EBOOK

Where Is Your Golden Copy?

Protecting Your Unstructured Data

Datadobi

Where Is Your Golden Copy?

As more organizations put their critical data on NAS and object systems or in the cloud, a complete, well-thought-out data protection strategy is critical. Data is constantly at risk from cyberattacks, insider threats, storage platform-specific hacks and bugs, and natural disasters.

In a recent survey, 65% of companies reported lost productivity due to data loss or outages – suggesting that current protection efforts are inadequate.

Traditional data protection strategies, though essential, are insufficient in completely protecting an organization's most critical NAS and object data.

Companies are simply creating a copy of their business-critical unstructured data in a bunker site, either in a remote data center or in the cloud. This "golden copy" serves as a way to get their business up and running as quickly as possible if the worst does occur.

Your Unstructured Data Protection Strategy

You can think of unstructured data protection strategies as layers of protection. Each layer provides a copy of data that can be used to restore availability should the worst happen. And each layer also provides for different recovery point objectives and recovery time objectives, as well as different recovery scopes ranging from individual bits of data, or files, to complete data sets.

These data protection strategies typically consist of local copies (RAID, snapshots, local backups) and off-site copies (replicated copies and synchronized copies).

Beyond the production copy, a typical strategy is to regularly copy data to a secondary system in an off-site facility. From this system, data can be restored or the users and applications can be redirected to that secondary system if the production system becomes unavailable or its data unusable.

But for most organizations, loss of data from their primary and disaster recovery (DR) site means a complete loss. Even if they can restore from a backup, it could be days or weeks later, creating a substantial business impact given that average data outage costs run at over $300K/hr.

Data outage costs run at over $300K/hr.

What’s Special About Unstructured Data?

While replicating block data is relatively straightforward – simply copy raw blocks from one system to another with no need for an understanding of the actual data or how it is used, structured or controlled – the replication of unstructured data is far more complicated.

With NAS storage, beyond just having a copy of the actual data, the remote site must also contain a regular mirror of the filesystem structure, the shares and exports required for end-user and application access, and all permissions (NFS and/or SMB) that control file access.

Object storage protocols, such as S3, support attaching a small amount of metadata to each object. To allow fast recovery and correct functioning of applications consuming the S3 storage, the object metadata is as important as the object data itself.

Without all this information applied, it would be impossible for a golden copy to be quickly restored at a granular level or act as a failover target.

The Limits of Backup and Array-based Replication

Although many organizations protect their NAS data with local backups and implement array-based, off-site replication, data and backups still remain at risk due to the inherent limitations of data protection strategies.

Local Copies

Typically local copies of data used in a data protection scheme fall into two categories, snapshots and backups.

Snapshots offer near-instant recovery but are very expensive as they consume primary storage space for each copy created, so relatively few copies are kept, meaning limited options in terms of historical copies.

Backups take far longer to recover and present an extended outage event as data must be restored to the NAS platform from a different storage media such as another disk system or tape.

Both of these solutions create point-in-time copies, which is ideal, but recovery depends on a valid copy being available, which may not be the case as both snapshot and backup copies are expensive to maintain and will be aged out and discarded over time. In addition, both solutions are vulnerable to a single site loss or outage.

In addition, malicious attacks will often focus on destroying backups first, since a corrupt backup often goes unnoticed for a very long time.

Off-Site Copies

Array-based replication solutions provide the ability to instantly failover to a replica in case the production data system is unavailable or its data unusable – this solves the issue of a loss of the primary production site.

But, the very nature of this replication process, while solving the site loss issue, introduces a new risk: the remote copies are typically not point-in-time copies but ones that are constantly updated.

This means that if corruption or an accidental or malicious update or deletion takes place on the source system, it is quite likely going to immediately propagate to the off-site target, making the copy of no value in a recovery effort.

The very nature of this replication process introduces a new risk.

Where Is Your Golden Copy?

Every organization should have an unstructured data protection strategy that includes not only local recovery and in-family replication recovery options, but also a mirrored golden copy. This golden copy should be located at a site outside the primary and DR data sites in its native format, providing an option in the event of a disaster.

Ideally the data would exist behind an air gap or in a bunker site (a break in the network that only allows selective access during a replication window) or in the cloud.

There are two distinct replication strategies to be considered, and potentially both implemented, depending on your requirements:

Replicating within the **same protocol** (NAS-to-NAS or S3-to-S3) enables the exact mirroring of the production system data – the filesystem structure, the shares and exports required for end-user and application access, and all permissions that control file access. This allows the replicated environment to be used as a failover target, in addition to being able to recover the data to any heterogeneous storage within the same protocol. In a NAS-to-NAS environment, versioning could be attained by taking snapshots.

Replicating between **different protocols** (NAS-to-S3) enables the creation of a golden copy on cost-effective storage – either on-premises or in the cloud – that can be restored to any NAS platform. Versioning allows data to be restored at a point in time before corruption or unwanted updates took place.

The Datadobi Solution

Creating a mirrored golden copy of your business-critical data should be a simple and easy task without complexity. However, as has been mentioned earlier, copying unstructured data in enterprise environments is not a simple task.

DobiProtect

DobiProtect® simplifies the data copy and recovery process and is built on Datadobi's tried-and-true architecture that is used in many of the world's largest and most complex environments.

DobiProtect allows customers to identify their core unstructured data, creating a mirrored golden copy to any heterogeneous NAS or S3 target.

DobiProtect copies all data, plus filesystem structure, shares, and exports required for end-user and application access, and all permissions that control file access.

Object storage protocols such as S3 support attaching a small amount of metadata to each object. To allow fast recovery and correct functioning of applications consuming the S3 storage, the object metadata is as important as the object data itself.

For NAS-to-S3 replication, DobiProtect keeps as many versions as required so that data can be restored before the point of corruption.

DobiProtect makes creating a mirrored golden copy of your core unstructured data quick, easy, and cost effective allowing you to prepare for the inevitable and recover from it fast.

The Datadobi Solution

PRODUCT BENEFITS	DobiProtect
Purpose-built to successfully carry out the data copy tasks with the greatest flexibility, including heterogeneous targets and scheduled replication.	✓
Ability to plan and design the data copy process within the same solution that will execute it.	✓
All aspects of data copy execution and recovery is managed from a single pane of glass.	✓
Provide real-time monitoring and easily consumed, on-demand and automated reports.	✓
Can perform discovery and analysis to provide a clear picture of the source and target systems' data.	✓
Natively store and retrieve/restore data from a cloud repository.	✓
Execute a **failover** to a target system automatically with full management and outcome tracking of the process.	✓
Execute a **failback** to a target system automatically with full management and outcome tracking of the process.	✓

When considering implementing the use of off-site data copies as part of your overall data protection and availability strategies, it is crucial to understand that while a secondary copy of data eliminates some risk, the need for a mirrored golden copy of unstructured data is vital to protect an organization's critical data and to maintain business continuity.

Implementing a golden copy should be simple, fast, cost effective, and above all reliable. However, due to the disparity between NAS and object platforms and the complexity of keeping that data synchronized, without proper tools and methodology, you will likely encounter numerous challenges.

Datadobi has created a purpose-built solution for the needs of unstructured data protection, file and object, in your data center and in the cloud.

Protect Your Unstructured Data Now

Contact Datadobi today to learn more about how we can help with your company's data copy and off-site protection efforts: sales@datadobi.com.

Datadobi

datadobi.com