

---

## Data Domain Increases De-Dupe Performance

**Date:** May, 2007

**Author:** Heidi Biggar

**Abstract:** Data Domain may be the hands-down leader in de-duplicating disk-based backup systems, but it hasn't been able to shake questions about the scalability of its systems beyond the midrange. With the release of its new dual-core system—the DD580/DD580g—Data Domain looks to silence skeptics once and for all.

### The Question of Scalability

Data de-duplication has quickly become one of this decade's most important—and hence most-talked-about—new technologies. It alone has the potential to revolutionize data protection (and primary storage) from both a technology, and an end-user adoption standpoint—by simply making disk-based backup, recovery and remote replication much more efficient and cost-effective than it is today.

What's game-changing about data de-duplication is its ability to alter the economics of disk-based protection. It not only allows users to keep more data online longer (i.e., extend retention periods), but it can also help them improve or expand their remote backup and DR efforts as well as reduce power, cooling and footprint costs in the data center by reducing back-end capacity requirements by 10x to 20x, or more.

While there is no doubt about the benefits of de-duplicating disk-based data protection solutions from vendors such as Data Domain and others (why wouldn't you reduce back-end capacity requirements if you could?), there are still a number of questions users should ask vendors when evaluating products to ensure that the products they choose are the best fit for their environments today and in the future. Such questions center on performance, capacity, scalability and ease of use.

In the case of Data Domain, the scalability of its systems, specifically in terms of performance, has continually been brought into question—rightly or wrongly—by the media, competitive vendors and well-read users. While Data Domain has seen great traction in midrange markets, where performance demands are less, the enterprise space has been comparatively elusive for the company to date. With the release of the DD580 dual-core system, Data Domain raises the bar on de-dupe performance, and makes its systems an even better option for enterprise users.

### The Power of the DD580/DD580g

While ESG Research shows that larger organizations are less likely to replace their existing tape-based backup environments with de-duping disk-based systems than smaller organizations (whose tape investments are significantly smaller), there is still significant opportunity at the high-end to help users augment their existing environments with tape going forward. For Data Domain, its new DD580/DD580g systems are the means to an enterprise end.

The DD580 is an Intel-based dual-core system, and represents a 2x increase in performance (400 GB/hour to 800GB/hour) and about a 30% increase in capacity over the DD560. Like previous models, the DD580 can be used in conjunction with internal disk storage (DD580) or as a gateway to third-party storage (DD580g). Data Domain claims aggregate backup and recovery throughput to 800 GB/hour with the DD580. A fully configured DDX array (with 16 DD580 controllers) has a physical capacity of 20PB and a throughput to 12.5 TB/hour. ESG Lab has reviewed the DD560-based DDX arrays and has validated the test process Data Domain uses for reaching its maximum throughput level (400 GB/hour). Data Domain has used the same test process to confirm the 800 GB/hour throughput of the DD580.

While the performance and capacity boosts are key to the scalability of Data Domain's DDX line, what's really worth noting is how this scalability is achieved. Unlike competitive vendors who achieve higher performance rates by adding disk drives (i.e., spindles), Data Domain scales performance by adding CPU power. Data Domain's underlying architecture is CPU-centric versus disk-bound.

Why is this important? CPU performance is increasing at a significantly faster rate than disk drive performance, and has since the 1980s. This means that to reach the same level of performance, vendors of non-CPU-centric systems have to add a lot of SATA disk. Doing so has the potential to:

- Add significant cost to the disk-based backup system. Not only does this potentially mean a lot more disk drives to buy but there is also the issue of housing, powering and cooling these drives. In today's economic climate, this is a key consideration.
- Increase the complexity of the storage environment. Simply put: More drives mean more drives to manage and maintain. Of course, just how easy it is to manage these drives depends on the system architecture. Some disk-based products are inherently more "modular" and hence more "scalable" than others.
- Defeat some of the gains made by de-duplication. One of the immediate benefits of de-duplication is a reduction in back-end capacity requirements by 10x to 20x or more, depending on the data type and the de-duplication process. The irony is that, depending on the system architecture, users may have to actually buy more disk (before they need it) to see the performance improvements they're looking for, thus potentially eroding some of the core benefits of de-dupe.

## The Bottom Line

Data Domain has been trying to move upstream into the enterprise space for a year or so but has seen most of its successes to date in the midrange market. The DD580 may be the technology to get it there. Speeds and feeds aside, Data Domain's CPU-centric story is likely to resonate with large end-users who, among other things, are grappling with today's data demands and energy realities. Beyond that, Data Domain, with more than 750 customers under its belt, is in the enviable positioning of being viewed as the de facto de-dupe leader and is on a high in light of its recent IPO filing. These factors combine to give Data Domain a very decent chance of gaining some real traction in the enterprise.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. and is intended only for use by Subscribers or by persons who have purchased it directly from ESG. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482-0188.