



Research Perspective

DataDirect Networks: A Storage Platform for the Extreme Demands of Today, and Tomorrow

Abstract:

Enterprises are generating massive amounts of unstructured data these days. Storing all that data is putting a strain on IT budgets and staffs, and traditional storage systems are buckling under the load, often failing to provide acceptable levels of performance or availability. A different kind of storage system is badly needed. DataDirect Networks (DDN) makes storage appliances that scale easily to multiple petabytes of capacity, provide the highest available I/O performance, and deliver consistent service levels. Moreover, DDN's S2A storage appliances were specifically designed to work well with large content files and other types of unstructured data.

This DMG Report describes today's changing data environment and explains why storage administrators and architects need to familiarize themselves with DataDirect Networks' S2A storage platform.

A different kind of data environment

Both the *amount* and the *type* of data that organizations are creating has changed dramatically in the past few years. Two trends stand out:

- 1. An explosion of unstructured data.** In the past, most electronic data consisted of *structured* data, i.e., records in a database. In most organizations today, however, the vast majority of data is *unstructured*—office documents, email, images, audio, video, etc.—and the volume of

Copyright © 2002-2008 Data Mobility Group, LLC. All Rights Reserved. Reproduction of this publication without prior written permission is forbidden. Data Mobility Group believes the statements contained herein are based on accurate and reliable information. However, because information is provided to Data Mobility Group from various sources, we cannot warrant that this publication is complete and error-free. Data Mobility Group disclaims all implied warranties, including warranties of merchantability or fitness for a particular purpose. Data Mobility Group shall have no liability for any direct, incidental, special, or consequential damages or lost profits. The opinions expressed herein are subject to change without notice.



such unstructured data has been growing at a truly astounding rate. This is not a trend that is likely to soon reverse; indeed, rather than slowing down, the pace of data growth appears to be still accelerating. The second trend is a contributing factor.

2. Web 2.0 and user-generated content. The term “Web 2.0” (for the purposes of this discussion) refers to web sites where users create or submit content to be stored and possibly shared with other users. There are of course businesses that exist solely to provide Web 2.0 web sites, but the phenomenon is more widespread than that. Many other businesses and organizations are implementing web services that invite users—employees, partners, and customers—to create and share content. User-generated content ranges in size from the very small (e.g., text-based messages, blogs, comments, or reviews), to the large (digital photographs), to the very large (high-definition video). Users expect content to be stored safely and served up fast, the faster the better.

The sheer amount of unstructured content being created poses a serious challenge. Organizations have been investing heavily in storage, but the volume of data is threatening to overwhelm both staff and budgets. Backup windows are inadequate; storage devices are proliferating; management headaches are more painful and more numerous. Another challenge is that traditional storage arrays simply weren’t designed to serve up large content files (e.g., streaming video) and, not surprisingly, they don’t handle it very well.

Organizations need a storage system that can scale to petabytes but won’t break the bank, won’t become a nightmare to manage (due to multiple arrays with silos of information, constant data migrations, forklift upgrades, tedious and error-prone storage management chores, etc.), and won’t exhaust data centers’ limited power and floor space. Organizations also need that storage system to work well with unstructured content, delivering the quality of service that users and applications demand.

Storage vendors have been scrambling to develop systems to meet these needs, but one vendor—DataDirect Networks—got a significant head start on the competition and has already developed a storage platform uniquely well-suited to this new data environment.



A different kind of storage system

DataDirect Networks' S2A storage appliances were designed for extreme environments—environments that absolutely require ultra-fast I/O throughput, ultra-reliable service levels, and ultra-high capacity. The S2A is short for “Silicon Storage Architecture,” and, as the word silicon suggests, the S2A appliances employ custom-programmed FPGAs and embedded processors to facilitate high-bandwidth I/O and on-the-fly error correction. S2A has a highly parallel architecture, providing simultaneous access, both read and write, to multiple disk drives from multiple hosts.

Key features and benefits of the S2A platform

- **Extreme storage density.** The S2A9900 packs 1.2 petabytes into only two racks. Storage administrators have fewer devices to manage and data centers can conserve floor space and energy.
- **Extremely high throughput.** An S2A couplet (two appliances together, the standard S2A configuration) can sustain both writes and reads, *simultaneously*, at a rate approaching 6 gigabytes per second. Multiple S2A appliances operating together in a cluster formation can reach mind-boggling levels of aggregate throughput; some of DDN's supercomputing clusters sustain more than 200 GB/s.
- **Writes are as fast as reads.** An S2A appliance can ingest data at the same speed that it spits it out.
- **Offers RAID 6 protection without degraded mode.**
- **Can use SAS, SATA and SSD drives in the same enclosure.** Provides flexibility to implement a form of tiered storage inside the appliance, or to maximize either performance or capacity.
- **Corrects the silent data corruption that plagues SATA drives.** S2A's real-time parity checks on both reads and writes allow it to catch and correct silent corruption errors.
- **Choice of interconnects:** Fibre Channel, InfiniBand, or Ethernet.
- **Choice of file systems.** There are S2A installations using GPFS, Lustre, IBRIX, PolyServe, StorNext, NFS, NTFS, GFS (Red Hat), Microsoft Cluster Server, etc.



S2A uses hardware-based parity checks on every read and write, so it can detect and correct corrupted data before sending it to a host. S2A also uses parity data to fill in missing blocks of data (if for instance one disk in a RAID group is lagging behind due to some routine error condition), so hosts receive a consistently high quality of service. Because the parity calculations are hardware-based, and because this is the normal mode of operation for an S2A appliance, the error corrections and missing data replacements happen in real-time and do not degrade performance.

DDN has an interesting perspective on quality of service. For DDN, QoS is not just about counting the 9's of availability; QoS also includes consistent, predictable I/O performance. This emphasis on guaranteed delivery (i.e., providing consistently high throughput) translates into no hiccups when streaming video, no doubts that data can be fully ingested into an S2A system in a given amount of time, no frustrated users pounding their fists on their desks, etc.

DDN has devoted a substantial amount of R&D effort to making inexpensive SATA drives function as reliably as expensive enterprise-class drives, and whether an S2A appliance is using SATA drives or enterprise drives, or both, S2A employs a battery of techniques to eliminate as much as possible the necessity for human intervention (e.g., to replace failed drives). The technical details—fixing the silent data corruption problem, drive journaling, diagnostics and error recovery without performance degradation, power cycling, etc.—are fascinating, but we won't go into them in depth here. Suffice it to say that S2A appliances have been painstakingly engineered to be self-healing in most situations, sparing storage administrators' time for more important matters.

A proven solution

DDN has been in business for more than a decade, has long been profitable, and has annual revenue exceeding \$100 million. Until recently, DDN has focused most of its sales efforts on vertical markets and selected customers that have the greatest need for the kind of high-end performance and massive capacity that is DDN's specialty. That strategy has led to considerable success:

- DDN dominates the high-performance computing market, providing the storage for more than 40 of the world's top 100 supercomputing facilities; BlueGene/L, one of the fastest supercomputers in the world, is powered by DDN storage.



- DDN has secured hundreds of customers in broadcasting, postproduction, animation, and related fields, by meeting those customers' stringent demands for consistent, high-bandwidth streaming and ingest of large media files.
- Savvy Web 2.0 companies have found a competitive edge with DDN. For instance, Shutterfly uses DDN to store more than two billion photographs, totaling more than 13 petabytes (and growing).
- Microsoft relies on DDN for the storage underlying Xbox Live, serving music and video files to more than 10 million users.

DDN has established a reputation for being able to meet the most extreme requirements for throughput, capacity, and quality of service. Given the proven ability of the S2A platform to satisfy the demands of supercomputing and other extreme environments, it should be obvious that S2A appliances can handle the lesser demands of general purpose storage. Accordingly, DDN has begun expanding its focus to encompass more mundane (and far more numerous) types of customers.

S2A as a general purpose storage platform

S2A appliances are already being used for a wide variety of “mundane” purposes. Two mainstream storage applications that leverage the strengths of the S2A platform in a noteworthy way are archival storage and backup.

1. Archival storage. Compliance and eDiscovery concerns are driving organizations to keep all or nearly all of their data online, on disk, where it can be immediately accessible. Aside from compliance and eDiscovery, there are also usually business or operational reasons for keeping some content in online archives. Immense amounts of online archived data call for a platform like S2A, with its ability to scale storage capacity to multiple petabytes and its space-saving, energy-saving storage density. In some cases, when archived content is requested, it needs to start streaming instantaneously. In other cases, a short delay before the beginning of streaming would be acceptable. For the latter scenario, S2A has a Dynamic MAID (D-MAID) option that spins down inactive disks, saving on energy costs; when content on a disk in sleep mode is requested, the disk is spun up and data starts to flow after a delay of approximately 12 seconds.



2. Backup and recovery. S2A appliances make a superb backup destination. The S2A platform is highly optimized for ingesting streams of sequential data. S2A's write speeds are fast—*really* fast—and, just as important, they are consistent. Making sure a backup finishes in the allotted window is much easier when the destination's write speed is both fast and entirely predictable. Recoveries also benefit from S2A's emphasis on a consistent quality of service. In addition, S2A is optimized to handle several write and read streams in parallel, so multiple backup jobs and recoveries can be running simultaneously. DDN sells S2A in a VTL configuration and has customers using S2A appliances for disk-to-disk backup as well as disk-to-disk-to-tape. This should continue to be a major growth area for DDN, because the S2A platform has features that make it especially useful in backup environments.

Investing in future-proof storage

Organizations should try to anticipate the new and greater demands that will be made on their storage systems. Financial prudence dictates investing in storage platforms that are guaranteed to be able to grow to meet new requirements. Many an organization has purchased a SAN or NAS device with the idea that it would provide plenty of storage space for everyone, only to discover, usually very quickly, that they would need to buy a second device, and then a third, and a fourth, etc.—this tends to irritate the CFO, and also tends to result in overworked storage administrators feeling a bit like sorcerer's apprentices.

As an example of the new and greater demands that should be anticipated, consider the growing ubiquity of video. The next few years will see a sharp rise in video chat, video conferencing, video surveillance, video instructional material (not just computer-based training for employees and students—customers will also expect instruction manuals to be in video format), video marketing, video podcasts, video simulations, video medical imaging, and so on. Anyone who has called for customer service is familiar with the notification that says “this call may be recorded for quality assurance”; in the future, expect to see “this interaction may be recorded on video for quality assurance” and “your every movement at work is being recorded on video for performance optimization purposes.” Essentially, video will be everywhere. Storage systems that are unable to stream video, both in and out, with a high quality of service, or that lack sufficient capacity to store large video files, will become a source of pain.



Regardless of the role that video ends up playing in a given organization, the volume of unstructured content is sure to continue growing. What would have been considered an unfathomable amount of data only a few years ago is today fairly common. In the future, we will see volumes of data that will stretch traditional storage systems beyond their breaking points. Indeed, the future is arriving ahead of schedule for many organizations. It is already becoming untenably expensive to purchase enough storage capacity and increasingly difficult to meet expected service levels.

In short, organizations are already falling behind the curve of rising storage demands—in order to catch up to the curve and have *any hope at all* of staying ahead of it, organizations need a storage system *designed* for extreme performance and scalable capacity.

Alternatives to DDN?

There are a number of high-capacity storage systems on the market, typically used for file archiving, but the I/O performance of such systems tends to be fairly mediocre. On the other hand, there are high-performance storage systems on the market, but scaling such systems to petabytes of capacity, if it can be done at all, is astronomically expensive. DDN's S2A platform is an interesting hybrid of both high-performance and high-capacity, and not in a compromised, watered down kind of way—S2A combines the highest performance available with the highest density and capacity available.

DDN's primary competitors include LSI, EMC, NetApp, HP, and Isilon. Frankly, the storage systems currently available from those vendors (and others) do not match up particularly well against DDN's S2A platform. For certain applications—video production is a good example, or basically any application where high throughput and capacity are crucial—the advantages of DDN's S2A appliances are so stark that an organization that buys storage from a different vendor is almost certainly making a mistake. For more general purpose applications, DDN's advantages are not as dramatic, but they are still quite real, and buyers are advised to assume that high throughput and capacity *will* be crucial, in the not too distant future, even for general purpose applications. (Even if S2A offers more performance and capacity than is needed *now*, will that still be the case in a year or two? Rarely is “enough” performance or capacity *really* enough.)



**Data
Mobility
Group**

www.datamobilitygroup.com

Summing Up

The extreme amounts of unstructured content that organizations are grappling with, along with the performance and service levels that are expected, have created a need for the exact kind of storage system that DataDirect Networks provides. DDN's S2A storage platform is truly compelling technology and currently without any real equal in the marketplace. Storage architects and administrators should familiarize themselves with DataDirect Networks' S2A platform. 