

The Importance of Operational Simplicity in Selecting Disk-to-Disk Data Protection Solutions

May 2007



Without question, Disk-to-Disk (D2D) backup has emerged in recent years to become a critical storage component of an enterprises' data protection environment. With enterprises experiencing accelerating data growth rates, they typically will use tape media for long-term archives, but use disk for online and nearline backup and restores. Sensing a new market opportunity, a multitude of vendors has responded to the lure of this enterprise market, resulting in a wide range of alternatives when it comes to D2D data protection. However, whether or not those alternatives are well suited to solve an end users' specific D2D challenges can be difficult to ascertain.

D2D purchasing decision-making has evolved very rapidly from "point purchase" to "strategic initiative." Enterprises are looking at this increasingly established technology category and demanding that it go beyond mere "speeds and feeds." In fact, we believe that enterprises are holding vendors to a higher standard – no longer can a solution just be cost effective as measure by \$/TBs or be high performance as measured by MB/s of throughput. Specifically, D2D solutions must be *simple and cost effective to deploy and manage*. It is ironic that the one strategic criterion that many vendors have ignored seems the least technical of all: simplicity of operation. Understanding and measuring an offering's operational simplicity is crucial criteria for selecting a D2D solution.

Accordingly, Taneja Group has identified a locus of buying criteria that we believe compose the overall simplicity of operation of a D2D solution. Solutions that meet this definition of operational simplicity can be deployed in complex enterprise backup environments, allowing businesses to sharply reduce management overhead, attain very strong ROI, and fully protect data integrity.

As we will explore in this brief, we believe that Data Domain has delivered an offering that personifies this notion of operational simplicity. We think that the positive market momentum that we continue to see from Data Domain is directly correlated to the company's focus on simplifying the deployment and ongoing management of its devices within D2D environment.

Operational Simplicity Defined

While a range of vendors have provided powerful D2D solutions, few have thought about the critical role simplicity of operation must play in a D2D environment. We view this as a major oversight in the industry to date. Based on our conversations with end users, we put forth a set of criteria by which end users can evaluate potential D2D offerings on not only their technological merits, but also on their ease of deployment, usage and scalability merits.

From our research, we define simplicity of operations to have three core dimensions - *plug and play* capability, cost-effective, transparent scalability, and non-disruptive, reliable *data protection*. It is the presence of all three core attributes in the same D2D solution that creates a powerful solution that meets our definition of operational simplicity. This cost-effective and comprehensive approach to the D2D backup environment delivers dramatic ROI benefits to the enterprise.

Simplicity Criterion #1: Plug and Play Capability

Plug and Play capability refers to a D2D solutions ability to transparently and non-disruptively fit with existing data protection infrastructure. Traditionally, D2D solutions are architected in two

dramatically different models: 1) a disk-based target store that works with existing backup applications or 2) a backup software-based model. The target store model easily slots into the existing backup infrastructure, while the software-based alternative can be disruptive to the existing backup environment, requiring large-scale changes – and even wholesale replacement -- of existing backup applications.

End user have found that replacing or heavily modifying the backup environment results in the loss of valuable features, requires new RTO evaluations, and raises serious concerns about using tape and supporting older backups. These concerns lead Taneja Group to conclude that the target store implementation is superior to the alternative software-based D2D approach with respect to plug and play deployment. Key features of the target store model include:

1. *Target stores interoperate with any leading backup software and more.* Since they work with multiple backup applications, target stores support most users' environments. Software-based D2D is limited to supporting one or two backup applications. D2D systems that only support a VTL interface support only backup software and are to some extent limited in flexibility. Finally,

TECHNOLOGY BRIEF

systems which approach capacity optimization by requiring application-dependent parsing approaches will always be behind on which applications they can support. It is optimal to be application independent.

2. *Target stores are packaged as an easy-to-deploy appliance.* These deceptively simple systems are in fact purpose-built for D2D optimization. Software-only alternatives present driver and tuning issues, only support single systems and fabrics, and are challenging to deploy and manage.
3. *Target stores are fabric agnostic.* The target store model should be capable of supporting multiple fabrics, such as Ethernet and Fibre Channel. Since many users utilize heterogeneous fabrics for storage connectivity, the target store can then be deployed easily within any existing fabric.

Simplicity Criterion #2: Cost-Effective, Transparent Scalability

Cost-effective, transparent scalability refers to how D2D solutions scale with little or no administrative overhead to meet ever increasing data storage requirements and shrinking backup windows. Our research shows that end users are not necessarily looking for a lower cost solution than what they are

already spending on tape, but rather D2D solutions should have a comparable cost to tape automation systems, while offering much higher capacity and performance. Performance should support shrinking backup windows in the presence of growing data requirements and should also efficiently support day-to-day data operations including fast restores. Inline de-duplication, per-stream speed, and fewer spindles characterize cost-effective, transparent scalability factors.

1. *Inline de-duplication* efficiently stores de-duplication data on the target store and factors backup data as it streams into the appliance. Inline placement dramatically cuts disk capacity and retains timely access to data for DR purposes. In contrast, post processing de-duplication technologies store all backup data first, regardless of redundancies. The post processing solution then runs de-duplication and stores the de-duplication information separately, requiring far more capacity and delaying access for DR purposes.
2. *Per-stream speed* is critical to successfully introducing D2D solutions into the backup environment. De-duplication technologies often slow per-stream speed. D2D solutions that deliver both capacity management and high

TECHNOLOGY BRIEF

performance for both backup and recovery meet the cost-effective, transparent scalability requirement. End users often overlook recovery throughput when evaluating data de-duplication solutions. We believe that users should consider both sides of the coin - backup and recovery throughput – as critical metrics in assessing performance.

3. *Tuning for fewer spindles* results in fewer disk interactions and minimizes disk IOs. Many D2D solutions require hundreds of disk spindles to deliver high levels of aggregate throughput, resulting in large disk farms, long seek operations, and the use of more expensive FC disk instead of SATA disk. This entire process soundly defeats the de-duplication advantage of using fewer disks.

Simplicity Criterion #3: Non-Disruptive, Reliable Data Protection

Data integrity and safety are critical to a successful D2D implementation, and are the third critical factor in our definition of what constitutes operational simplicity. This type of protection cannot be an afterthought. Data integrity and reliability must be designed in from the ground up to accommodate the highest levels of data protection. Two core areas that end

users should evaluate vendors against are:

1. *Data Fault Tolerance – more important than HA.* End users should determine whether the D2D solution preserves data protection, integrity and recoverability and is capable of fault detection and healing. For example, RAID 6 support is a crucial data protection feature that allows the D2D system to survive two disk failures. Since backups often fail and restart with tapes, the bar on High Availability is low, but the premium on recoverability is high.
2. *Flexible, timely replication* includes performing data de-duplication inline as the data is backed up, observing backup and disaster recovery policies and preferences, and being auditable. Replication should work at both the directory levels and across sites. Lacking these options, a D2D solution is dependent on backup array replication or off site tape storage. It retains no replication state information of its own, so it must remount an entire volume to restore a file.
3. *Time matters.* When de-duplication is performed as a post-process, there can be a long lag time between completing the backup, de-duplication of the backup data, and

TECHNOLOGY BRIEF

then replicating that data offsite. As a result, the restore point on the DR site can become obsolete. High performance inline approaches do not suffer this problem.

Comparing Data Domain against Our Definition of Operational Simplicity

As we track the vendor landscape in D2D data protection, it is clear that Data Domain has demonstrated breakaway customer momentum in the past 12 months. Our hypothesis is that Data Domain has enjoyed this market success not only due to its core data de-duplication technology, but also because of the way that it has packaged its solution to provide an easy to deploy and operate appliance that meets our definition of operational simplicity

The Data Domain arrays, appliances and gateways are purpose-built systems that offer plug-and-play capabilities, cost-effective, transparent scalability, and non-disruptive, reliable protection. These factors taken in aggregate create a truly simple to deploy and manage device.

Data Domain: Plug and Play Simplicity in Enterprise Environments

Data Domain is not limited to supporting one or two backup applications, but offers rich support and

interoperability with all types of backup applications. Its appliance architecture is purpose-built for D2D optimization. It comes pre-tuned and pre-configured for this environment. The Data Domain appliance can be installed and up and running in a matter of minutes and ongoing management is straightforward and requires minimal additional training. Data Domain supports both Gigabit Ethernet and Fibre Channel, making it a highly flexible and global choice in the enterprise.

Data Domain: Cost-Effective, Transparent Scalability

Data Domain's inline de-duplication technology dramatically reduces the storage required for backup data by pooling redundancies within backup images and storing only unique data patterns.

The Data Domain appliance supports inline de-duplication (de-duplicating before storing to disk), not a post-processing capability. Inline architectures should minimize hardware requirements. As a result, Data Domain requires two to eight times less hardware than other D2D alternatives. Most D2D alternatives require more capacity because they first store all data sets from the backup and only then run de-duplication on the data, or else they seek to disk so often during their de-duplication verification process that small numbers of disks cannot handle

TECHNOLOGY BRIEF

the load. Data Domain's inline de-duplication works during the backup, meaning that its first full backup is only one-third the size of the original. Subsequent backups, whether full or incremental, are only 1% the size of the original backup.

Inline de-duplication can also speed up disaster recovery (DR) access to data by replicating during the de-duplication process instead of waiting until after the backup job is complete.

Data Domain can store several months of recovery copies in an extremely small number of disk drives, lowering the price per GB and greatly simplifying administration. In fact, the same amount of capacity that other D2D solutions require for one day's worth of retention, would take Data Domain an entire two months to fill.

Data Domain's de-duplication technology does not reduce its per-stream speed. The Data Domain appliance can meet the needs of database and system backups and restores, and also speed up disk-to-tape (D2T) copies for long-term archiving.

Data Domain also optimizes disk interaction. Optimized processes take place in RAM, resulting in higher transfer speeds with fewer disk IOs. Data Domain's per-stream speed tops 200 MB/sec, which is five to ten times

faster than most de-duplication alternatives.

Data Domain: Non-Disruptive, Reliable Data Protection

Data Domain's purpose-built data invulnerability architecture ensures data integrity and guarantees data safety. Making stored data invulnerable requires four critical areas of focus: end-to-end verification, fault avoidance and containment, continuous fault detection and healing, and file system recoverability. Data Domain provides all four.

For example, Data Domain protection options allow the appliance to warn storage administrators about a backup error in time to fix it, and also keeps metadata reference information to enable active, ongoing file system consistency checking. A recent, separately stored snapshot copy of the file system is internally maintained so that there is a reliable file system rollback point at all times (available through Data Domain support personnel if required).

Data Domain also practices flexible replication including bi-directional, many-to-one, and selective options. These replication processes allow Data Domain to maintain fine-tuned control over replicated data, enabling it to perform granular restores.

T E C H N O L O G Y B R I E F

Taneja Group Opinion

The D2D product category is rapidly evolving. End users are no longer purchasing D2D data protection solutions solely based on cost per capacity metrics (\$ per TB) or raw performance (\$/MB of throughput). In short, the purchasing decision making process has matured to the point where D2D solutions are being evaluated on a new set of operational simplicity criteria. From our perspective, this evolution in purchasing behavior represents a natural maturation process of the entire D2D product category.

A simple D2D product that does not provide an easy to deploy, easy to manage experience is no longer adequate for the complex enterprise backup environments of today. Operational simplicity attributes such as plug and play capability, cost-effective, transparent scalability and non-disruptive, reliable data protection are becoming top of mind buying criteria for end users.

We believe that Data Domain's focus on operational simplicity has had a direct

correlation with its business performance and market success over the past couple years. In a Taneja Group report entitled "Next Generation Data Protection Market Forecast 2006-2010", Taneja Group recently calculated the market for Capacity Optimized Storage (e.g. products that utilize data de-duplication) was \$87M in 2006. Based on Data Domain's published financials, Data Domain enjoyed over 50% market share in 2006. We believe that its maniacal focus on simplifying the end user experience has allowed it to make significant inroads to date.

Data Domain stands out in its ability to deliver a solution that not only dramatically shrinks required capacity of a secondary storage environment, but also personifies this concept of operational simplicity: an easy-to-use package that offers significant results quickly across a wide variety of use cases. We believe this focus explains Data Domain's strong market traction over the past 12 months, and this same focus should serve it well into the future.

***NOTICE:** The information and product recommendations made by the TANEJA GROUP are based upon public information and sources and may also include personal opinions both of the TANEJA GROUP and others, all of which we believe to be accurate and reliable. However, as market conditions change and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. The TANEJA GROUP, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors which may appear in this document.*