Established 1988

- Technology Company (ASICs, FPGA, Firmware, Software)
- System Experts (Optimization, Clusters, Interconnects, Protocols, File Systems, Streaming, Video)
- S2A Introduced June 2000 (Developed from 1997 to 2000, Shipping 6th Gen)

Focused

- High Throughput, High Scalability
- HPC and Media & Entertainment

More Than 1,000 Systems Shipped

6th Generation S2A9500 in Q4’05

Recent Gartner Dataquest report stated: DDN is 5th largest independent storage provider in terms of Market Share, and 3rd largest independent storage provider in volume.

#1 Fastest Computer in the World

- DDN Powers IBM’s BG/L @ LLNL
- S2A Delivers 320 TFlops w/ 1PB of SATA

Top 5 and 36 of the HPC Top50 Sites

- DDN Powers Clusters from IBM, Dell, HP, Cray, SGI, Bull, others…

#1 Tapeless Newsroom in the World

- DDN Powers CNN

300 TV Stations and Media Sites

- DDN Powers Systems from Sony, SGI, Autodesk/Discreet, Pinnacle, Thomson, …

Petabytes Shipped

- 2003
- 2004
- 2005
- 2006
- 2007

02.11.2006 12th ECMWF Workshop
Sample End User Customers

High Performance Computing

Lawrence Livermore National Laboratory

Rich Media

AOL Time Warner

TIME WARNER CABLE

TIME WARNER NY 1

HARPO STUDIOS

FOX SPORTS

WALT DISNEY

LASERPACIFIC MEDIA CORPORATION

STAR
Like straws in a glass of water

- No Switching Latencies
- Greatly reduced Port contention
- No Striping Overhead
- Tested up to 53% improvements just due to host parallelism and PowerLUNs with only 8 hosts

- Congested, Complicated Fabrics
- Lots of Switching Latencies
- Lots of Port Contention
- Host Striping robs CPU Performance
- Small I/O size per Storage Device
- Many more components (higher complexity)
S2Axx00 Host Interfaces

S2A3000 : 8 * FC1 = 800 MB/s peak
S2A8500 : 8 * FC2 = 1600 MB/s peak
S2A9500 : 8 * FC4 = 3200 MB/s peak
**Disk Drive Progress**

<table>
<thead>
<tr>
<th>Cheetah 1 FC</th>
<th>Cheetah 7 FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual ported at 100MB/s</td>
<td>Dual ported at 200MB/s</td>
</tr>
<tr>
<td>1GB capacity</td>
<td>300GB capacity</td>
</tr>
<tr>
<td>Sustained reads at 5MB/s</td>
<td>Sustained reads at 50+MB/s</td>
</tr>
<tr>
<td>6.5mS full stroke seek</td>
<td>6.5mS full stroke seek</td>
</tr>
<tr>
<td>Block reassign in ~1.5s</td>
<td>Block reassign in ~2.5s</td>
</tr>
</tbody>
</table>

**Challenge**: How to achieve dramatic performance increases with no change in disk random performance

**Solution**: High Performance
Silicon Based Storage Controller

- Parallel access for hosts and parallel access to a large number of disk drives
- True performance aggregation and scalability
- Reliability from a parallel pool and QOS
- Drive error recovery in real time and True State Machine Control
PowerLUNs can span arbitrary number of tiers

directRAID
- Equivalent READ & WRITE performance
- No performance degradation in crippled mode
- Tremendous back-end performance for very low-impact rebuild, disk scrubbing, etc.

RAIDed Cache
- Parity Computed on Writes AND Reads

Multi-Tier Storage Support, Fibre Channel, SATA and SAS Disks

Global Hot Spares

Up to 1250 disks total
- 1000 formattable disks
Custom Host Adapters
- FC-4, 10Gb Infiniband
- Others Possible

Custom PCI Bridge FPGAs
- Separates Commands and Data
- Serial to 4KB Parallel Conversion

Custom FPGA Parity Engine

Custom FPGA Disk Controller Engines (DCEs)

Disk Queue Cache and Cache Controllers

Disk Interface Adapters
DirectRAID Data Flow

- **FC-4 or IB Serial Host Interfaces**
- **Parallel Processing Parity Engine**
  - **FPGA PCI Bridge**
    - Generates 512B Parallel Segments
  - **FPGA Parity Engine**
    - Generates One or Optionally Two Parity Segments Synchronously
- **Disk Controller Engines**
  - Queue Command Re-Ordering in Queue Cache
  - Vertical Striping
  - Disk Interfaces

**DirectRAID Data Flow Diagram**

- Host I/F's
- Serial I/F Data Streams
- FPGA PCI Bridge
  - 512Byte Parallel Data Segments
- FPGA Parity Engine
- Disk Controller Engines
  - Queue Cache
  - Disk I/F

02.11.2006 12th ECMWF Workshop
S2A Added Capabilities

- LUN Aliasing (virtualization)
- LUN-in-Cache
  - Solid-State Disk Functionality
- LUN Zoning by Host WWN
- LUN Zoning by Port
- LUNs Permissions
  - Read/Write
  - Read-Only
- Optional directMONITOR Management Console
  - Phone Home, Remote Logging, E-Mail, etc.
- GUI, API
- 8+2 Parity Mode

- Advanced Low-Latency Optimization Modes
- Place Holder LUNs
  - Zero-capacity LUNs
  - “Real” LUNs can be assigned to a host later and mounted without requiring a host reboot to see the LUN
- Performance Analysis
  - I/O Profiling
  - Fibre-Channel Analyzer-Type Functionality
- READ Parity on the fly
- Tunable Background Data Scrubbing
- directMirror LUN Cloning
### Performance, GB/sec

<table>
<thead>
<tr>
<th></th>
<th>WRITES</th>
<th>READS</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDN S2A9500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDN S2A8500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard RAID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DDN 2-3x Faster than Other RAIDs**

### Capacity, Usable TB

<table>
<thead>
<tr>
<th></th>
<th>S-ATA</th>
<th>FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDN S2A9500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDN S2A8500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard RAID</td>
<td>S-ATA</td>
<td>FC</td>
</tr>
</tbody>
</table>

**DDN 10x More Scalable Than Other RAIDs**

(FC:300GB, SATA:500GB)
S2A9500 Single RAID Mode

Tier 1
Tier 2
Tier 8
Tier 16
Tier 32

Ten SxB2016
Nine SxB2016

Full JBOD Redundancy
S2A9500 Double RAID Mode

Parity Enclosures

8*FC4

Tier 1
Tier 2
Tier 8
Tier 44
Tier 48
Tier 49
Tier 74
Tier 96

Ten SAF2248
Ten SAF2248
Disk Enclosure Specifications

- 2Gb/s FC-FC (SFBx016) or FC-SATA (SABx016)
- 3U rackmount
- Single, double, or 6-Channel dual loop
- Sixteen 1” drives
- Fully redundant
- Hot-swappable
New Extreme-Density SATA JBOD

SAx248 SATA Chassis
- 48 Slots
- 4U
- Daisy-Chainable
- 480 Disks per Rack
- 240TB per Rack (500GB SATA Disks)
Five and 20 Chassis Configuration

S2A9500 with

- Five 48-Slot JBODs
- Two Dual Loop per JBOD 240 Disks
- 120TB SATA using 500GB Drives

or

- Twenty 48-Slot JBODs
- Two Dual Loop per JBOD 960 Disks
- 480TB SATA using 500GB Drives
<table>
<thead>
<tr>
<th></th>
<th>FC (10Krpm)</th>
<th>FC (15Krpm)</th>
<th>SATA (7.2Krpm)</th>
<th>SAS (15Krpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Today</td>
<td>73GB</td>
<td>73GB</td>
<td>250GB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>146GB</td>
<td>146GB</td>
<td>400GB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>300GB</td>
<td></td>
<td>500GB</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td>300GB</td>
<td>750GB</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td>450GB</td>
<td>1000GB</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>450</td>
</tr>
</tbody>
</table>
S2A Enables Shared File Systems

• No Switching Latency
• No Port Contention
• No LUN Bottlenecks
• Minimal or No Striping
• Easier to Manage
• Easier to Add Capacity
• Need 3 to 5 Standard RAIDs to match one S2A9500 in a Shared Storage Network

S2A9500
• 3GB/s
  R & W
Only The S2A Enables and Simplifies End-to-End HPC and Media File System Environments
Bull NovaScale
OST
- 8-CPU
- 16 FC-2 Host Connections
- Export via Quadrics
- FC Disks
- 2.2GB/s to 2.8GB/s R&W per OST
- Lustre
## Sample HPC Installations

<table>
<thead>
<tr>
<th>Year</th>
<th>Site</th>
<th>Site</th>
<th>of S2As</th>
<th>File System</th>
<th>Cap</th>
<th>CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Sandia</td>
<td>Albuquerque</td>
<td>14</td>
<td>Lustre, PVFS</td>
<td>250TB</td>
<td>Intel, AMD</td>
</tr>
<tr>
<td>2003</td>
<td>NCSA</td>
<td>Champaign, IL</td>
<td>26</td>
<td>Lustre, XFS, Solaris</td>
<td>145TB</td>
<td>Intel</td>
</tr>
<tr>
<td>2003</td>
<td>LLNL</td>
<td>Livermore</td>
<td>48</td>
<td>Lustre, GPFS AIX</td>
<td>560TB</td>
<td>Intel, AMD, IBM</td>
</tr>
<tr>
<td>2004</td>
<td>Sandia</td>
<td>Albuquerque</td>
<td>14</td>
<td>Lustre, PVFS</td>
<td>500TB</td>
<td>Intel, AMD</td>
</tr>
<tr>
<td>2004</td>
<td>Cray</td>
<td>Multiple Sites</td>
<td>250</td>
<td>Lustre</td>
<td>800TB</td>
<td>AMD</td>
</tr>
<tr>
<td>2004</td>
<td>NCSA</td>
<td>Champaign, IL</td>
<td>20</td>
<td>Ibrix</td>
<td>250TB</td>
<td>Intel, AMD</td>
</tr>
<tr>
<td>2004</td>
<td>LLNL</td>
<td>Livermore</td>
<td>80</td>
<td>Lustre</td>
<td>1.2PB</td>
<td>Intel, AMD</td>
</tr>
<tr>
<td>2004</td>
<td>NASA</td>
<td>Goddard, MD</td>
<td>12</td>
<td>XFS, ADIC</td>
<td>120TB</td>
<td>Intel, AMD</td>
</tr>
<tr>
<td>2004</td>
<td>NERSC</td>
<td>Berkeley</td>
<td>12</td>
<td>Lustre, GPFS AIX</td>
<td>70TB</td>
<td>Intel, AMD</td>
</tr>
<tr>
<td>2004</td>
<td>CINECA</td>
<td>Bologna, IT</td>
<td>4</td>
<td>GPFS AIX</td>
<td>25TB</td>
<td>IBM</td>
</tr>
<tr>
<td>2004</td>
<td>FZK</td>
<td>Karlsruhe, GE</td>
<td>2</td>
<td>GPFS AIX, SNFS</td>
<td>40TB</td>
<td>IBM</td>
</tr>
<tr>
<td>2004</td>
<td>CEA</td>
<td>Paris, France</td>
<td>54</td>
<td>Lustre</td>
<td>1PB</td>
<td>Intel</td>
</tr>
<tr>
<td>2005</td>
<td>NASA</td>
<td>Goddard, MD</td>
<td>6</td>
<td>CXFS</td>
<td>220TB</td>
<td>Intel</td>
</tr>
<tr>
<td>2005</td>
<td>Sandia</td>
<td>Albuquerque</td>
<td>6</td>
<td>CXFS, GPFS Linux</td>
<td>100TB</td>
<td>Intel, IBM</td>
</tr>
<tr>
<td>2006</td>
<td>CEA</td>
<td>Paris, France</td>
<td>22</td>
<td>Lustre</td>
<td>5PB</td>
<td>Intel</td>
</tr>
<tr>
<td>2006</td>
<td>ZIH</td>
<td>Dresden</td>
<td>7</td>
<td>CXFS, Lustre</td>
<td>150TB</td>
<td>SGI, LNXI</td>
</tr>
<tr>
<td>2006</td>
<td>LLNL</td>
<td>Livermore</td>
<td>63</td>
<td>Lustre, GPFS AIX</td>
<td>5PB</td>
<td>Intel, AMD, IBM</td>
</tr>
<tr>
<td>2006</td>
<td>AWE</td>
<td>London, UK</td>
<td>6</td>
<td>Lustre</td>
<td>200TB</td>
<td>Cray</td>
</tr>
<tr>
<td>2006</td>
<td>Sandia</td>
<td>Albuquerque</td>
<td>12</td>
<td>Lustre</td>
<td>500TB</td>
<td>Intel, AMD</td>
</tr>
<tr>
<td>2006</td>
<td>Sandia</td>
<td>Albuquerque</td>
<td>6</td>
<td>Lustre</td>
<td>50TB</td>
<td>Cray</td>
</tr>
<tr>
<td>2006</td>
<td>NERSC</td>
<td>Berkeley</td>
<td>2</td>
<td>Lustre, GPFS AIX</td>
<td>70TB</td>
<td>IBM, Cray</td>
</tr>
<tr>
<td>2006</td>
<td>NOAA</td>
<td>Washington</td>
<td>16</td>
<td>SNFS, GPFS</td>
<td>1.2PB</td>
<td>IBM, Intel</td>
</tr>
<tr>
<td>2006</td>
<td>NCSA</td>
<td>Champaign, IL</td>
<td>2</td>
<td>Lustre</td>
<td>100TB</td>
<td>Intel</td>
</tr>
</tbody>
</table>
S2A Unique Benefits

- Easy to Install and Manage
- Minimum Level of Daisy Chaining
- Double parity (8+P+P’) in hardware
- Parity calculation on Writes and Reads
- Sustained performance up to 2.4 GB/s in Reads and Writes
- Minimal performance degradation in crippled mode
- Up to 896 usable (1120 total) drives behind 1 couplet
- S2A solutions considerably reduce TCO
The Leading Provider of Networked Storage and Clusters for High Performance Computing

Thank You