

STORAGE SWITZERLAND

THE STATE OF DEDUPE 2012



George Crump, Senior Analyst

Deduplication has transformed itself as a technology. What was once thought of as only valuable in the backup-to-disk market has now moved into almost every type of storage and networking device. Eventually, anything that handles or stores data will likely implement some form of deduplication.

Over the last two years the big emphasis from deduplication suppliers like [Permabit Technology](#) has been to move deduplication out of the backup and archive market segments and into primary storage. What these vendors found were the obvious benefits of implementing deduplication in storage area networks (SAN) and network attached storage (NAS) systems. But they also found new markets that could benefit more directly from deduplication technology.

Dedupe in 2012 - beyond SAN and NAS

While traditional hard drive based SAN and NAS storage systems have been slowly adopting deduplication and other optimization technologies, 2012 may mark the year

that other storage vendors pass these legacy vendors in the adoption process. The most aggressive push is likely to be in enterprise SAN and NAS systems based on solid-state, flash technology. Here the payoff for a deduplication investment has the same potential that it did in the disk backup market.

One of the key drivers for deduplication in that market was the cost differential between disk and its competitor at that time: tape. In the early 2000s there was at least a 5 to 10x price differential between even the most economical disk systems and a comparable tape library. While disk was attractive to all because of its ability to decrease backup windows and increase recovery speeds, the price differential made its application as more of a staging area and/or cache, rather than a final storage target.

Compare this to solid-state storage. Today, because of its dramatic price differential compared to high-speed mechanical disk drives (also 5 to 10x), flash storage systems are typically relegated to a small staging area or tier, or used as a cache. The similarities are fascinating.

Deduplication, when applied to enterprise flash, has the potential to level the price playing field with enterprise disk. The result is that enterprise flash systems come well within the price points of high-speed disk-based storage systems when deduplication is included. 2012 looks to be the year that we see massive adoption of deduplication technology by flash vendors. While the deduplication can be applied to high speed 15K RPM disk, the performance of solid-state in these configurations is still unmatched.

The adoption of deduplication in solid-state devices will not be limited to just flash systems vendors though. In 2012, look for flash controller chip vendors to also adopt deduplication in a large way. These vendors already have intimate knowledge of the data traveling through their controllers to be stored on NAND flash and today they use that knowledge to optimize writes and increase the flash device's life expectancy. Adding a deduplication algorithm to their controller logic is certainly not out of the realm of possibility. Most flash controller vendors that Storage Switzerland speaks to openly admit interest in implementing deduplication in their controllers. The state of dedupe in 2012 though will not be limited to memory-based storage, but will go well beyond those devices.

Another area in which we expect adoption will be in RAID controllers, which, similar to flash controllers, have intimate knowledge of the data they handle and send to their attached mechanical storage devices. Imagine buying a controller with a deduplication option and being able to turn any server or workstation into a high-performance deduplicated SAN, NAS or even backup target.

Dedupe Reboot

Another area we expect significant activity is with vendors that have implemented a revision 1 of deduplication technology and are realizing that implementation is not going to scale to meet customer demands long-term. Many vendors both large and small attempted to implement

deduplication via an internal effort, often leveraging open-source code.

In 2012 as these vendors, especially startups, begin to grow their customer bases and their implementation of deduplication encounters real workloads, scaling issues and the need to provide better dedupe performance will drive them to look at alternative methods for implementing deduplication. Vendors like Permabit, with their Albireo Software Development Kit (SDK) and turn-key Virtual Data Optimizer (VDO) module, equip these vendors to quickly position themselves to capture this "revision 2" market.

Who is embracing Deduplication

Another interesting development that we will see coming in 2012 is that this evolution in deduplication will come primarily from smaller vendors and startups. And, these vendors will likely go outside to source that deduplication technology. Legacy vendors will try to "roll their own" deduplication implementation. And some eventually may very well be successful, although many will need several iterations.

Smaller vendors or startups on the other hand, are used to outsourcing everything or at least as much as possible. This gives them the right attitude when looking at implementing a technology like deduplication. For them, outsourcing an API set (via the SDK) to get deduplication will not be such of a strategic stretch for them.

Also, the case could be made that these vendors will have an easier time implementing and integrating a deduplication API set because, at this point in their existence, their code base is not overly complicated with added features. There is also the chance that the core engineering team that wrote the original code for the appropriate product is still with the company. Their intimacy with that code should allow for easier and more rapid adoption of an external API set. Integration with the SDK can be done in as little as one week versus the thousands of man-hours to build internally.

For vendors of Linux-based systems that do not want to integrate with an API set, the Permabit VDO solution offers a turn-key solution that can be incorporated directly into the Linux file system in less than a day. Albireo VDO provides data deduplication capabilities while enabling OEMs to continue leveraging all of their storage solutions' existing features, including existing file systems, virtualization features, and data protection capabilities.

VDO supports all major Linux versions used by the storage industry and delivers everything needed for deploying enterprise data optimization capabilities, including binaries, open source code modifications, management scripts, and full documentation. With Albireo's performance, OEMs can extend their deduplication capabilities and out-compete even the largest of storage players by providing data optimization capabilities for mission-critical application storage while effectively leveraging Linux open source to maximize value.

Summary

Over the last two years, deduplication was something that we recommended vendors in the primary storage space seriously begin evaluating, in order to understand how to implement it into their current technology. At that time we recommended the API set approach, as it seemed to be the shortest path to delivering functional products. Last year we strongly advocated that flash storage vendors begin implementing the technology given the potential for return on investment. VDO provides a faster turn-key approach for those vendors.

In 2012, deduplication is here. Storage systems that can't offer deduplication won't just be at a serious disadvantage, they won't even be considered. There are the obvious cost advantages to a deduplication-enabled storage system, as well as the practical advantages of better leveraging a current resource with no performance impact. For companies that are not within a few months of rolling out a viable deduplication strategy the time to embrace an API set is now.

About Storage Switzerland

Storage Switzerland is an analyst firm focused on the virtualization and storage marketplaces. For more information please visit our web site: <http://www.storage-switzerland.com>

Copyright © 2011 Storage Switzerland, Inc. - All rights reserved