a separate system is required. A truly scalable solution should allow you to add backup nodes that take advantage of current technology and scale to your entire environment. This allows you to find data from a single source and to take advantage of global deduplication. Adding nodes should be an easy process that does not require days of data rebalancing or professional services to manage.

Data efficiency is another important component of a scalable solution. Modern backup solutions use coding methods such as erasure coding that are fault tolerant and increase storage capacity without impacting performance.

Data Efficiency via Erasure Coding



Modern protection methods enable faster rebuilds with lower storage space overhead

COST VS. VALUE

Understanding the true cost of backups is extremely difficult. You need to know how much data you have, the type of data (structured or unstructured), the amount of granularity required for RPOs, and how long the backups will be stored. The cost of backups includes the software and hardware, the cost of the WAN for replication and backup, co-location costs, and the cost of business revenue loss and productivity during the recovery window.

IMMUTABILITY

As discussed above, ransomware attacks are becoming increasingly common. Having immutable backups that cannot be encrypted by ransomware are a critical part of a protection strategy. Ask your backup provider if they can guarantee that backups are immutable and can not be encrypted by ransomware even if the system is misconfigured.

BEYOND PROTECTION

Data protection is no longer just for insurance. Most businesses now expect to leverage their backup and recovery platform for additional use cases - archive to cloud, test/dev environments, migration to the cloud, and more.

At the same time, these new and often secondary use cases must not detract from the primary purpose - adding functionality that diminishes from the primary purposes of backup, and recovery is not feasible operationally.

In short, look for platforms that can help you leverage your data for use cases beyond backup and recovery to drive business initiatives. Make sure the additional capabilities are not simply “checkbox” features but can provide true value to your business.

CONCLUSION

When looking at backup and recovery solutions, the solution must be simple and scalable. It should support data portability and accessibility in a cloud era. The total cost of ownership should be less than what you are paying for your legacy system and offer innovative features that allow you to do more with your data.

Rubrik delivers instant application availability to hybrid cloud enterprises for recovery, search, cloud, and development.

Use the Rubrik Cloud Data Management platform to mobilize applications, automate protection policies, search and analyze application data at scale, and even recover from ransomware – all on one platform.

To learn more, visit www.rubrik.com.