DataDirect™
NETWORKS

High Performance Storage Solutions

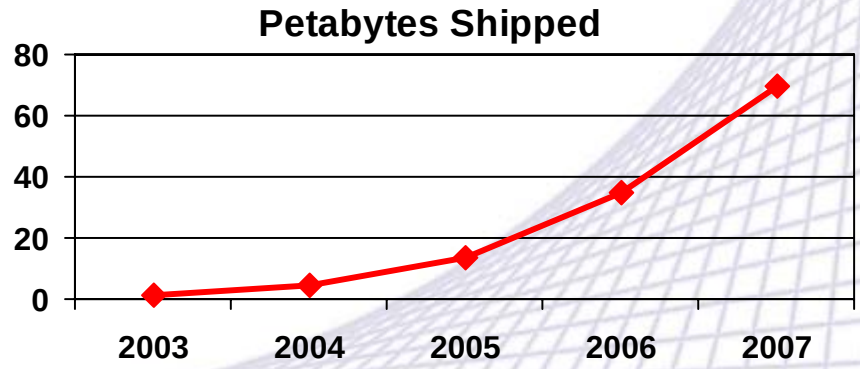
Toine Beckers
tbeckers@datadirectnet.com

www.datadirectnet.com

DataDirect Leadership

- 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, …



Sample OEM & Channel Customers

High Performance Computing

CRAY
IBM
sgi
Linux NetworX
Bull
NEC
CSC
hp invent
DELL
NORTHROP GRUMMAN

Rich Media

SONY
sgi
discreet
SOBEY
PINNACLE SYSTEMS
kasenna
THOMSON MULTI MEDIA BROADCAST & NETWORK SOLUTIONS
INCITE
DAYANG
TWI interactive
TOSHIBA

Sample End User Customers

High Performance Computing

Lawrence Livermore National Laboratory

NOAA

Sandia National Laboratories

National Geospatial-Intelligence Agency

cea

Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft

CINECA

AWE

DKRZ

ZIH
Zentrum für Informationsdienste und Hochleistungsrechnen

NERSC

NASA

Berkeley Lab

ARL

NCSA

Oak Ridge National Laboratory

Rich Media

EUROSPORT

WB

AOL Time Warner

TIME WARNER CABLE

NY 1

CNN

HARPO STUDIOS

SWR

E!

BBC

FOX SPORTS

CCTV

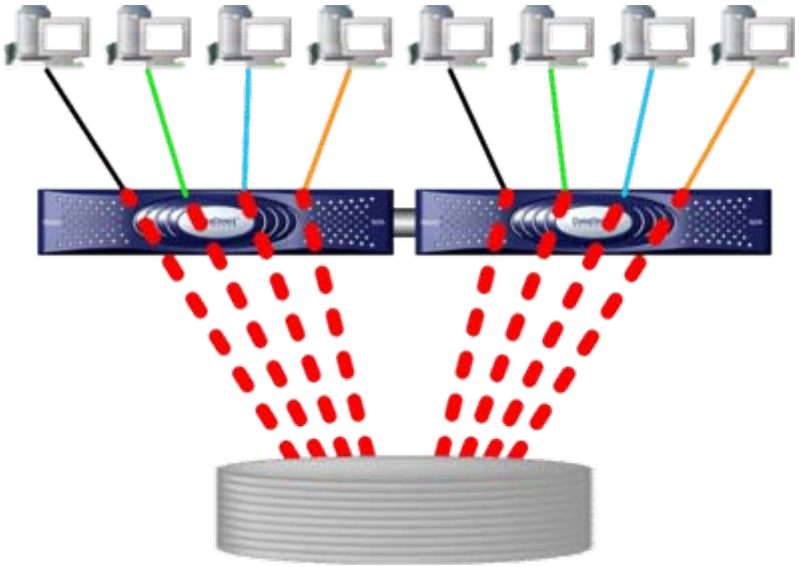
Walt Disney

LASERPACIFIC MEDIA CORPORATION

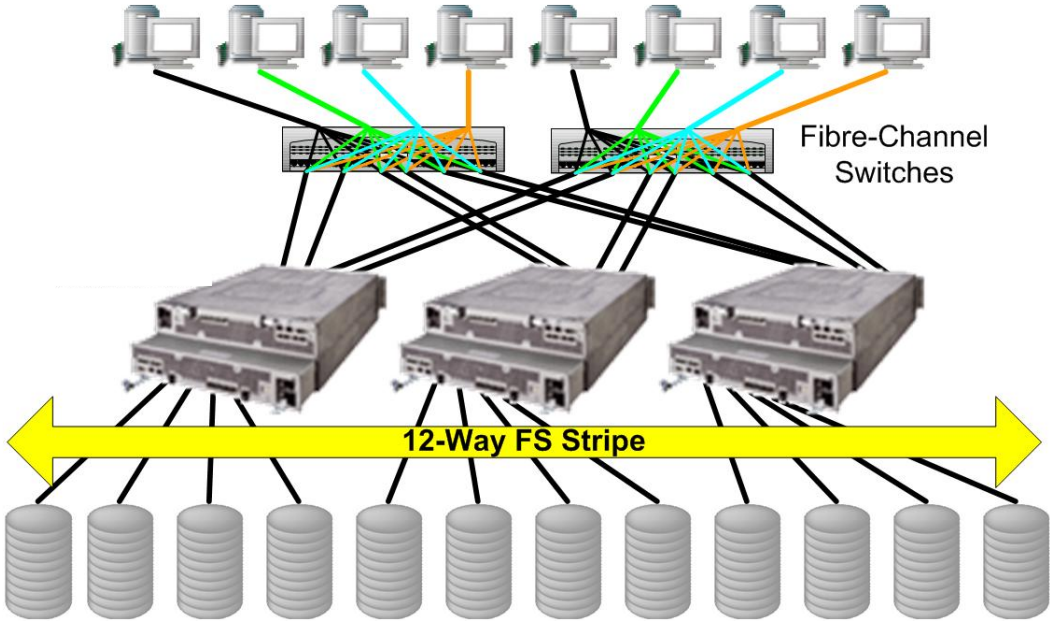
STAR

The S²A: Architecturally Unique

DDN S²A9500 Content Access: Host Parallelism and PowerLUNs



Generic RAID SAN Architecture

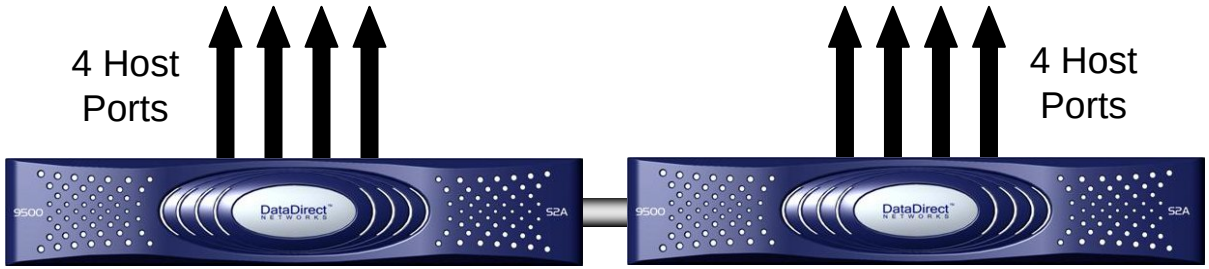


Like straws in a glass of water

- 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

Cheetah 1 FC

- Dual ported at 100MB/s
- 1GB capacity
- Sustained reads at 5MB/s
- 6.5mS full stroke seek
- Block reassign in ~1.5s

Cheetah 7 FC

- Dual ported at 200MB/s
- 300GB capacity
- Sustained reads at 50+MB/s
- 6.5mS full stroke seek
- Block reassign in ~2.5s

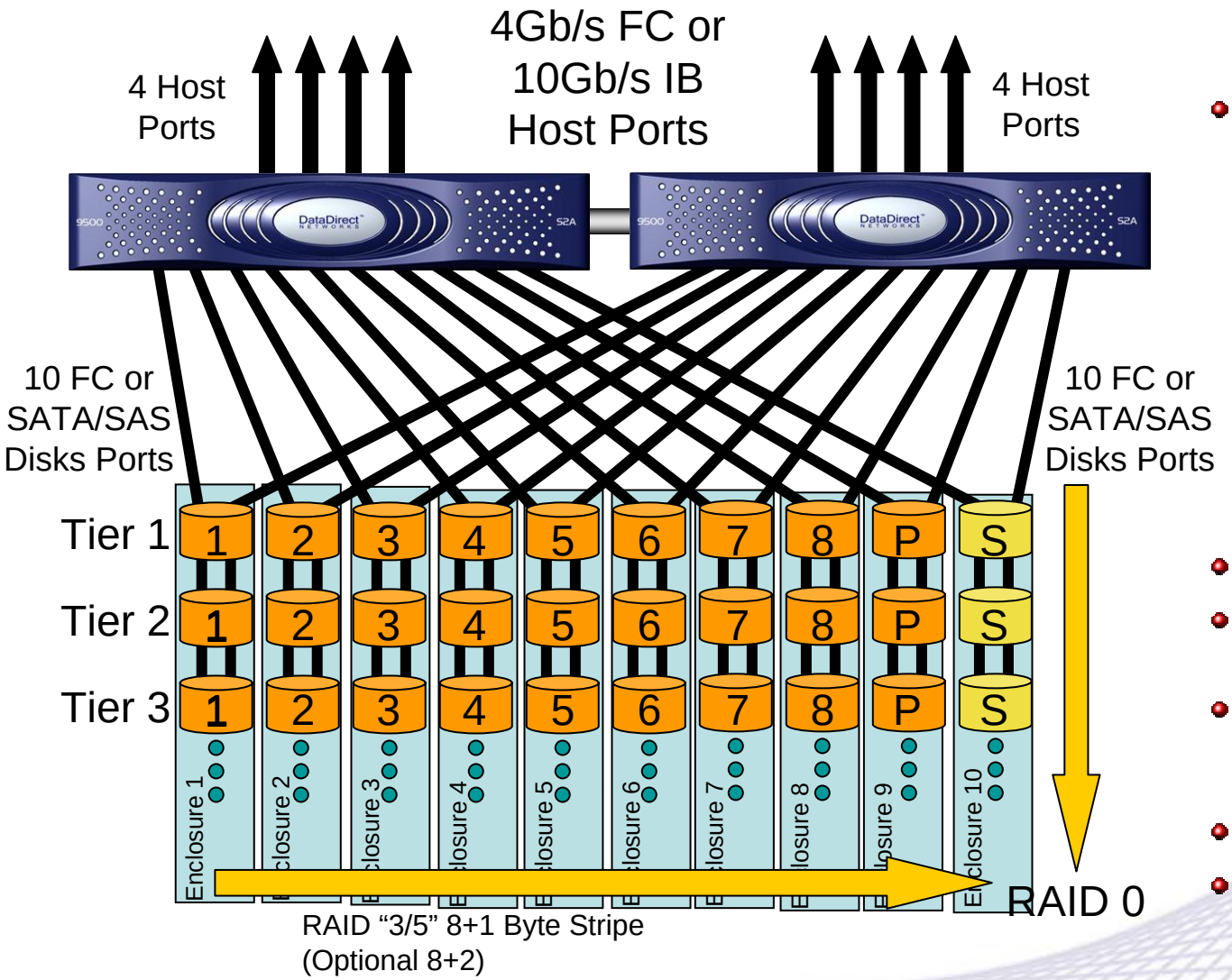
Challenge: How to achieve dramatic performance increases with no change in disk random performance

---

Solution: High Performance Silicon Based Storage Controller

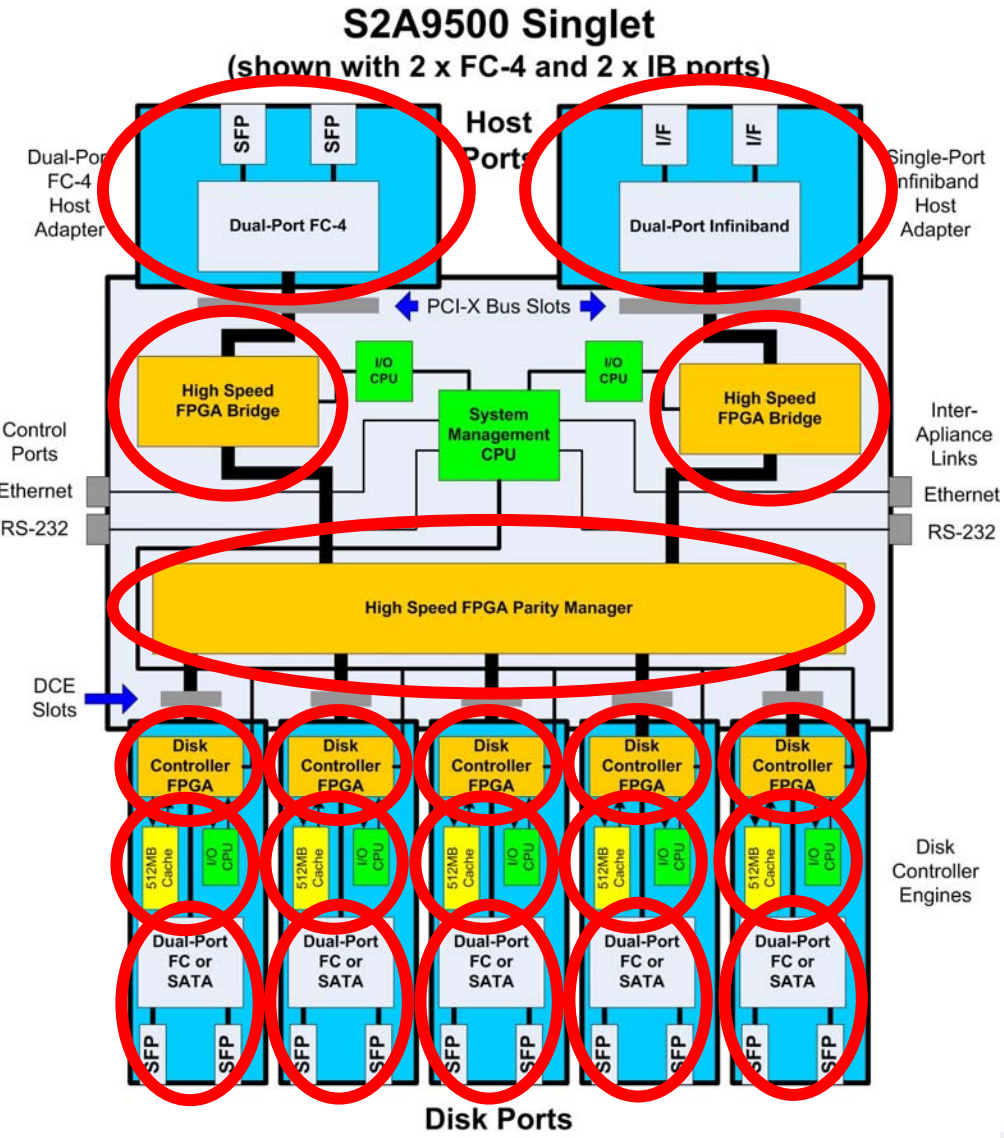
- **Parallel access for hosts and parallel access to a large number of disk drives**
  - **True performance aggregation and scalability**
    - **Reliability from a parallel pool and QOS**
  - **Drive error recovery in real time and True State Machine Control**

S2A9500 Basic Configuration



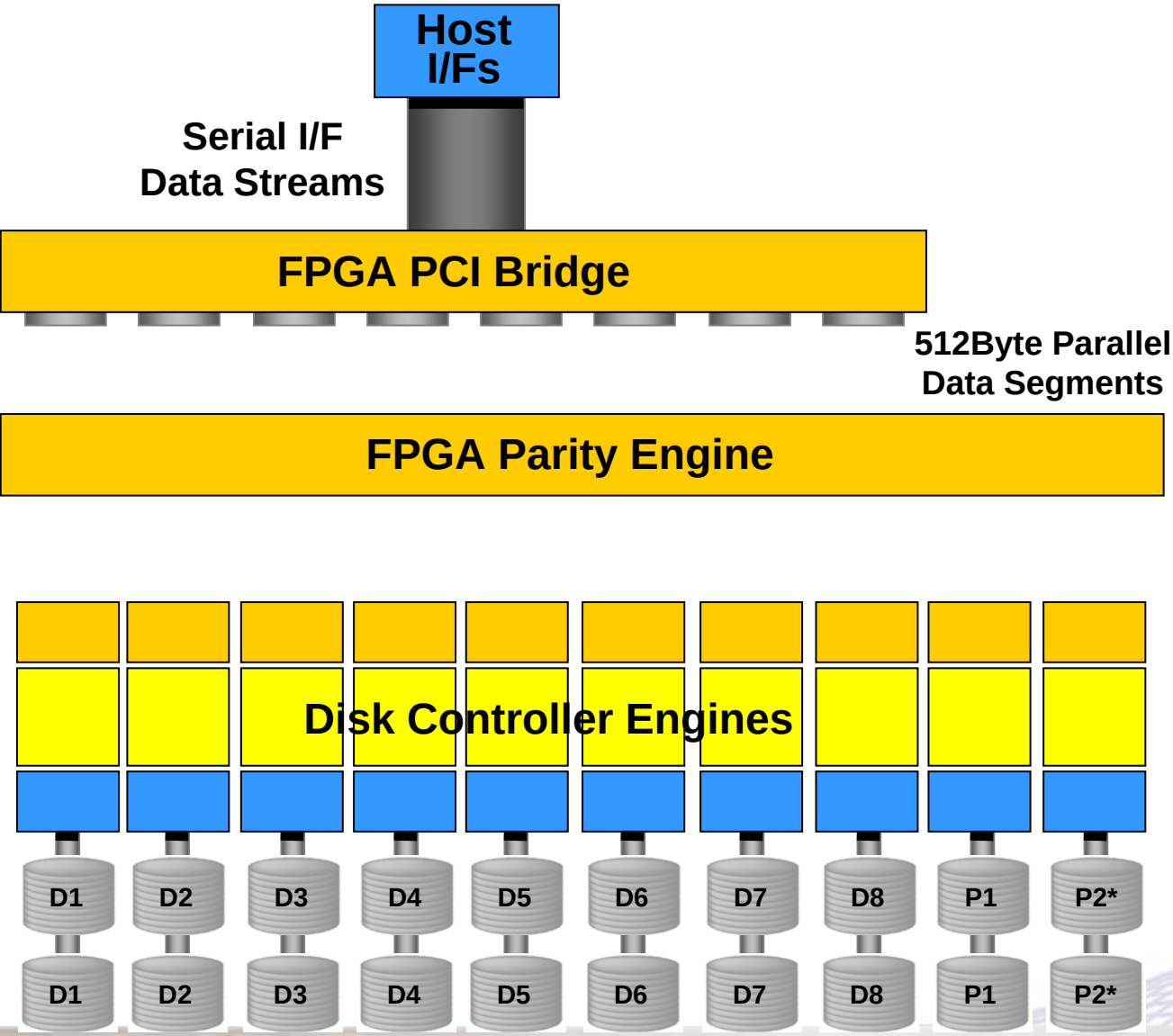
- **PowerLUNs can span arbitrary number of Tiers**
- **directRAID**
  - **Equivalent READ & WRITE performance**
  - **No performance degradation in crippled mode**
  - **Tremendous back-end performance for very low-impact rebuild, disk scrubbing, etc.**
- **RAIDed Cache**
- **Parity Computed on Writes AND Reads**
- **Multi-Tier Storage Support, Fibre Channel, SATA and SAS Disks**
- **Global Hot Spares**
- **Up to 1250 disks total**
  - **1000 formattable disks**

S2A9500Singlet Block Diagram



- 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



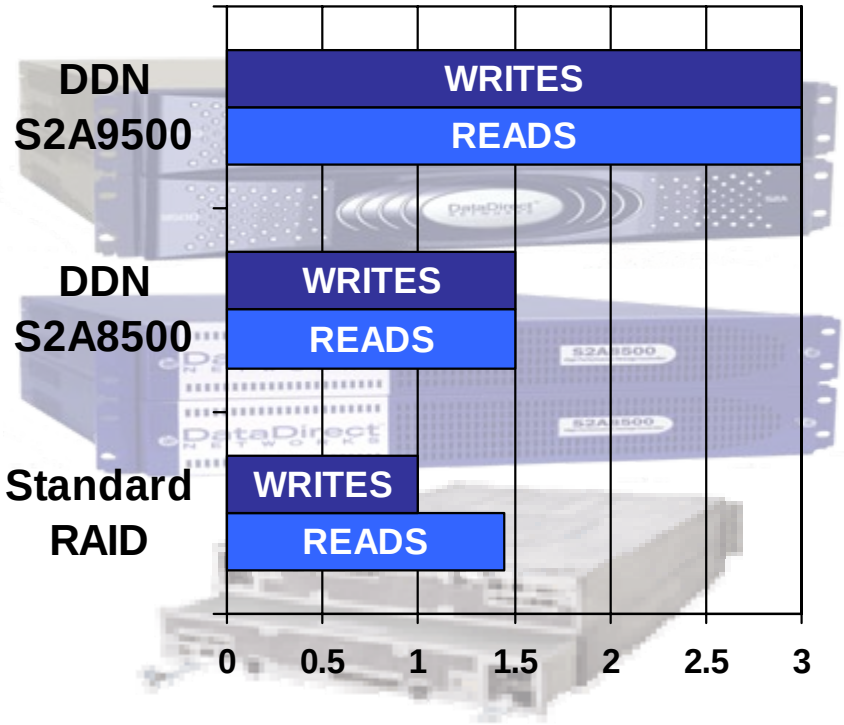
- FC-4 or IB Serial Host Interfaces
- Parallel Processing Parity Engine
  - FPGA PCI Bridge
    - Generates 512B Parallel Segments
  - FPGA Parity Engine
    - Generates One or Optionally Two Parity Segments Synchronously
  - Disk Controller Engines
    - Queue Command Re-Ordering in Queue Cache
   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

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

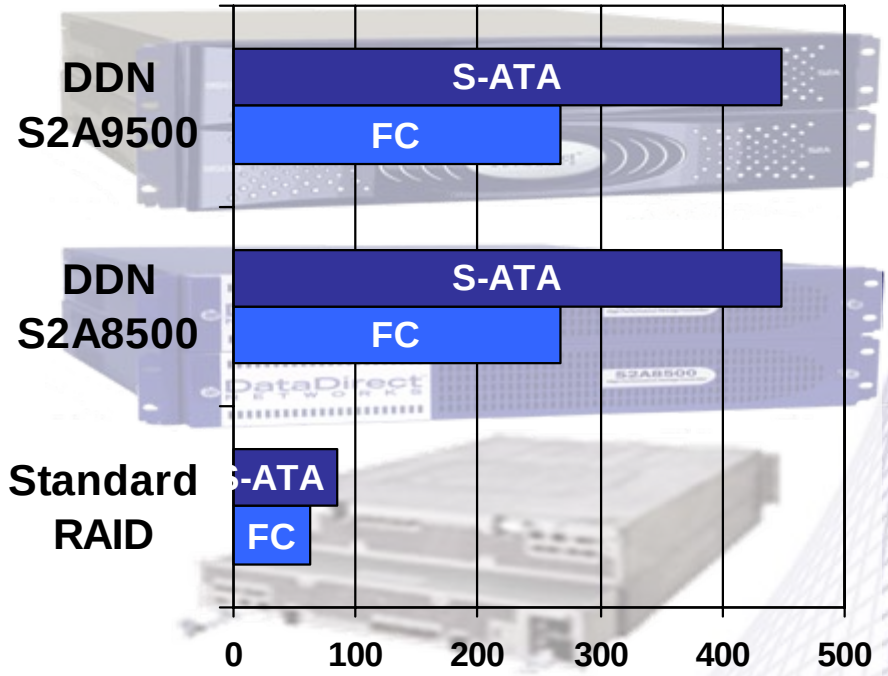
Nominal Performance & Capacity Scalability

Performance, GB/sec



DDN 2-3x Faster than Other RAIDs

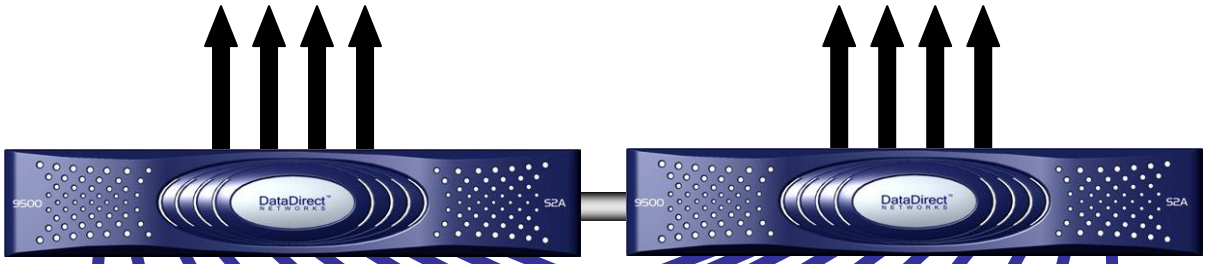
Capacity, Usable TB



DDN 10x More Scalable Than Other RAIDs

(FC:300GB, SATA:500GB)

S2A9500 Single RAID Mode



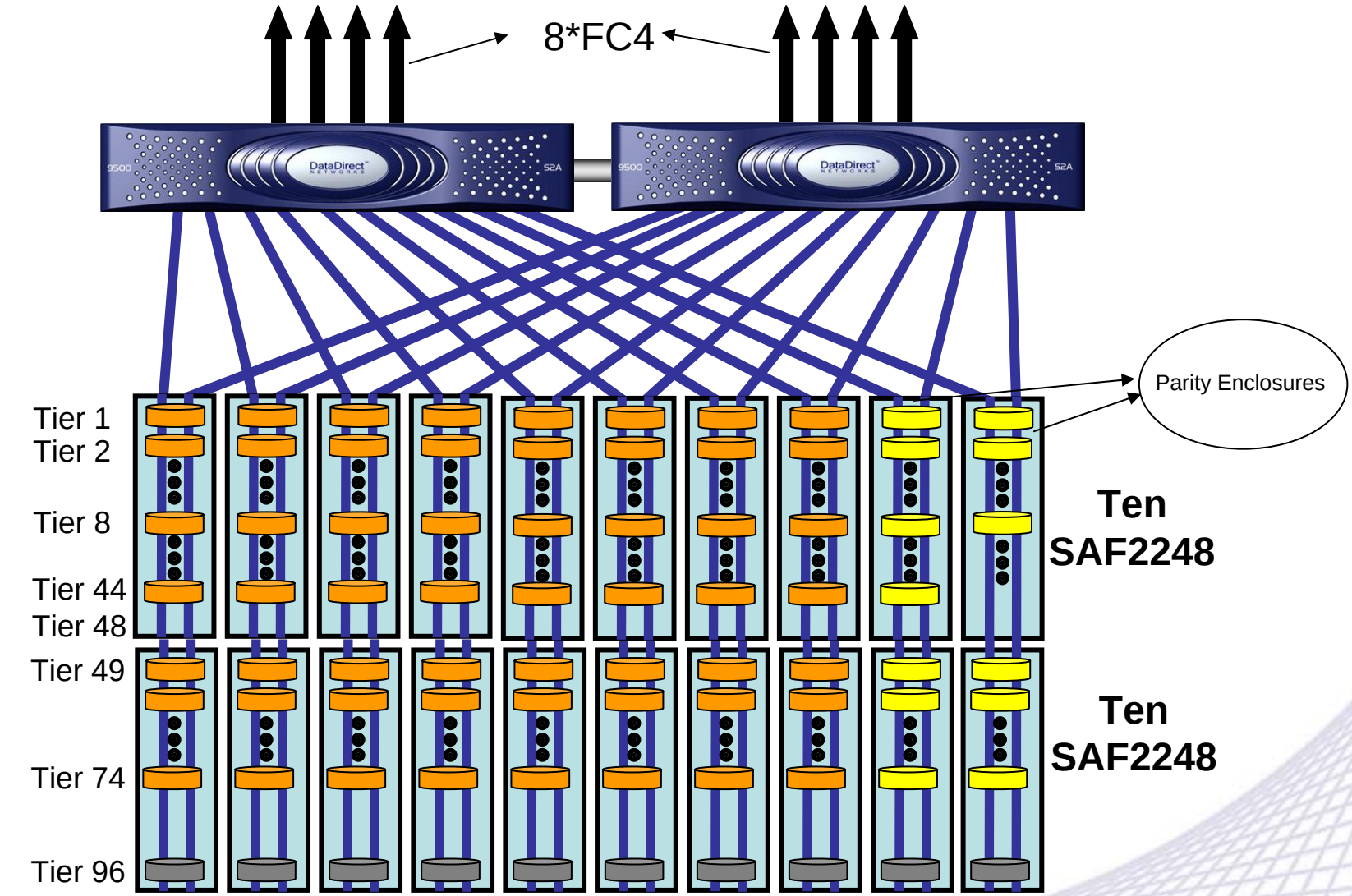
Tier 1
Tier 2
Tier 8
Tier 16
Tier 32

Ten SxB2016

Nine SxB2016

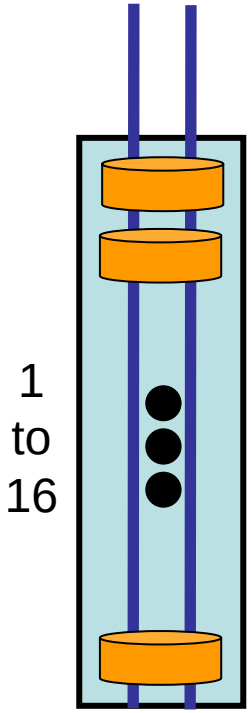
Full JBOD Redundancy

S2A9500 Double RAID Mode

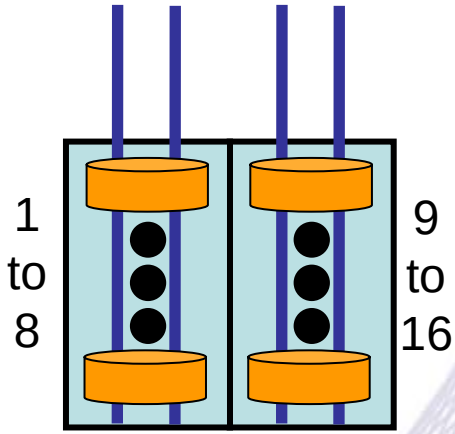


Disk Enclosure Specifications

- 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

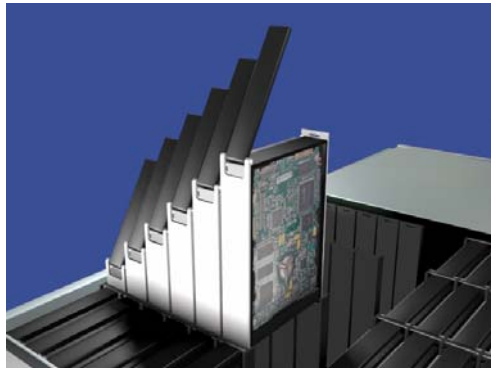


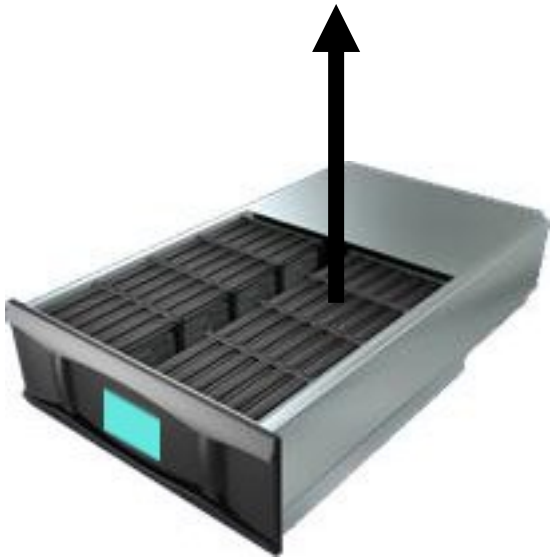
SxB2016

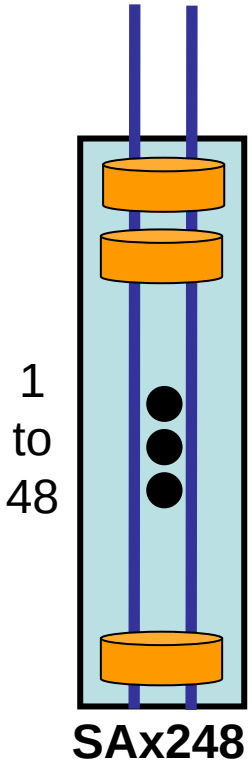


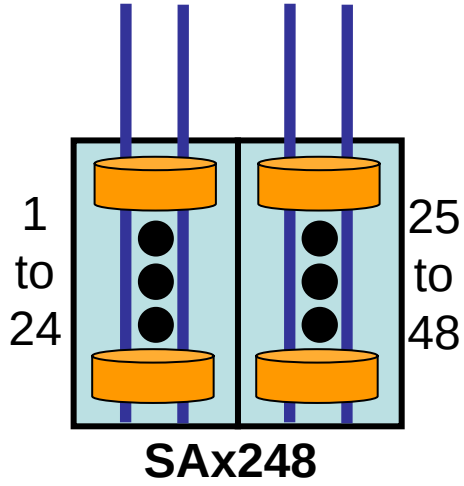
SxB4016

New Extreme-Density SATA JBOD





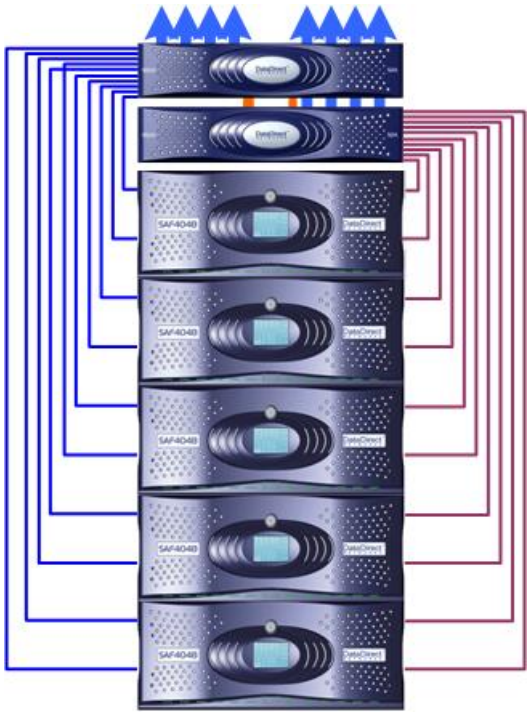




SAx248 SATA Chassis

- 48 Slots
- 4U
- Daisy-Chainable
- 480 Disks per Rack
- 240TB per Rack (500GB SATA Disks)

Five and 20 Chassis Configuration

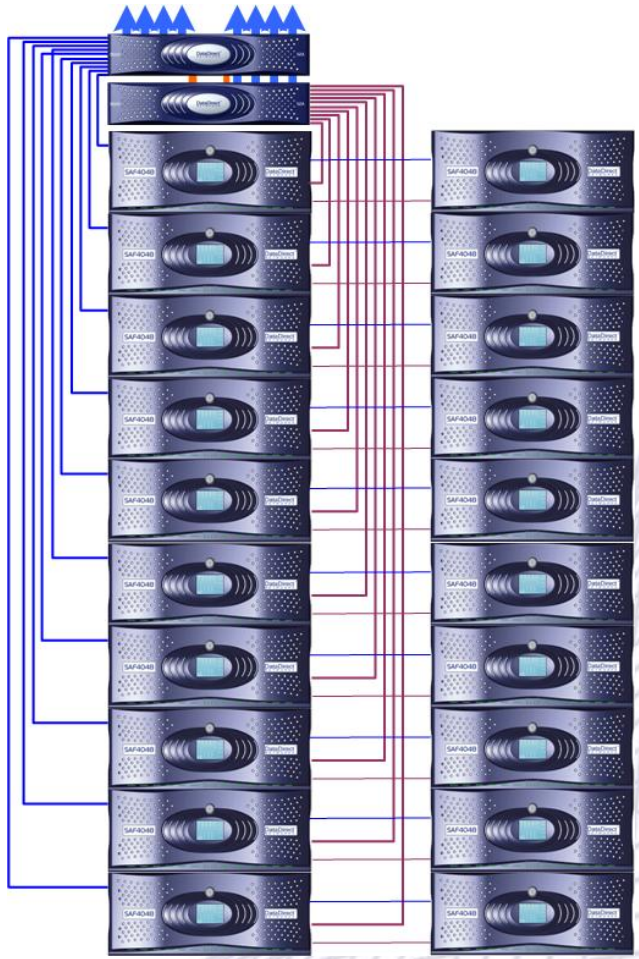


S2A9500 with

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

or

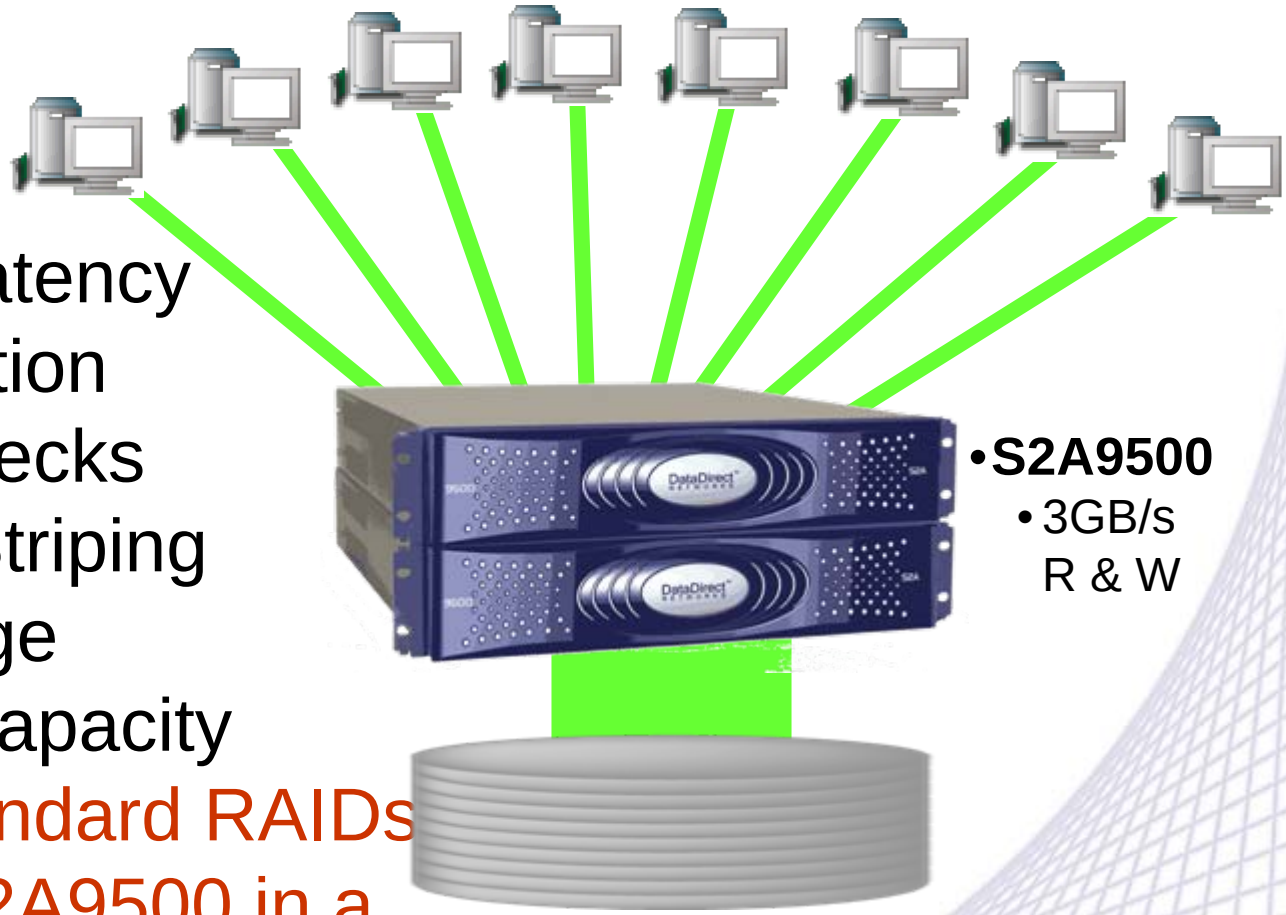
- Twenty 48-Slot JBODs
- Two Dual Loop per JBOD 960 Disks
- 480TB SATA using 500GB Drives



Drive Densities

	FC (10Krpm)	FC (15Krpm)	SATA (7.2Krpm)	SAS (15Krpm)
Today	73GB	73GB	250GB	
	146GB	146GB	400GB	
	300GB		500GB	
2006		300GB	750GB	
2007		450GB	1000GB	146
				300
				450

S2A Enables Shared File Systems

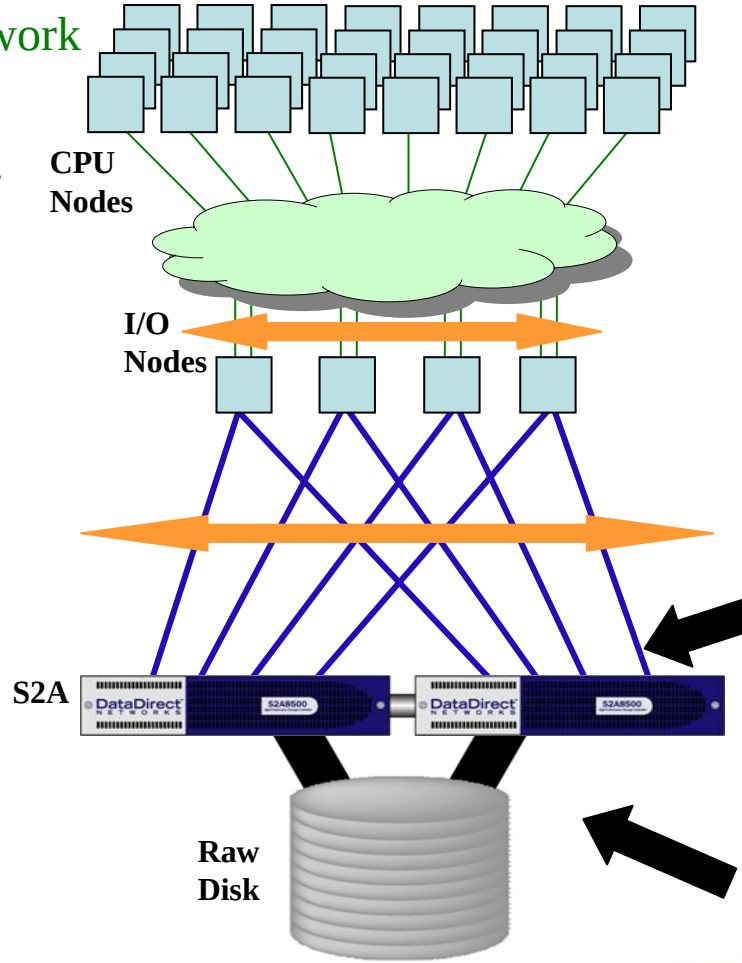


- No Switching Latency
- No Port Contention
- No LUN Bottlenecks
- Minimal or No Striping
- Easier to Manage
- Easier to Add Capacity
- Need 3 to 5 Standard RAIDs to match one S2A9500 in a Shared Storage Network

- **S2A9500**
  - 3GB/s R & W

S2A Enabling File System I/O Stack

- File I/O Services over Network
  - NFS,CIFS/SMB, FTP
  - Custom (Lustre Client), etc.
- Physical I/F Options
  - Ethernet, Infiniband, Quadrics, Myrinet, etc.
- Block I/O RAID Services over Channel I/F
  - SCSI, RDMA
- Physical I/F Options
  - FC, IB, AS
- Block I/O Disk Services
  - FC, SATA, SAS



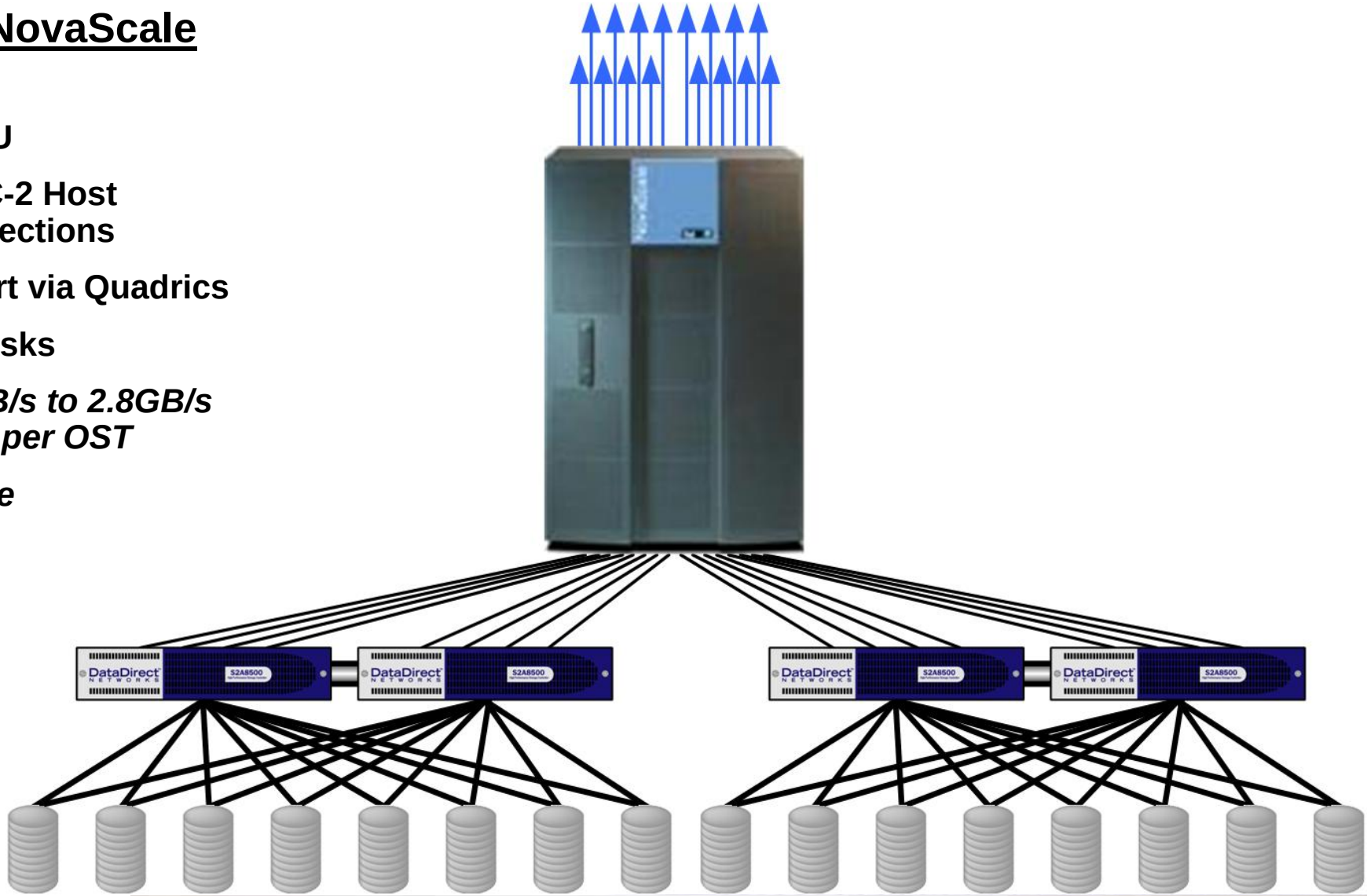
- Parallel File System
  - Lustre, GPFS, SNFS
  - Striped/Aggregated at File or Object Level
- SAN File System
  - Ibrix, GFS, CXFS, StorNext, Polyserve
  - Striped/Aggregated at Block Storage Level
- S2A Parallel Host Access
  - Contentionless
  - Zero Switching Latency
  - Zero-Time Failover
- S2A Parallel Disk Access
  - PowerLUNs
  - Huge Scalability

Only The S2A Enables and Simplifies End-to-End HPC and Media File System Environments

Benchmarked CEA OST Configuration

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

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

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

DataDirect NETWORKS

DataDirect™
NETWORKS

The Leading Provider of Networked Storage and Clusters for High Performance Computing

Thank You