White Paper

Silicon Storage Architecture (S2A) series from DataDirect Networks

Josh Krischer
April 2008

Silicon Storage Architecture (S2A) series from DataDirect Networks

A new disk storage vendor appears on the radar, or, more precisely, emerges from special applications to more wide-spread usage.
DataDirect Networks was founded ca. ten years ago and is a privately owned, profitable company with over $100M in annual revenue. The company has sold over 150 Petabytes world-wide, installed at tier one customers such as Boeing, CNN, Lawrence Livermore National Laboratory, Microsoft, Time-Warner, and AOL. Global Partners include Cray, IBM, SGI, Sony and Dell, which mainly use DataDirect Networks products for very high performance applications. Forty out of the top 100 supercomputers in the world are attached to DataDirect Networks products, including the most powerful supercomputer in the World, BlueGene/L.

Products

DataDirect Networks products (in their eighth generation) include four subsystems, called the Silicon Storage Architecture (S2A) series, which differ in performance and scalability:

- S2A9900 – maximum scalability of 1.2 PetaBytes, and the ability to read and write at the same speed—6 GBps—which is unique in the industry. This top model was announced in November 2007 at SC07, which is the international conference for high performance computing, storage and networking. The S2A9900 supports eight ports of either 4 or 8 Gbps Fiber Channel or 4 InfiniBand DDR host connections. The systems can be configured for block-level access, or as clustered NAS devices with 2 to 16 NAS heads.
- S2A9700 – Similar scalability as the S2A9900 but lower performance, 3 GBps reads and writes via 8 4GBps FC or 4 InfinBand SDR ports.
- S2A9550 – Scaled-down capacity (up to 960 TBytes) with similar performance as the S2A9700
- S2A6620 – an entry model with up to 660 TBytes optimized for random access which can support up to 30 000 I/O per second. This newest subsystem is planned for general availability in Q3CY08.

S2A9900/9700/6620 models support a mixed configuration of SAS and SATA disk drives and the S2A9550 supports a mix of SATA and Fibre Channel drives. The S2A9900/9700/9550 models are optimized for high performance sequential operations and support optional features such as S2A Clustered NAS, the S2A HPC Storage Solution, S2A Shared SAN File System, and the S2A VTL Solution.

Design concepts

As opposed to the vast majority of the disk storage subsystems, which are designed for databases and transactional applications, the S2A series is designed to deliver very high, constant performance for streaming data and large content files. Instead of employing a traditional RAID 1, 5 or 6, it uses a Reed-Solomon error correction code (ECC), which is also used for error correction in mainframe memory or CDs, for example. The basic data

transfer block is 8 data bytes and two ECC bytes which are striped on 10-disk drives. All 10 bytes are read at once with a two-level error correction circuitry verifying the data and correcting it if necessary. Since on write operations, the data is written to the caches before the full 10 bytes block is staged to the disk, there is no so-called "RAID write overhead", which occurs in traditional RAID 5 and RAID 6 techniques as a result of reading *old parity* and *old data* to calculate the *new parity*.
The hardware is also designed to deliver very high performance. Instead of using commercial processors as most of the competing storage subsystems do, the S2A 9900 models include 16 Intel special purpose processors and 18 FPGAs[1] (divided between both controller units) that facilitate real-time parallel operation.
As a result, the S2A series can deliver very high, consistent performance for write and read operations even during drive failures/rebuilds, a business-enabling capability not available on other storage systems.

Power efficiency

A majority of the S2A product lineup uses SATA and/or SAS disk drives. SATA drives require substantially less energy per Gbyte than the commonly used Fiber Channel disks. In addition, the S2A series supports the *S2A SleepMode* power management feature, which increases the power efficiency of the subsystem. A "sleeping" drive requires 68 percent less energy than a spinning one. A timer can be set to put the drive to "sleep" after an idle time of between one hour and seven days.

Reliability, Availability, and Serviceability (RAS) features

Despite using SATA disks with lower MTBF than the Fiber Channel or SAS drives, the Reed-Salomon based error recovery technique allows for real-time correction of up to two disk failures, resulting in an increase of the Mean Time Between Data Loss (MTBDL) to enterprise-level requirements. A substantial amount of disk problems generally are temporary malfunctions, which usually disappear after a short time. To avoid unnecessary disk replacement and data rebuilding, the S2A uses *controller journaling*. In this process, the presumably failing disk are taken off-line and analyzed by internal diagnostic routines. During this time (usually 3-4 minutes) the writes are routed to special non-volatile memory. If the disk is found functional, the modified data is written from the memory to the disk. These and some other functions are combined into *SATAssure™ Plus*, a unique feature in the industry.
Additional data availability is ensured by an active-active failover / failback structure, dual active-active storage managers, redundant, hot swappable power supplies, and "hot-swappable" components such as storage managers, drives, and drive controller cards. Failure-pointing LEDs (Light Emitting Diodes) help end-users perform some basic maintenance tasks such as the replacement of failing disks if they are interested in partial self-maintenance.

[1] Field Programmable Gate Array – a semiconductor device which can be (re) programmed to implement any (specialized) logical function.

The S2A6620 StorageScaler model

This is the latest model from DataDirect Networks. As opposed to the S2A9xx0 series, this entry level subsystem is optimized for animation rendering, OLTP, e-mail, and IOPs-centric applications. It supports 8 Fiber Channel host interfaces only and is controlled by AMD processors. The S2A uses RAID 5, and optionally, RAID 6 Dual Parity Protection with redundancy groups configurable to 8D+1P (RAID 5) and 8D+2P (RAID 6), respectively. With the S2A6620, DataDirect Networks has ported technology from its flagship S2A9xx0 series into a more compact form factor, and plans to use this new product to target the broader midrange marketplace with an aggressive channel program.

Market and conclusions

The vast majority of currently available storage subsystems are designed for structured data and text, whereas the highest growth in storage demand is for unstructured data and large content files such as digitized audio, video, images, scanned documents, medical imagery, computer simulations, seismic data, etc. Such types of data are usually large files requiring high bandwidth and highly concurrent access to data. Typical Web 2.0 sites include capabilities such as video gaming, social networking, and image and video sharing, animation. Such applications require high availability, high bandwidth, consistent performance, load balancing and a single image view - all guaranteed by DataDirect Networks' S2A series.

New additions to the management team in global sales, marketing and engineering is moving the company from the prestige but niche HPC market to broader availability. This seems to be the right strategy at the right time. Moreover, DataDirect Networks has the right products and enough prestigious references to support its claims regarding performance and availability. What is still on the drawing boards of other vendors, the S2A series already delivers today.

Josh Krischer April, 2008

Josh Krischer is an expert IT advisor with 38 years of experience in high-end computing, storage, disaster recovery, and data center consolidation. Currently working as an independent analyst at Josh Krischer & Associates GmbH, he was formerly a Research Vice President at Gartner, covering enterprise servers and storage from 1998 until 2007. During his career at Gartner, he was responsible for high-end storage-subsystems and disaster recovery techniques. He spoke on these topics and others at a multitude of worldwide IT events, including Gartner conferences and symposia, industry and educational conferences, as well as major vendor events.