

Deduplication has been one of the hottest technologies in the storage industry for almost three years. During that time, it has generated marketing wars, industry consolidation, and comments and controversy from vendors. IT managers in most midrange data centers typically have limited staff and few backup specialists, and it can be hard to figure out how deduplication might fit into their situation. Following are important questions for IT managers to ask as they consider deploying deduplication in a midrange data center.

### 1. Is data deduplication now a mainstream technology?

**Yes.** Deduplication appliances have absolutely made the transition from experimental to mainstream. Analysts tell us that a little over 30% of IT departments use it for at least part of their data, and vendors now offer products with a couple of technology generations behind them that are optimized for simplified, non-disruptive deployment.

However, this doesn't mean that every solution is equal. Most deduplication vendors go through a learning curve, so it pays to ask about experience, references, and support when evaluating solutions.

### 2. What does deduplication really do?

Generally, deduplication is a method for finding redundant data at a sub-file level, and substituting a pointer for the repeated data. It can be used to reduce disk requirements as well as the bandwidth needed to transmit data.

There are several different and legitimate ways of doing that—block level deduplication is the most typical, but some products find differences between file-sets at a Byte level. Different approaches may have implications for performance, the amount of working space required, how easily they can support different software applications, and ease of setting up replication. The specific approach is less important than proven results and how well the approach matches with the problem you are trying to solve.

### 3. What problems are best addressed by deduplication?

The greatest leverage, and the most widespread adoption, involves backup data. That's natural since backups contain more redundancy than any other datasets and get retained longer. Most common types of office data—including email, data bases, and flat files—benefit from high deduplication rates.

Quantum recently surveyed users of its DXi-Series appliances to quantify results on the effects of deduplication when it is added to users' backup strategies. Compared to traditional storage systems, users reported an average increase in backup speeds of 125%, an 87% reduction in failed backups, and a huge change in restore profiles—restores that used to take several hours or days are typically reduced to minutes using deduplication. Costs are also reduced, often dramatically. Users reported that overall removable media costs dropped by an average of nearly half, tape recall costs were down by 97%, and the amount of time required to manage backups was reduced by 63%.

Users that adopted remote replication for disaster recovery (DR) protection saw an increase in recovery points, automating the process and eliminating tape (and tape management) in smaller offices.

### 4. Does it matter what backup software I use?

Most deduplication vendors have tested their systems with different backup applications and achieved effective results. Some vendors can even optimize data storage for more than one backup application. It is worth asking a deduplication supplier whether there are applications that they have optimized around.

Be sure to check for support for specific backup software interfaces. Symantec, for example, has developed an OpenStorage interface that works with backup appliances to provide an additional level of operational advantages—increased performance, better replication management, even direct, off-line tape creation. Ask deduplication appliance vendors about their strategic relationships with backup application suppliers. You will want to understand how closely they work together, and what their plans are for interoperability and integration in the future.

## 5. What is the easiest way to implement deduplication?

The choice facing most IT departments is between deploying deduplication appliances or carrying out deduplication within the backup software. There is no universal answer about which approach will be easiest to deploy. There are some guidelines, however.

With appliances, currently the most wide-spread approach to deduplication, the backup data is all sent to the device and deduplication occurs at the target. With appliances, users can add systems in place of, or along side of, existing backup targets and make very little change in the overall backup methodology. Because the deduplication is carried out on a purpose-built appliance, it never increases the load on backup clients or media servers, and it makes the deployment of operations like replication straightforward. As the most common method, it is also the most mature—which usually means faster deployment and fewer service needs.

With a software approach, the backup application adds deduplication to the other tasks that it carries out, either on backup clients or on media servers. By deduplicating data before it is sent to a target, less data has to be transmitted over the network—the idea is similar to performing compression in software, and in fact deduplication processes almost always include compression as well. Since deduplication is a relatively high overhead operation, there's a chance that backup operations may slow down so deployment may require adding new servers or dedicated storage. This tends to increase the cost and complexity of integration.

Either approach can make sense depending on specific circumstances. To decide what is best for you, think about where bottlenecks are in your system today, whether or not your current media servers are underutilized, and what level of integration effort makes sense for your specific situation.

## 6. Should I eliminate my tape storage altogether?

Although most end users who adopt deduplication reduce their use of removable media, very few eliminate it entirely—and for good reason. Typically, users have roughly three tiers of needs for backup: daily backup and restore, near-term DR protection, and long-term data retention. It makes sense to look at different technologies for each tier and to talk to vendors who understand them.

**Daily backup and restore:** Many users find that disk read and write profiles give them advantages for day-to-day backup and restore. Deduplication adds the advantage of letting them store data on disk longer so that more restores can take advantage of those profiles.

**Near-term DR:** Replication enabled by deduplication lets users with multiple sites replace removable media with remote replication for DR. As a result, they see more restore points, reduce costs, and automate what is for many a very manual operation.

**Long-term retention:** Removable media continues to provide strong economic and security value. Tape consumes the least power, space, and cooling of any storage, making it the preferred medium for long-term retention. New technologies for tape, including encryption and media integrity analysis, have made it more secure and reliable.

## 7. Where can I get objective advice?

There are lots of ways to get objective advice about which approaches match best to your specific needs. Some independent analysts who spend time talking directly with end users provide very useful and objective information about others' experience. But if you aren't a client of the big-name analysts, there are other options.

One of the best is an experienced reseller. Good resellers, who have a track record of helping IT departments deploy technology, understand the reality of what will work for specific environments and they have a vested interest in helping you succeed.

You can also talk directly to vendors. If they offer multiple technologies, they are likely to provide a broader view than if they offer only one product. And if you have a vendor that you already trust for backup, it makes sense to see what kind of deduplication options they have.