



SearchStorage.co.UK

E-Guide

## Checklist and Tips to Choosing the Right Backup Strategy

Data deduplication is no longer just a cool technology, it's become a fairly common component of modern data backup strategies. Learn how to leverage data dedupe technology as part of your backup and disaster recovery strategy in this SearchStorage.co.UK eguide.

*Sponsored By:*

**EMC<sup>2</sup>**



SearchStorage.co.UK

E-Guide

# **Checklist and Tips to Choosing the Right Backup Strategy**

## **Table of Contents**

[Using data dedupe technology as part of your disaster recovery strategy](#)

[How to choose the best data replication product: A checklist](#)

[Resources from EMC Backup and Recovery Solutions](#)

## **Using data dedupe technology as part of your disaster recovery strategy**

By Pierre Dorion

With products from major storage vendors, it's clear data dedupe technology isn't just a niche technology. You may already have a dedupe appliance or dedupe software installed in your environment as part of your data backup strategy. But deduplication can also play an important role in your disaster recovery strategy.

### **How data dedupe technology works**

Because deduplication is based on the replacement of identical data segments with much smaller "stubs" that point back to a single, unique data segment common to all, it is easy to understand that deduplication ratios increase with history. What I mean by "history" is that the higher the number of similar files subjected to deduplication, the better the reduction ratio becomes because of the increased incidence of identical data segments. This is why deduplication is so popular for use on storage targets for backup and archives -- we typically store multiple copies of files for a specific amount of time, providing ample opportunities for the identification of duplicate data segments.

It's possible to implement deduplication on a primary storage device but the focus of this tip is on deduplication and disaster recovery. It is also important to point out that single-instance storage (SIS) is also considered by many to be another form of data deduplication. SIS discovers identical files or objects and store them only once with a reference to copies. For the purpose of this discussion, I will focus on variable-segment deduplication. I should also point out that deduplication-capable storage arrays are often transparently implemented either as a disk target or tape replacement in the form of a virtual tape library (VTL) when used in conjunction with backup software. There are some exceptions where the backup software itself has a deduplication capability, such as CommVault Simpana or IBM Corp. Tivoli Storage Manager, for example.

## Deduplication styles

Deduplication can be grouped under three main categories in terms of backup and disaster recovery:

1. Deduplication at the **source** where duplicate data segments are identified on the host before data is backed up. EMC Avamar and CommVault Simpana are two examples of product offering this capability.
2. Deduplication on the primary backup storage **target** (typically disk) where duplicate data segment are identified. This can be done using any of the many products, including CommVault Simpana, which can do both source and target deduplication.
3. Deduplication on a secondary or **copy** backup storage. This is common in environments where a deduplication-capable array is used as a replacement or complement to tape storage.

Different vendors might introduce variations in their approach such as in-band and out-of-band deduplication, or give their technology a slightly different spin to differentiate themselves but, essentially, the above can be considered the three main flavors. Some backup software also incorporates deduplication capabilities, but in the context of this discussion they can be grouped between source or target dedupe and, in some cases, both, like CommVault.

The differences between source-based and target deduplication are often the object of debate. Source-based reduces the amount of data sent across the network, so this is especially important for remote sites sending data back to a main site. However, target dedupe provides the opportunity to identify more identical data segments by comparing all data in a single location rather than at the individual host level. CommVault and EMC Data Domain, for example, are two products that can provide global deduplication, which means that data is deduplicated on an individual host at the source, or a smaller deduplication array can be replicated back to a single storage device and deduplicated globally.

## **Disaster recovery implications**

So far we have focused on backup, which obviously is an important part of a disaster recovery strategy, but without an offsite copy, backups alone cannot provide full protection in the event of a site-wide disaster causing the loss of both the production data and its backup copies. This is where data dedupe technology can improve the disaster recovery process via remote replication. Deduplication significantly reduces the amount of backup data stored by replacing identical data segments with stubs or pointers as mentioned earlier; this reduced data set can be replicated to a remote deduplication-capable storage array, and requires much less network bandwidth.

In addition, in the event of the total failure of the deduplication storage device at the primary site, deduplicated data at the remote site can be cross-replicated back to the primary site to rebuild the failed storage device. Because the deduplicated data and the pointers are kept together, this self-contained data set is unaffected and ready to use for restore operations on individual systems that used it as a backup target.

## **Data dedupe advantages in backup and disaster recovery**

The main advantages of deduplication in the context of backup and disaster recovery include:

- Reduced amount of disk required to store backup data
- Potential to increase retention for some backups
- Reduced network bandwidth requirement for remote replication (deduped data is replicated)
- Reduced network bandwidth requirement for backups when used at the source

## **Possible disadvantages**

While data dedupe offers many enhancements to a disaster recovery strategy, there are certain things to watch out for. Deduplicated data must still be restored in the event of a disaster. While deduplication leverages disk storage, it isn't the same as data mirroring or cloning, and even if deduplicated data can be replicated remotely, it doesn't become a

mirrored volume that can be mounted by a server to resume business processing. Data must be reconstructed and restored to a file system recognized by a server before it can be used.

In addition, if data was copied to tape from a deduplication-capable disk storage array using a standard backup software product, data was likely reconstructed in the process and lost all its deduplication attributes. This means files are back to their original size and number, and must be restored from tape media in the event of a disaster.

As you can see, data dedupe technology isn't just for data backup and primary storage; the benefits of using it as part of your disaster recovery strategy make it worth careful consideration.

## How to choose the best data replication product: A checklist

By Jacob Gsoedl

Data replication has been used for high availability and data protection of critical data and applications for a long time, and it has been slowly eating away at tape as the media of choice for data protection. As a consequence of shrinking recovery time objectives (and increased need for 24/7 availability of applications and data, this trend is likely to continue, if not accelerate. On the technology side, cloud computing, the virtualization of IT infrastructure, and a flurry of replication options and offerings are aiding this trend.

Choosing the best data replication product can be difficult. We've made it easier for you by providing a free, downloadable data replication product checklist. [Click here to download the data replication product checklist you see below.](#)

<b>KEY CRITERIA FOR SELECTING A REPLICATION PRODUCT</b>			
	<b>Type of replication</b>		
<b>Criteria</b>	<b>Array based</b>	<b>Host based</b>	<b>Network based</b>
<b>Storage infrastructure</b>	Requires homogeneous storage or storage virtualization	Storage agnostic	Storage agnostic; fabric-based products; require intelligent switches
<b>Server infrastructure</b>	Equally suitable for small or large number of servers and multi-OS environments	Very attractive for a small number of hosts; unfavorable for a large number of servers and multiple OS infrastructures	Equally suitable for small or large number of servers and multi-OS environments

**Checklist and Tips to Choosing the Right Backup Strategy**

<b>Management</b>	Very good, as replication is simply an extension of array management	Complexity increases with the number of hosts	One additional storage component to manage; complexity is independent of the number of hosts and arrays
<b>Scalability</b>	Good; it's directly linked to scalability of the arrays	Very good, as replication runs on hosts	More prone to becoming a performance bottleneck
<b>Robustness</b>	Very robust, as it's part of the storage array	Less robust, as problems on hosts and in applications can impact replication	Very robust because of dedicated appliances
<b>Cost</b>	Very good for existing arrays; unfavorable if new arrays need to be added	Cost increases proportionally with the number of hosts	Fixed cost for variable number of servers and arrays

## Resources from EMC Backup and Recovery Solutions



[IDC Workbook: Assess the Value of Deduplication for your Storage Consolidation Initiatives](#)

[ESG Report: Data De-duplication and Disk-to-Disk Backup Systems: Technical and Business Considerations](#)

[ESG Paper: What Data Domain is Doing to Storage](#)

## About EMC Backup and Recovery Solutions

EMC backup and recovery solutions are driving a fundamental shift in the way businesses protect and manage their information. With industry-leading deduplication storage systems and software, EMC has brought to market the next generation of data protection solutions. These cost-efficient disk- and network-based solutions are enabling organizations of all sizes to transform the way they protect and manage their information. For more information about EMC backup and recovery solutions, visit [www.EMC.com](http://www.EMC.com).