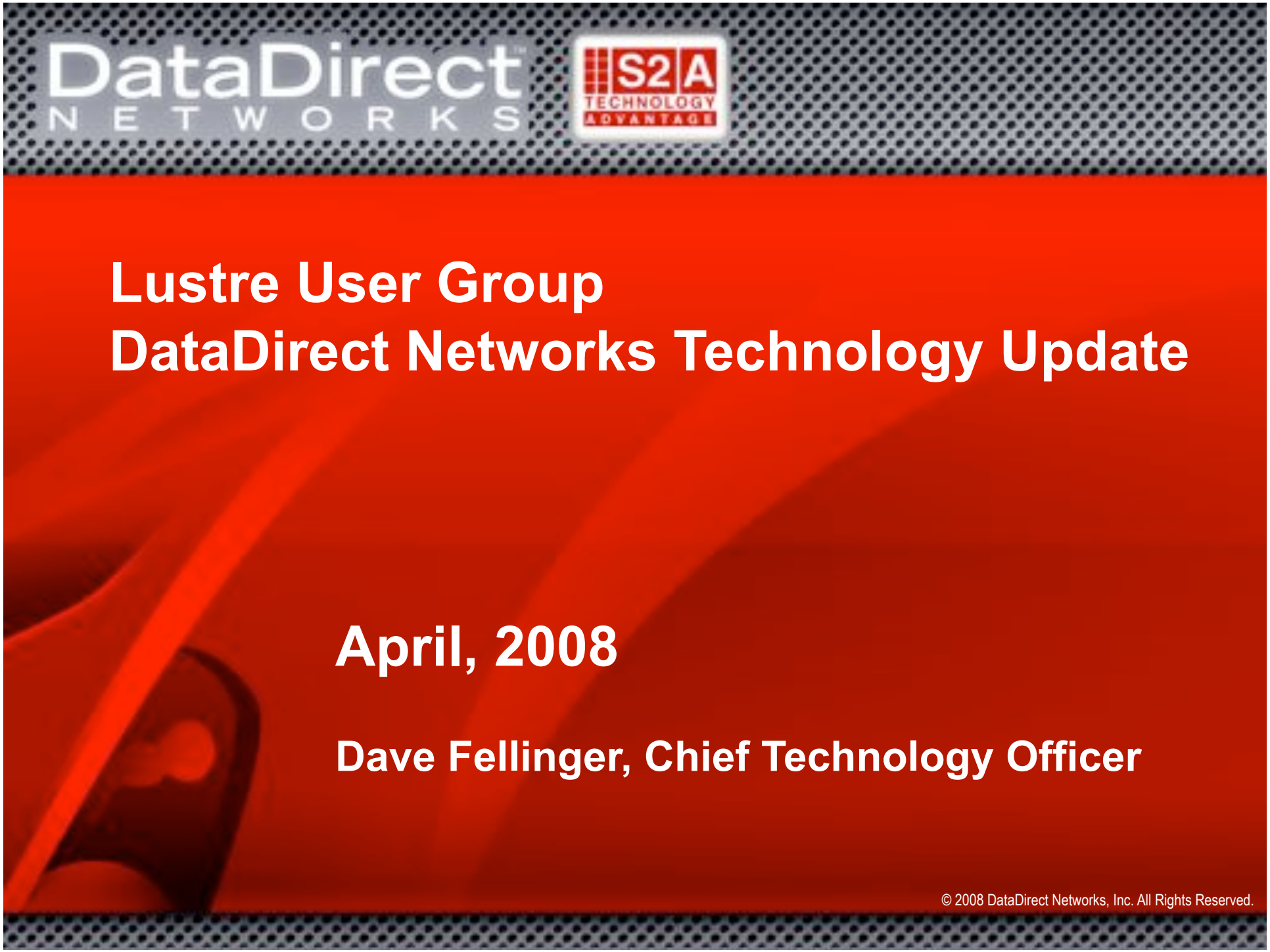

DataDirect
NETWORKS
S2A
TECHNOLOGY
ADVANTAGE
Lustre User Group
DataDirect Networks Technology Update
April, 2008
Dave Fellinger, Chief Technology Officer
© 2008 DataDirect Networks, Inc. All Rights Reserved.

Agenda

- **S2A Parallel Storage Architecture**
- S2A9900 StorageScaler
- S2A6620
- DDN HPCSS Lustre Offering
- Future Requirements

Parallel Storage Goals

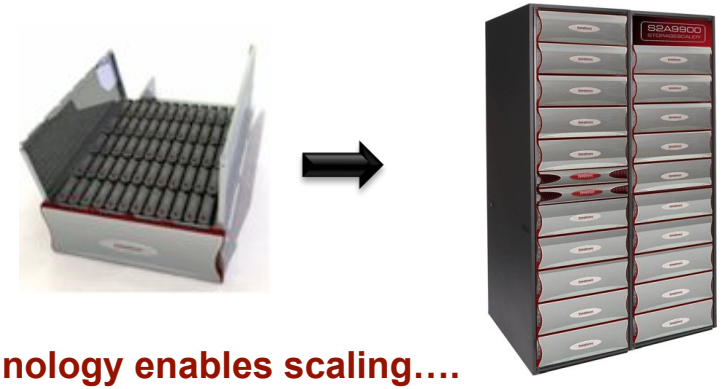
DataDirect NETWORKS S2A

Low Latency - High Performance, Silicon Based Storage Appliance

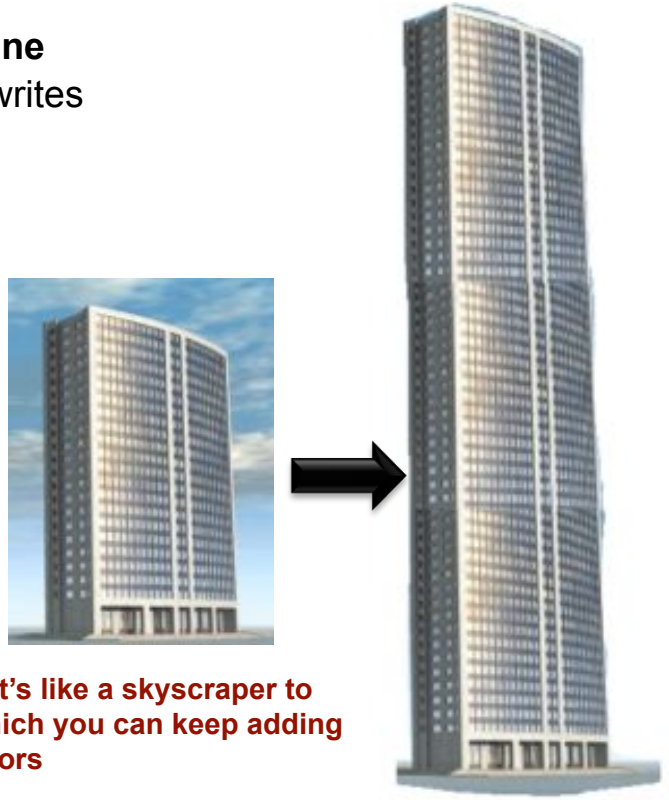
- Parallel Access For Hosts
- Parallel Access To A Large Number Of Disk Drives
- True Performance Aggregation
- Reliability From A Parallel Pool
- Quality Of Service
- Scalability
- Drive Error Recovery In Real Time
- True State Machine Control
  - 10 Virtex 4 FPGAs, 16 Intel embedded processors, 8 Data FPGAs

DataDirect Technology

- **DirectOS: Core S2A Operating System**
  - Storage Management Features (DirectRAID, PowerLUN, SATAssure, Partial Rebuild)
  - Network and Host Management (LUN Masking/Zoning, Infiniband RDMA, Real Time Mode)
  - System Tuning Utilities
  - Field Upgradeable
- **DirectRAID: Scalable, High Performance Data Protection Engine**
  - Parity and double parity calculated in real-time on reads and writes
  - Multiple paths to data
  - *Writes are as fast as reads*
- **SATAssure: Intelligent and Reliable SATA Drive Management**
  - Delivers enterprise-class data protection
  - Makes large SATA pools reliable (not just less expensive)
  - Detects and corrects silent data corruption

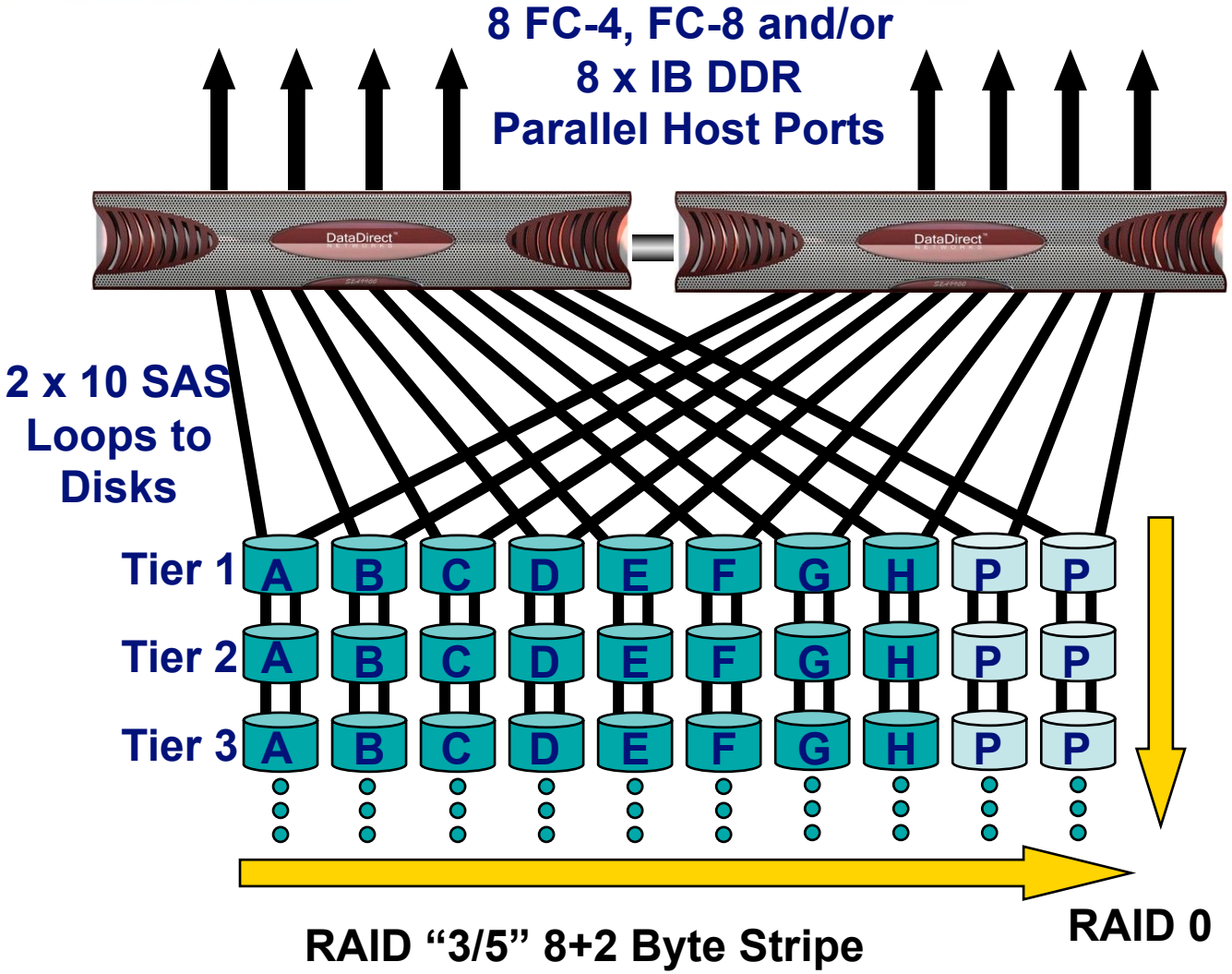


S2A Technology enables scaling….



…it's like a skyscraper to which you can keep adding floors

An Implementation of Parallelism w/ Double Parity RAID Protection

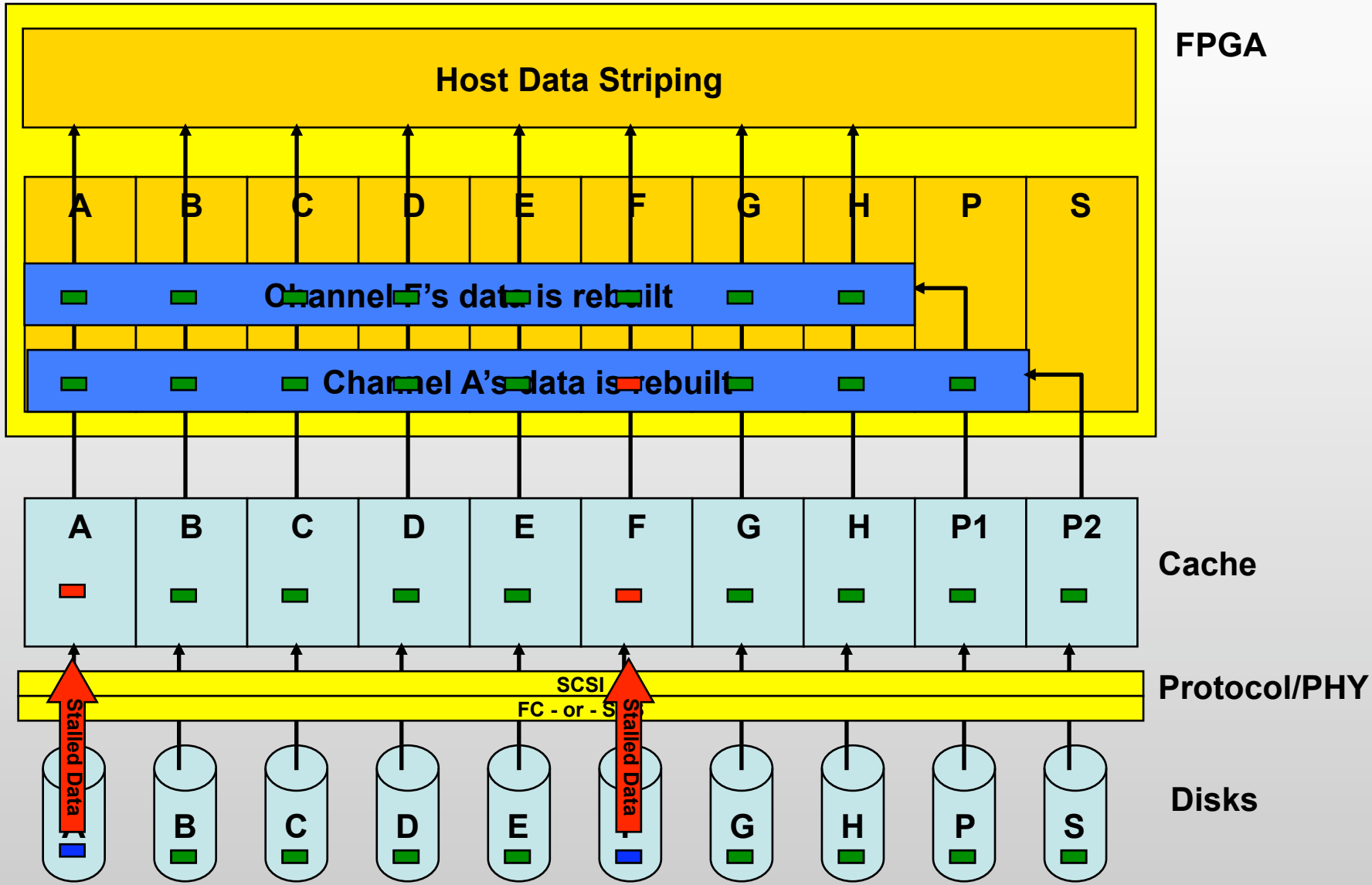


- Double Disk Failure Protection
- LUNs can span tiers
- All ports access all storage
- Implemented in Hardware State Machine
  - No penalty for RAID 6!
- Parity Computed On Writes AND Reads
- No loss of performance on any failure
- Multi-Tier Storage Support, SAS or SATA Disks
- Up to 1200 disks total
  - 960 formattable disks

Quality of Service

- S2A always reads (and writes) to all members of a RAID group
- FPGA designed to generate host data with missing elements
- If a single member of RAID group is slowed by internal error recovery S2A can still provide host data at a high level of QOS

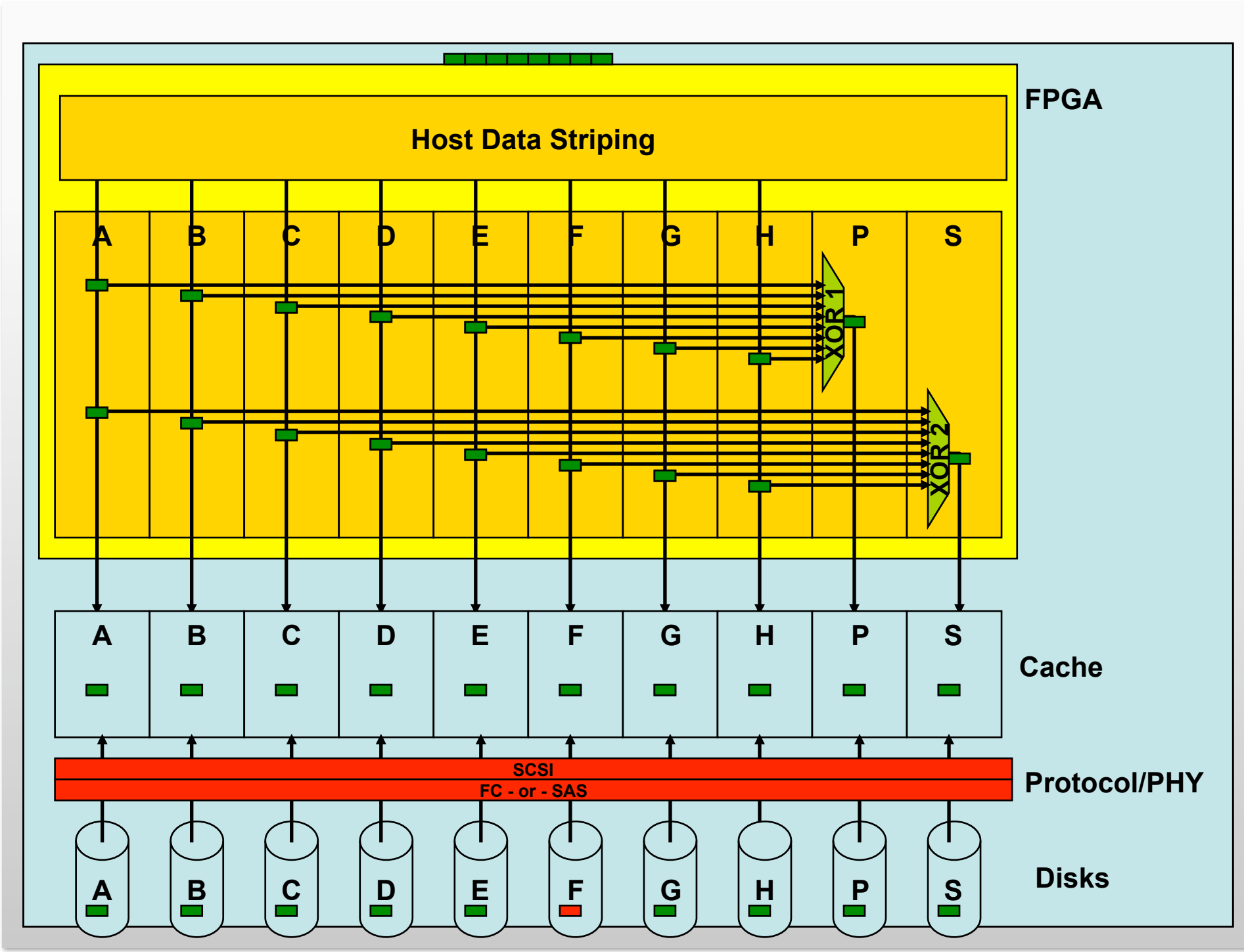
Quality of service



Data Corruption Error Handling

Note that the Cache and Disks have not corrected the data corruption.

We will need to rebuild the data into the cache and flush the data back to the disk in order to repair the problem fully.


FPGA
Host Data Striping
A
B
C
D
E
F
G
H
P
S
XOR 1
XOR 2
A
B
C
D
E
F
G
H
P
S
Cache
SCSI
FC - or - SAS
Protocol/PHY
A
B
C
D
E
F
G
H
P
S
Disks

SATAssure Data Integrity

- **SATAssure makes SATA drives behave like more expensive enterprise-class drives**
  - S2A hardware enables SATAssure software to verify all data read from the disks
  - S2A hardware allows SATAssure to send hosts “fixed” data (data integrity is assured)
  - S2A hardware enables SATAssure to correct data on the disk for future accesses (self-healing array)
  - Multiple levels of disk recovery attempted before failing drives (replace fewer drives)
  - S2A controller journaling allows partial rebuilds (less time in degraded mode)

Agenda

- S2A Parallel Storage Architecture
- **S2A9900 StorageScaler**
- S2A6620
- DDN HPCSS Lustre Offering
- Future Requirements

S2A9900 StorageScaler

The World's First Petabyte-Class Storage System

S2A9900

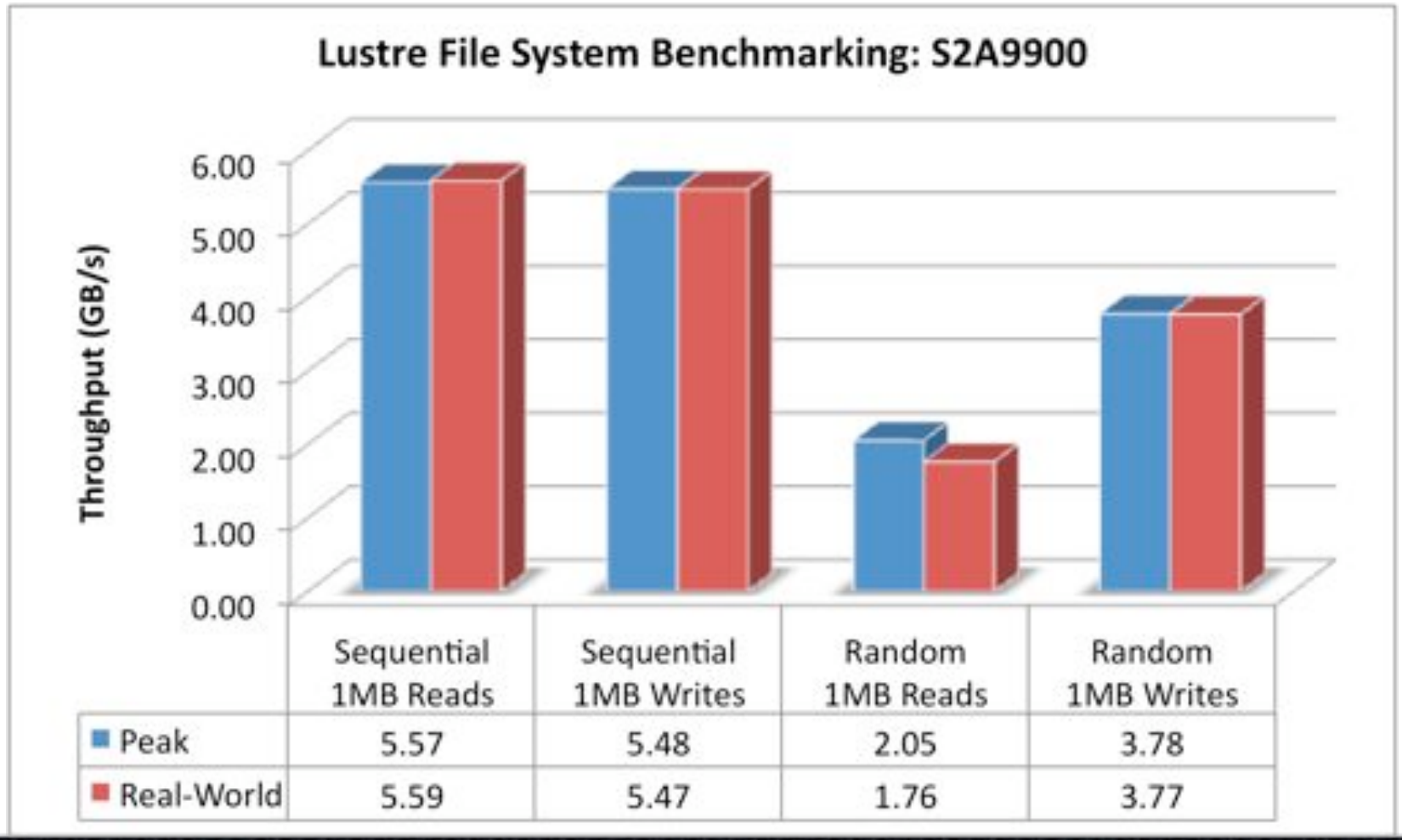
- **8th Generation Platform**
- **Design Goals:**
  - Double throughput performance and 3x IOPS over S2A9550
  - Provide extremely high disk-side bandwidth to enable file systems and storage applications
  - Allow enterprise-class and SATA drives within the same system for storage tiering and HSM
  - Improve density and maximum system capacity
  - Utilize SAS drives/interconnects
  - Further enablement of InfiniBand clusters
  - Continue DDN leadership in $/performance and TB/sq.ft.

Key Highlights

- **2.4-2.8 GB/s sustained bandwidth per singlet**
- **~3x IOPs of 9550**
- **PCIe connections to hosts**
  - DirectOS 5.00: 4Gb FC and 4x IB DDR
  - DirectOS 5.05: also supporting 8Gb FC
- **10 SAS (4x) connections to disks per singlet**
  - **24GB/s** of Internal Bandwidth
- **Internal Hard Drive**

S2A9900 + Lustre

- Checkpoint Faster! Future-Proof for Multi-Core!



	Sequential 1MB Reads	Sequential 1MB Writes	Random 1MB Reads	Random 1MB Writes
Peak	5.57	5.48	2.05	3.78
Real-World	5.59	5.47	1.76	3.77

Why SAS?

- **Drive and enclosure manufacturers moving from FC to SAS**
  - Lower cost infrastructure
- **Native support for SATA over SAS**
  - No FC←→SATA bridge; reduces cost and complexity
- **Potential to mix SAS and SATA drives in same enclosure**
  - Serial SCSI Protocol (SSP)
  - Serial ATA Tunneling Protocol (STP)
- **Potential for large configurations**
  - 16,384 devices theoretical maximum
  - Facilitated by SAS expanders
- **Excellent roadmap**
  - 6 Gb/s on the horizon

S2A9900 Specifications

Specification	S2A9900 Couplet	S2A9550 Couplet
Supported Disk Technology	SAS & SATA (in same unit)	Fibre Channel & SATA
RAID Parity Protection	RAID6 8+2	RAID3 (8+1+1), RAID6 8+2
Sustained Throughput	5.8GB/s – 5.9GB/s	2.4 GB/s – 2.8GB/s
IOPS	40,000	14,000
Cache	5.0GB ECC/RAID Protected	5.0GB ECC/RAID Protected
Disk Side Ports/Port Type : Total Back-End Bandwidth	20 / SAS 4 Lane : 24GB/s	20 / FC-2 : 4GB/s
Host Side FC Ports	8 x IB 4x DDR or 8 x FC-4 or 8 x FC-8	8 x FC-4 or 8 x IB 4x
Dimensions	7 x 19 x 28 in. (4U)	7 x 19 x 25 in. (4U)

Blue Text Denotes Change from S2A9550

S2A = Storage without Compromise

DataDirect NETWORKS S2A

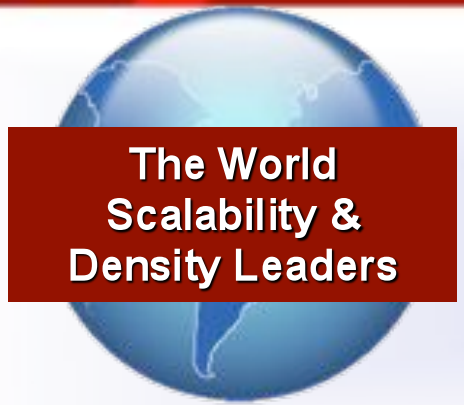
S2A9900 has all S2A9550 features

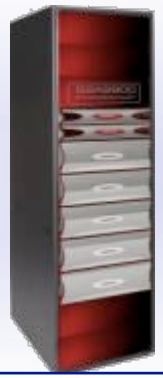
- Massive throughput performance
- Scalable capacity & dense footprint
- Full-speed RAID 6 data protection
- No “degraded mode”
- Writes occur as fast as reads
- “Self Healing” array
- All data available from all host ports

- **Turbocharges the storage network:**
  - Use fewer storage systems
  - Manage fewer devices and applications
  - Save power
  - Save floor space

Scalability & Density

DataDirect NETWORKS S2A





S2A TECHNOLOGY	5 Enclosures 24U: 1/2 Rack	10 Enclosures 44U: 1 Rack	20 Enclosures 84U: 2 Racks
S2A9900 S2A9700	Up to 300 Drives Up to 300TB	Up to 600 Drives Up to 600TB	Up to 1,200 Drives Up to 1.2PB
S2A9550	Up to 240 Drives Up to 240TB	Up to 480 Drives Up to 480TB	Up to 960 Drives Up to 960TB

- **Simple Cabling:** All Enclosures are direct connected (up to 10 enclosures) to the S2A Appliances for easy configuration and maximum reliability.
- **Maximum Availability:** S2A Storage Systems can lose **up to 20%** of the available drive enclosures without impacting host performance or data availability.

SleepMode: MAID Technology Ideal for Data Archiving

- **Leading Power Efficiency**
  - Only 4 x 30A 220V Drops per 600TB
  - Dense Packaging to Reduce Space/ Cooling
    - Up to 600TB/rack
- **Truly Green Storage!**

- **S2A SleepMode™**
  - Intelligent Power Management
  - Optimized for Backup/VTL/Archive
  - Spin Down Tiers of Inactive Drives
    - 12 seconds to spin up

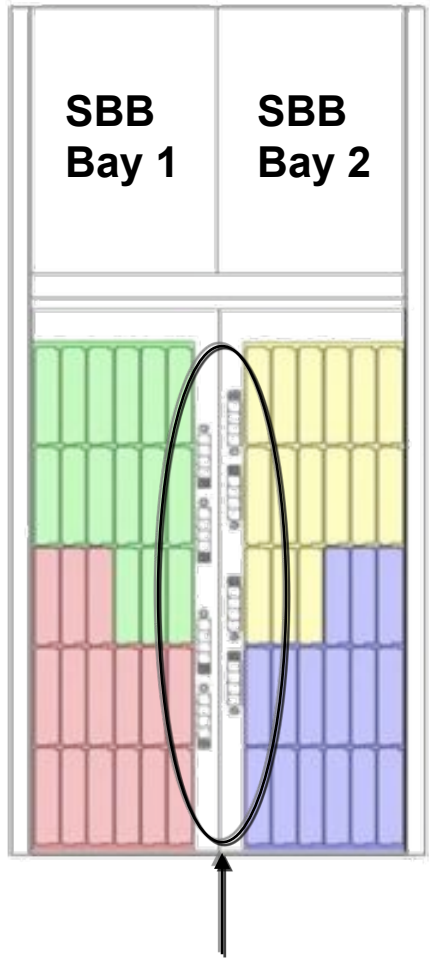
	S2A9900	
	Active	SleepMode*
300TB (300 x 1TB SATA)	7.1 kW	**4.5 kW**
600TB (600 x 1TB SATA)	13.5 kW	**8.29 kW**
1.2PB (1200 x 1TB SATA)	26.1 kW	**15.8 kW**
1.2PB SleepMode Savings		**Up to $36,000/yr**

	S2A9550/S2A9700	
	Active	SleepMode*
240TB (240 x 1TB SATA)	4.6 kW	**2.9 kW**
480TB (480 x 1TB SATA)	8.7 kW	**5.3 kW**
960TB (960 x 1TB SATA)	17 kW	**10.1 kW**
960TB SleepMode Savings		**Up to $23,900/yr**

**** S2A SleepMode Savings results assume 80% data dormancy for online archive, $0.20 kWhr***

StorageScaler 6000 Enclosure

- 1 x 60 drive or 2 x 30 drive channel options
- 1 Passive Baseboard
- 8 active SAS expander cards (4- “A” & 4 “B”)
  - Drive Expander Modules (DEMs)
  - Groups of 15 drives
  - Located in the middle of the enclosure drive section.
  - Top removable
- IO modules are SBB compliant and plug into the rear of the enclosure.
- Redundant Power Supplies
  - Hot-swappable
  - Plug into the rear of the enclosure
  - Provides system cooling
- Optional internal flash drive (under consideration)
  - Faster, persistent LUN for file system journaling



StorageScaler 6000 Enclosure

- **Power Cycling Capabilities**
  - Increase System Reliability
  - Reduce Drive Replacements & Rebuilds
    - Not all unresponsive drives are dead drives
    - S2A9900 performs a series of recovery techniques including command retries & drive resets
    - If unsuccessful, enclosure will have ability to *power cycle individual drives* to confirm the status of the specific device.
    - Capability complimented by journaled rebuild capability
      - Drives are back online in minutes
    - If the device cannot be revived it can be replaced online.
    - Reduce RMAs – ***No more "NO Trouble Found RMAs"***

SELF HEALING TECHNOLOGY FOR MAXIMUM UPTIME

Agenda

- S2A Parallel Storage Architecture
- S2A9900 StorageScaler
- **S2A6620**
- DDN HPCSS Lustre Offering
- Future Requirements

DataDirect Networks S2A6620 Appliance

Modular Storage Optimized for IOPS and Density Applications

- **Up to 30,000 IOPS (to disk)**
- **4 x Active/Active Host Ports: FC4, FC8**
- **Scales to Support 120 Hard Drives in 8U**
- **Up to 2.0 GB/s Performance**
- **Mix SAS + SATA For Storage Tiering**
- **Up to 11 Systems (660 TB) per Rack**
- **RAID 5 and RAID 6 Options**
- **Journaled Fast Drive Rebuild**
- **Active/Active Storage Managers with Failover**
- **Full SATAssure Data Protection**

Shipping 2H08

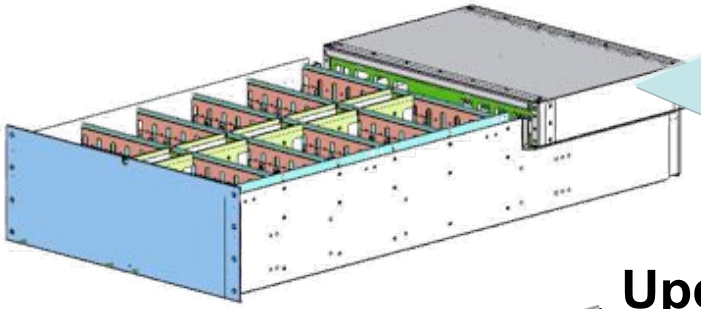
Industry-Leading Extreme Density

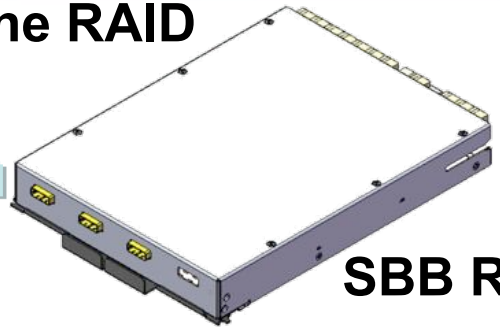
DataDirect NETWORKS S2A

60 Terabytes Per Drawer, 660TB Per Single Rack

StorageScaler Migration

Begin – Standalone RAID



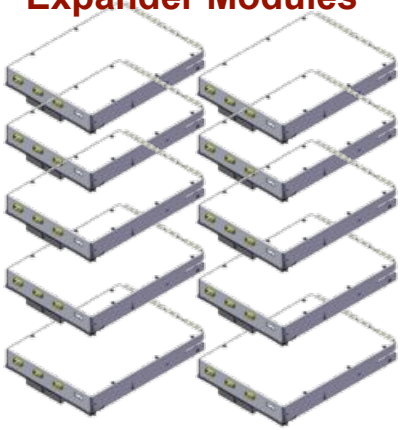


SBB RAID Module

Upgrade to….

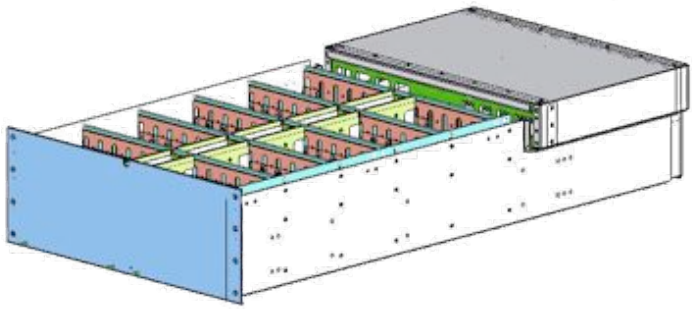
S2A9900

Replace SBB RAID Modules with SBB SAS Expander Modules



Future Storage Server w/ RAID

Replace SBB RAID Modules with SBB Storage Server Modules



Agenda

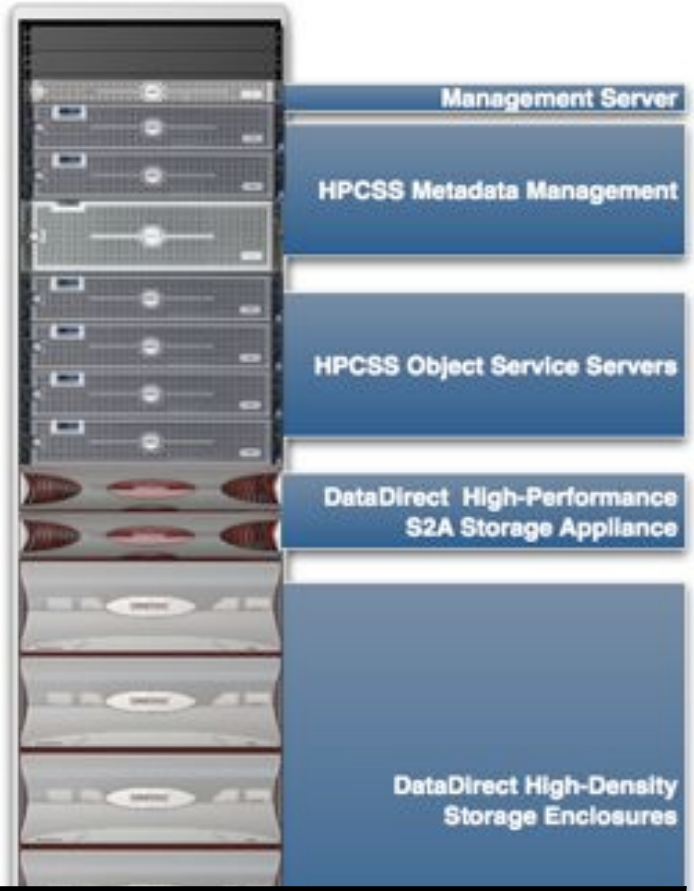
- S2A Parallel Storage Architecture
- S2A9900 StorageScaler
- S2A6620
- **DDN HPCSS Lustre Offering**
- Future Requirements

High Performance Cluster Storage Solution (HPCSS)

DataDirect NETWORKS S2A

- S2A technology enables Lustre scaling
- Linearly scalable (from 2.5GB/s to 200GB/s+)
- Single client throughput of 1GB/s+
- Fault-tolerant architecture
- Ideal for cluster & grid computing
- DDN developed tools for easy installation

Scale Performance Linearly by Adding OSS Building Blocks



Optionally sold with:

- S2A9900 Storage System
- S2A9700 & S2A9550 Storage Systems

High Performance Cluster Storage Solution (HPCSS)

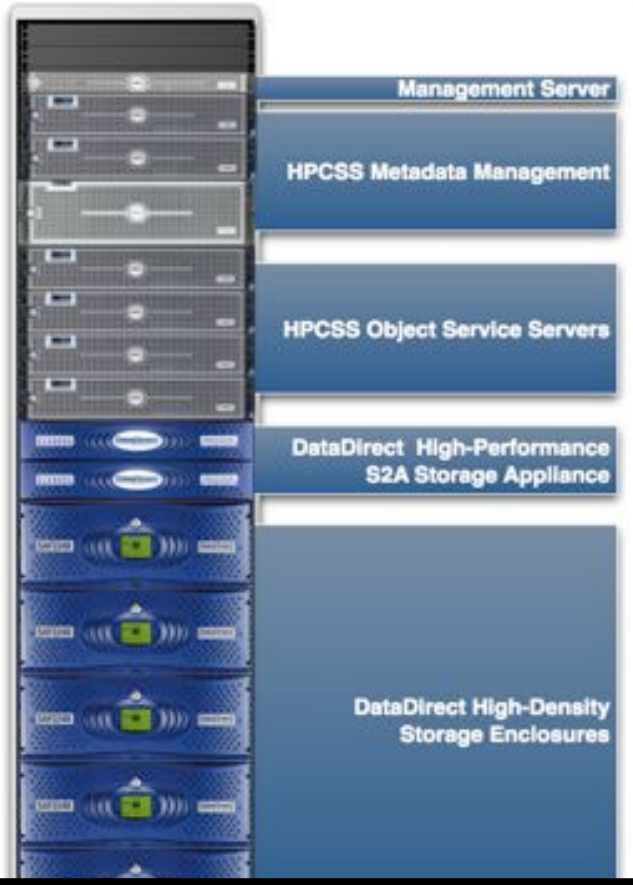
DataDirect NETWORKS S2A

- **Full System Performance Efficiency**
  - Results using S2A9550 (4OSSs; DDR IB LAN)
  - Recent IOZone Summary:

api = POSIX
access = file-per-process
ordering = sequential offsets
clients = 64 (1 per node)
repetitions = 3
xfersize = 1 MiB
blocksize = 32 GiB
aggregate filesize = 2048 GiB

access	bw(MiB/s)	block(KiB)	xfer(KiB)	open(s)	wr/rd(s)	close(s)	iter
write	2486.28	33554432	1024.00	7.44	836.06	73.76	0
read	2649.34	33554432	1024.00	0.005301	791.57	42.81	0
write	2443.70	33554432	1024.00	4.41	853.78	60.02	1
read	2672.79	33554432	1024.00	0.004316	784.63	43.50	1
write	2400.73	33554432	1024.00	12.60	860.96	66.45	2
read	2678.90	33554432	1024.00	0.005354	782.84	40.55	2

Max Write: 2486.28 MiB/sec (2607.05 MB/sec)
Max Read: 2678.90 MiB/sec (2809.03 MB/sec)
Using only 1 x Object Storage Server Building Block



Optionally sold with:

- S2A9900 Storage System
- S2A9700 & S2A9550 Storage Systems

Sample Lustre + DDN Customers

DataDirect NETWORKS S2A

Sandia National Laboratories

OAK RIDGE NATIONAL LABORATORY

cea

SYNOPSYS® Predictable Success

AWE

NCSA

MAUI HIGH PERFORMANCE COMPUTING CENTER

LSU

INDIANA UNIVERSITY

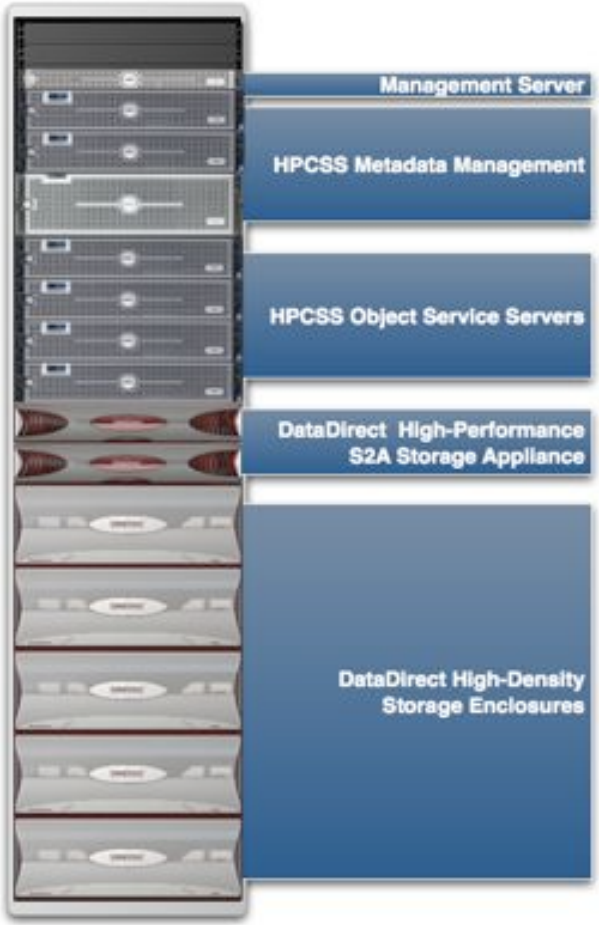
Lawrence Livermore National Laboratory

CLEMSON

ERSC

NASA

PSC



And many more…

Agenda

- S2A Parallel Storage Architecture
- S2A9900 StorageScaler
- S2A6620
- DDN HPCSS Lustre Offering
- **Future Requirements**

Future Requirements

DataDirect NETWORKS S2A

Storage Challenges

- Data transfer rates will range to TBs/s
- Drive transfer rates will not exceed 120 MB/s
- Average seek times for SAS will remain at 3mS
- Average seek times for SATA will remain at 11ms
- Any random activity greatly diminishes the effective transfer rate

Future Solutions

Evolving Technology

- **Faster physical transfer architectures such as IB 32x**
- **File systems with better transfer aggregation**
  - Lustre at 2MB? 4MB?
- **Storage integrated with file services to enable intelligent data transfer reordering**
- **Storage elements are getting faster, better, cheaper, and lower in power consumption**
  - SSDs are larger and more reliable and can be utilized in the same architecture
  - Smaller form factor disks are larger, cheaper, and more reliable
  - SRAM costs are decreasing with finer pitch implementations