DataDirect NETWORKS

S2A TECHNOLOGY ADVANTAGE

Exploring Mass Storage Concepts to Support Exascale Architectures

Cray User Group **May, 2009**

Dave Fellinger
CTO

DDN = HPC

- **DDN provides more bandwidth to the top500 list than all other vendors combined!**
- **8 out of Top10 systems choose DDN**
- **50 out of Top100 and more**
- **5 systems over 120GB/s**
  - top end = 3 x faster than rivals
- **Mix of applications:**
  - Government/University
  - Defense/Intelligence
  - Oil Exploration
  - Product Design
  - Archival, Backup

Rank	Site	Computer
1	DOE/NNSA/LANL United States	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 G Opteron DC 1.8 GHz , Voltaire Infiniband IBM
2	Oak Ridge National Laboratory United States	Jaguar - Cray XT5 QC 2.3 GHz Cray Inc.
3	NASA/Ames Research Center/NAS United States	Pleiades - SGI Altix ICE 8200EX, Xeon QC 3.0/2.66 G SGI
4	DOE/NNSA/LLNL United States	BlueGene/L - eServer Blue Gene Solution IBM
5	Argonne National Laboratory United States	Blue Gene/P Solution IBM
6	Texas Advanced Computing Center/Univ. of Texas United States	Ranger - SunBlade x6420, Opteron QC 2.3 Ghz, Infi Sun Microsystems
7	NERSC/LBNL United States	Franklin - Cray XT4 QuadCore 2.3 GHz Cray Inc.
8	Oak Ridge National Laboratory United States	Jaguar - Cray XT4 QuadCore 2.1 GHz Cray Inc.
9	NNSA/Sandia National Laboratories United States	Red Storm - Sandia/ Cray Red Storm, XT3/4, 2.4/2.2 Cray Inc.
10	Shanghai Supercomputer Center China	Dawning 5000A - Dawning 5000A, QC Opteron 1.9 Ghz, Infiniband, HPC 2008 Dawning

Petascale Storage Blueprint

- **World's First PFlop System Without Accelerators**
- **240GB/s Site-Wide File System: "Spider"**
  - ~2x the other fastest at that time: CEA & LLNL – Update… LLNL is catching up!
- **Uses 48 x S2A9900 Storage Systems**
  - DDR IB-Connected Arrays
  - Over 13.4K HDDs, 10PB Usable
- **Enables Site-Wide Scalable File System with High QoS**
  - One file system for all clusters
  - Supports 98,400 CPUs
- **Selection based on extensive storage bakeoff vs. Competition**
  - Storage Energy Consumption
  - Selected on Mixed I/O Capabilities More So than Sequential

DataDirect NETWORKS

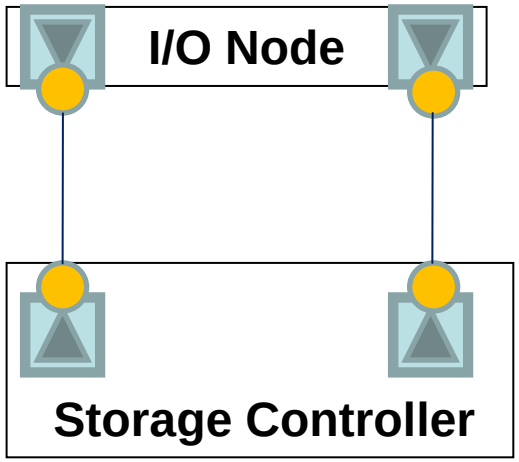
S2A TECHNOLOGY ADVANTAGE

Storage Fusion Architecture
The Next Generation

SFA Appliance Architecture

System Simplification/Cost Reduction

Today

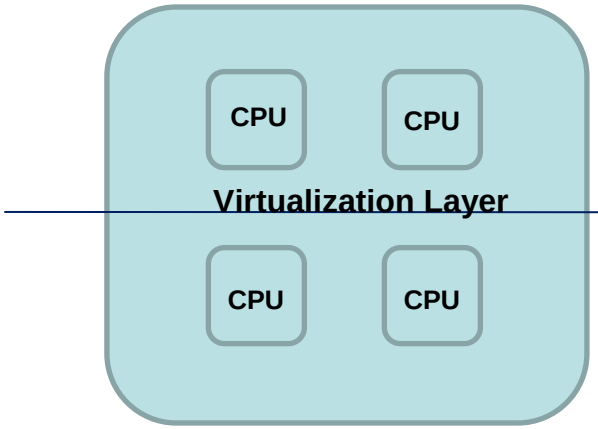


HBA/HCA

SFP/XENPAK/XFP

Cable

SFA



- All I/O Node to Storage Controller interconnect components eliminated.
- Protocol overhead and conversions reduced
- Shared memory space
- High-speed internal connections

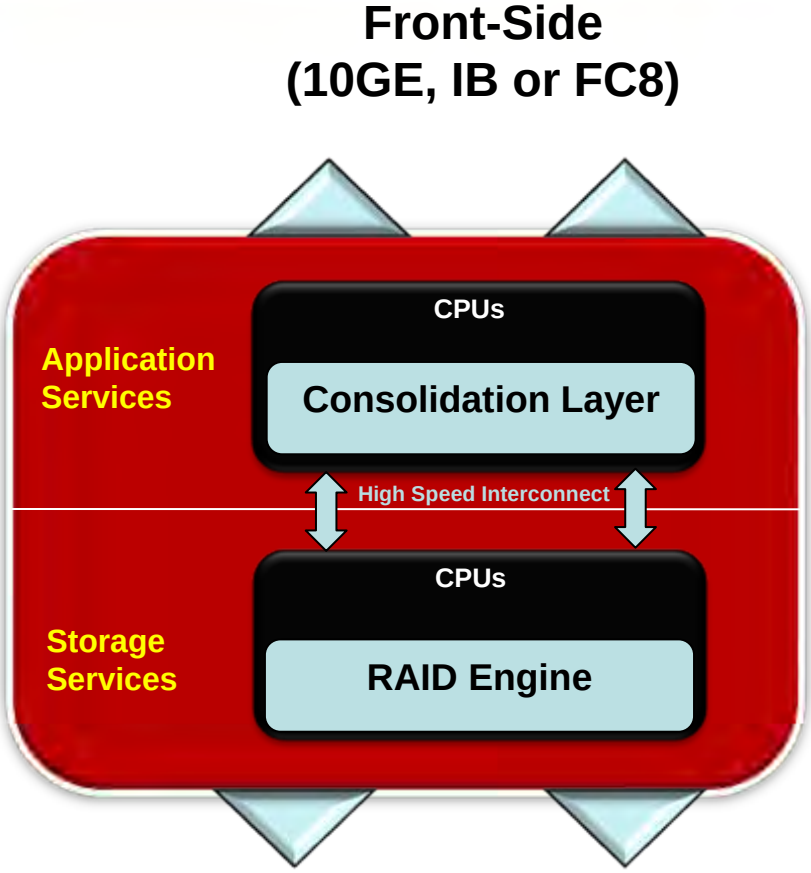
SFA Architecture

Features

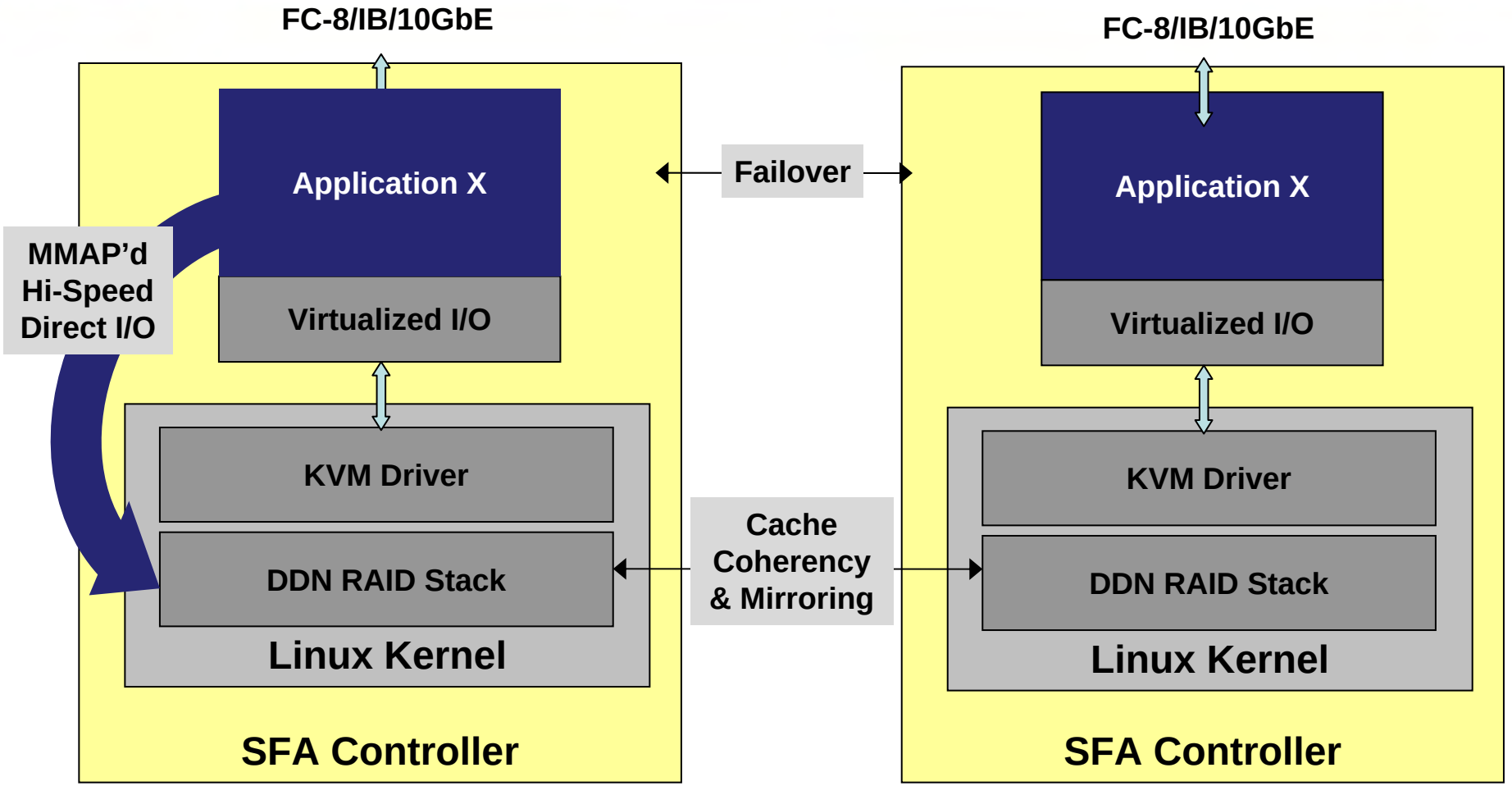
- Dedicated resources provided to Storage Services and Application Services
- High-speed internal connections and shared memory architecture
- Protocol conversions eliminated
- Massive and balanced front-side and back-end bandwidth

Benefits

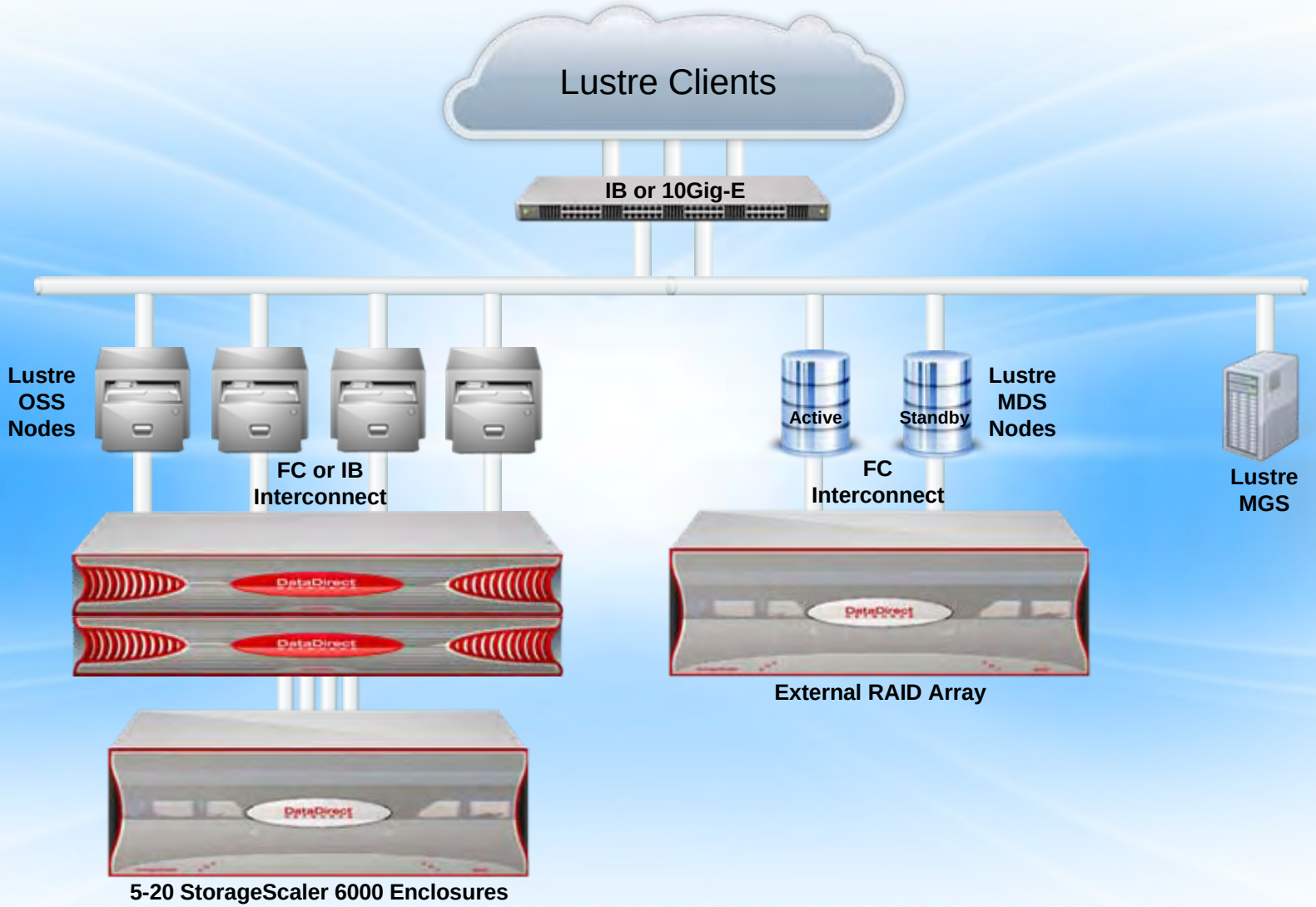
- High performance bandwidth ***and*** IOPS
- Stable performance for both Applications and Storage Services
- Reduced latency between application servers and storage
- Reduction in infrastructure and complexity
- Reduced number of individual storage systems required to scale capacity



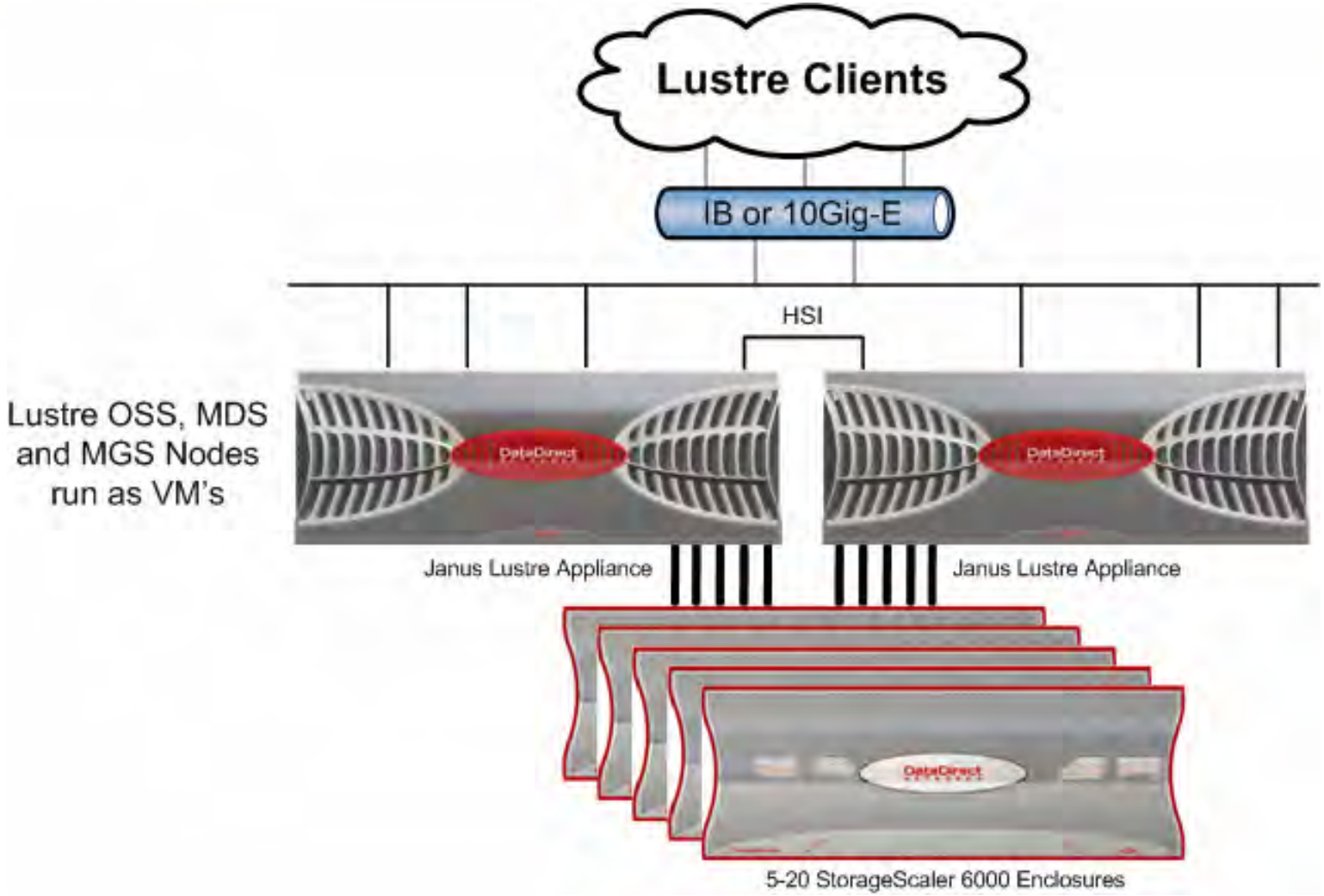
SFA Application Platform



Example: Today's State-of-the-art Lustre HPC Solution



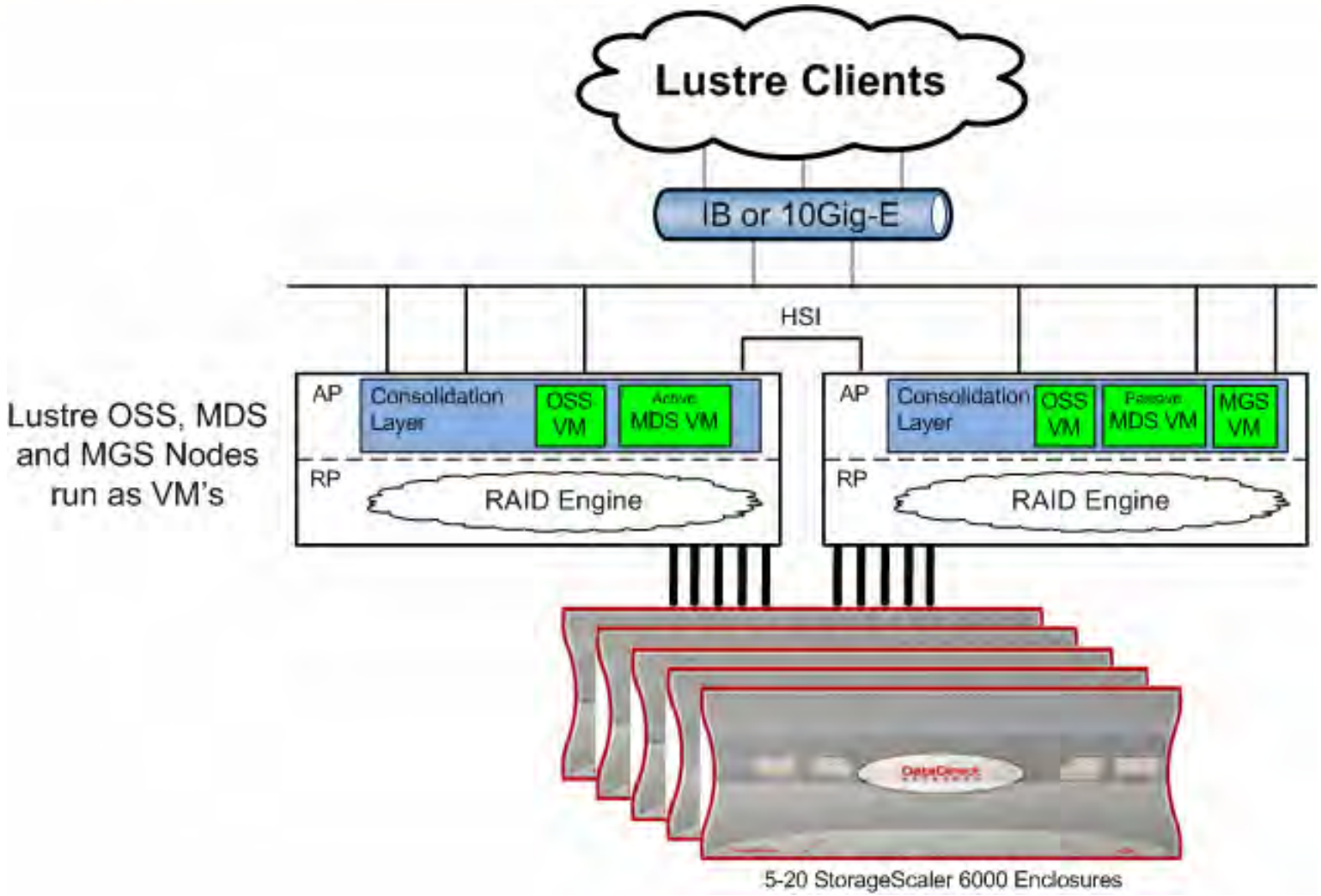
Example: Lustre HPC Storage with the SFA Platform



Example: Lustre HPC Storage with the SFA Platform

DataDirect NETWORKS

S2A TECHNOLOGY ADVANTAGE



Example: Lustre HPC Storage with the SFA Platform

5:1 Reduction in Active Managed Components!!

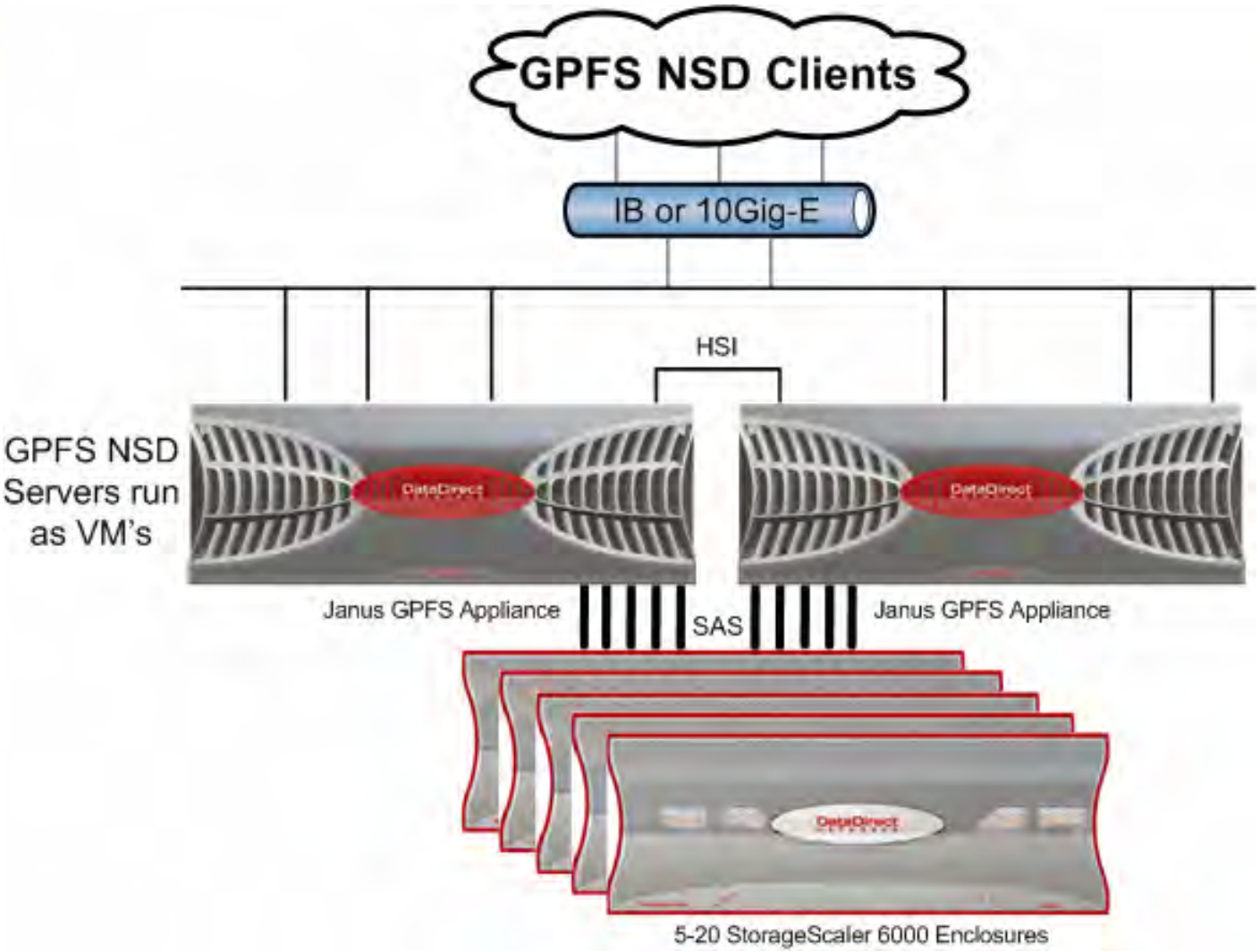
Reduced Management, power, cooling, floor space

SFA Lustre Solution

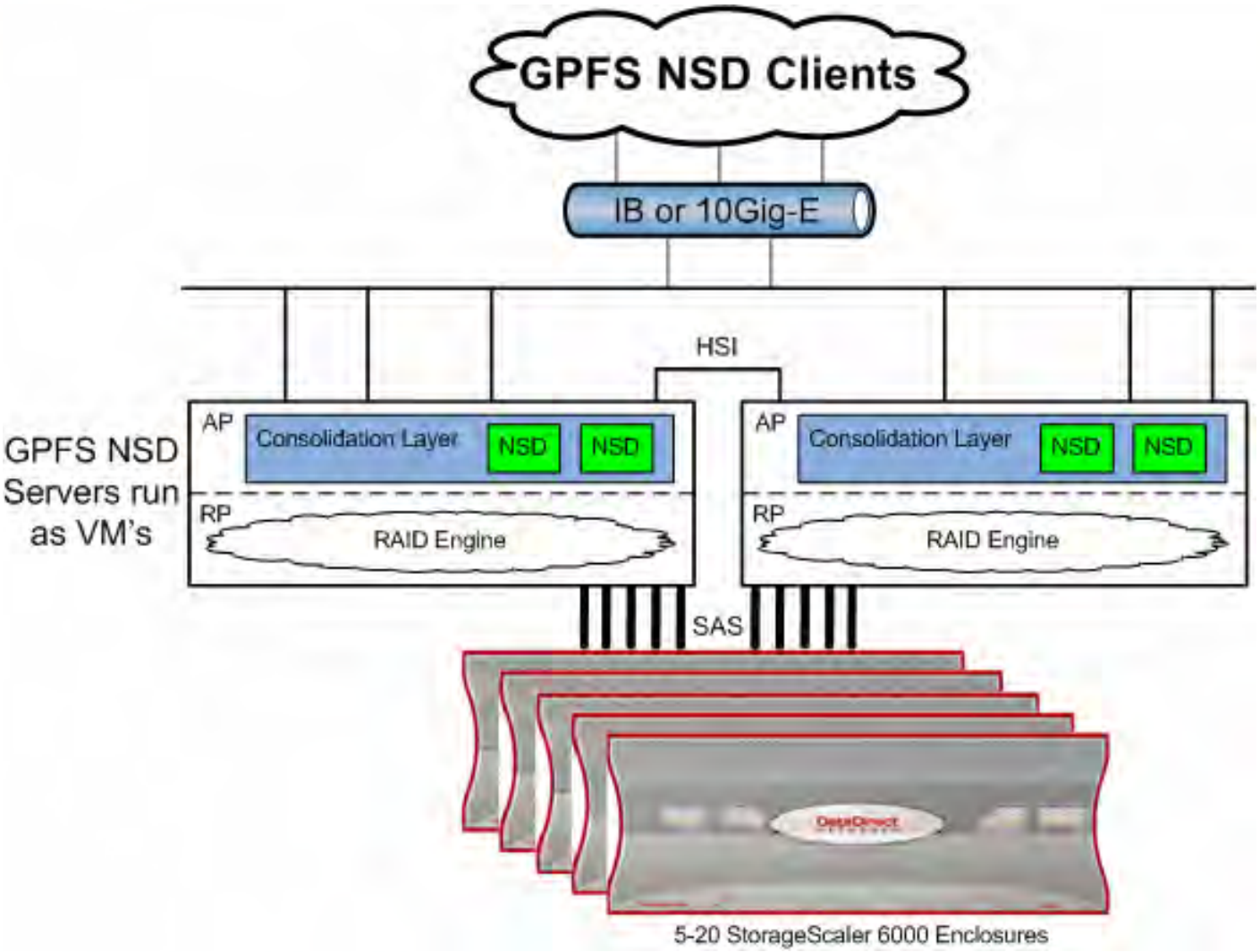
- **Today:**
  - I/O architecture of existing OSS server platform limits scalability of individual nodes
  - Large numbers of systems necessary to scale performance
  - Large numbers of OSSs potentially require a large FC switch infrastructure investment
  - Installation and administration of Lustre is difficult
  - Lustre failover is difficult to set up and to have work properly

- **SFA:**
  - **SFA hardware is optimized for I/O. Advanced software technology ensures maximum performance**
  - **Greatly reduces the number of active elements – as much as 5 to 1!**
  - **SFA eliminates the need for Fiber Channel infrastructure, significantly reducing overall cost & complexity**
  - **SFA Lustre Appliances will come with Lustre pre-installed**
  - **Lustre failover will be a configurable option within a SFA Lustre Appliance couplet**

Example: GPFS HPC Storage with the SFA Platform

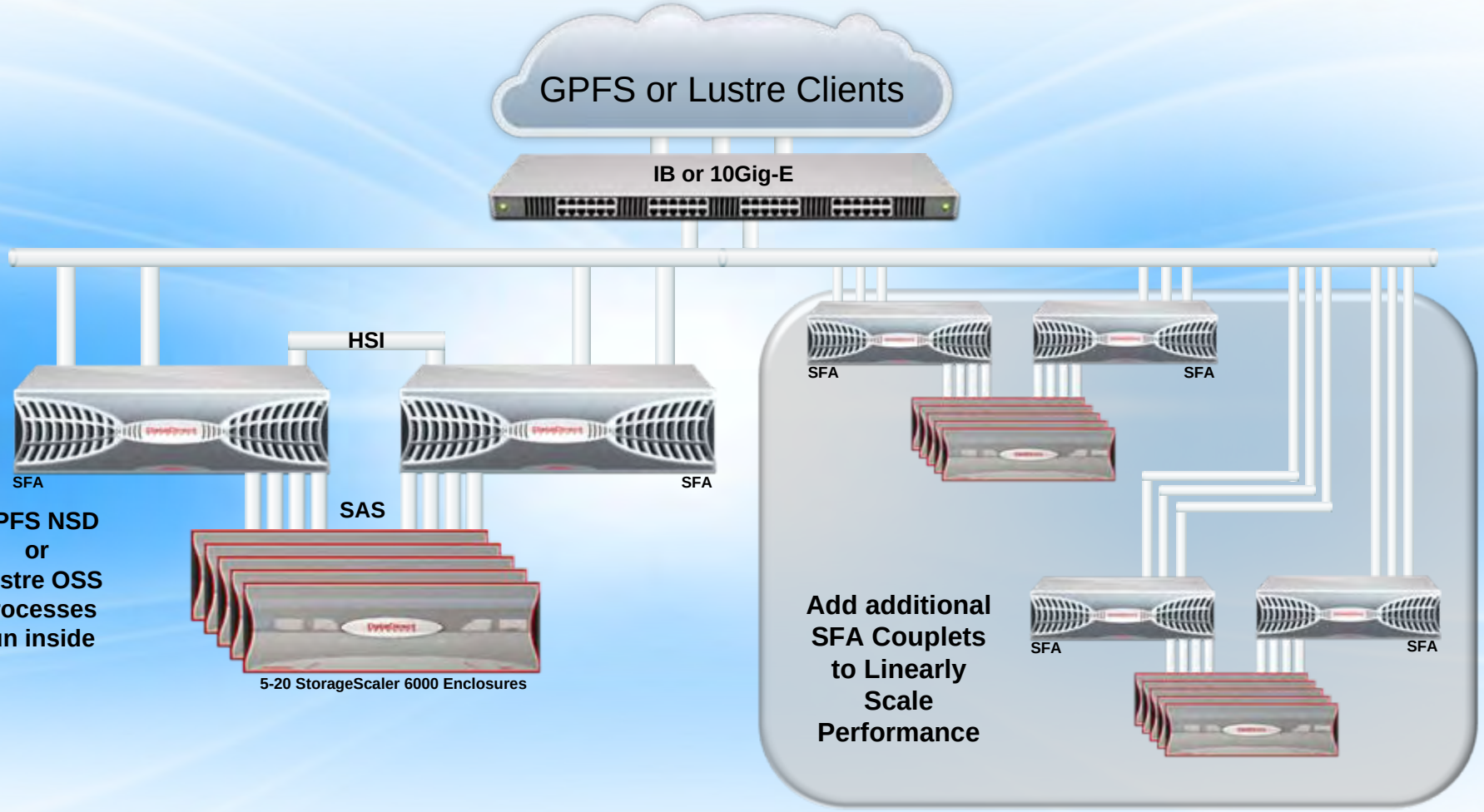


Example: GPFS HPC Storage with the SFA Platform



Example: GPFS HPC Storage with the SFA Platform

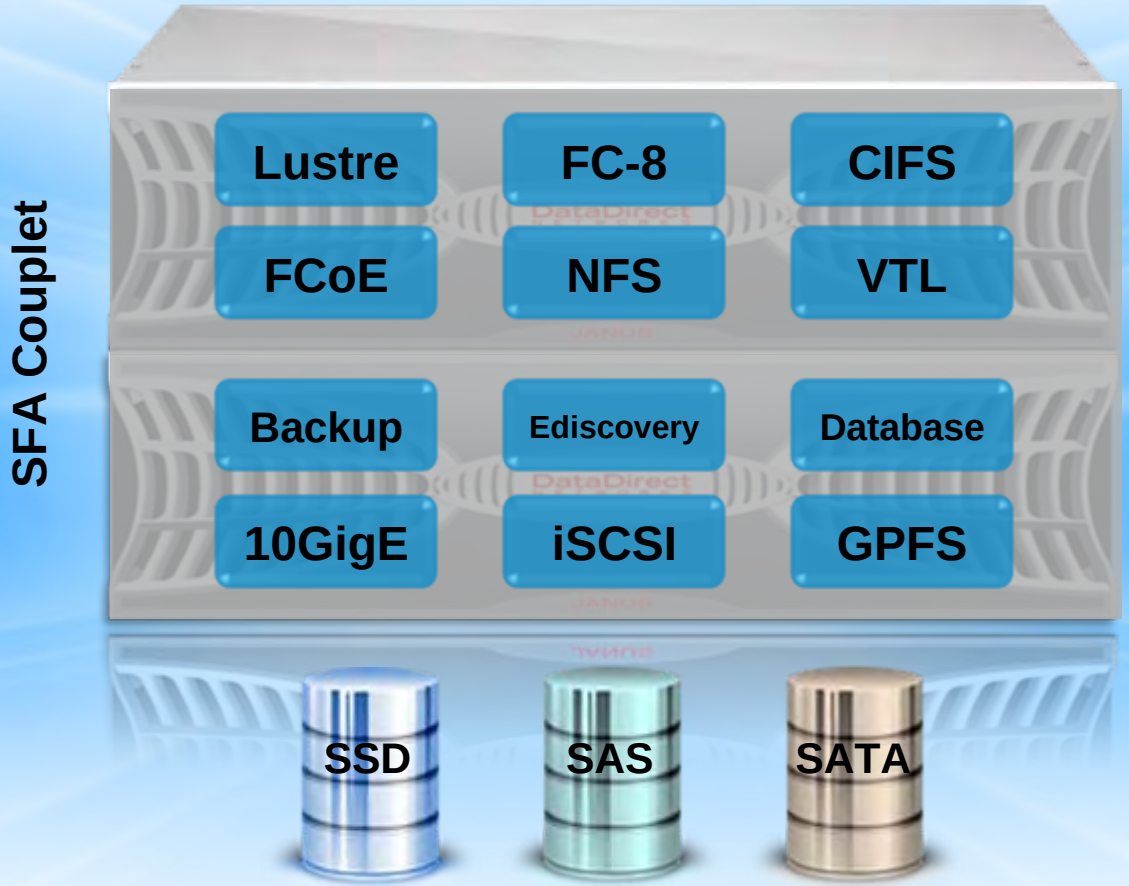
Scaling Performance with the SFA Platform



A Truly Flexible Storage and Application Platform

DataDirect NETWORKS S2A TECHNOLOGY ADVANTAGE

Designed for Maximum Storage Application Freedom



Freedom to Mix Physical Storage Tiers

Multi-Platform Architecture

DataDirect NETWORKS S2A TECHNOLOGY ADVANTAGE

Storage Array

Block Storage Target

Fibre Channel
Infiniband
iSCSI

Clustered Filer

DDN File Storage [Lustre, NAS, VTL, etc]

Block Storage Target

Open Appliance

Customer Applications

Storage Server Virtualization

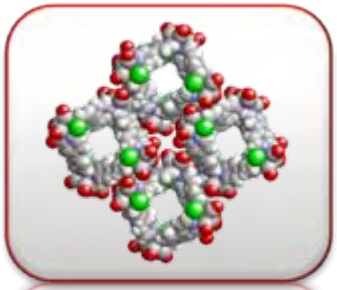
Block Storage Target

Flexible Deployment Options: 3 System Modalities

SFA Enhances Many Classes of Application

DataDirect NETWORKS

S2A TECHNOLOGY ADVANTAGE



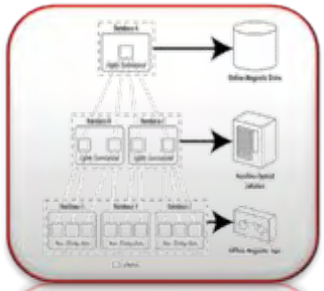
HPC

Rendering & Animation

Information Services

Database

Data Warehouse



Data Management

Data Protection

SFA Benefits

- **Extreme Performance**
  - Increased Application Performance
  - Mixed Workload Capable
  - 10 GB/s of throughput per Couplet
  - 300k IOPS (Burst to Disk) per Couplet
  - 1M IOPS (Burst to Cache) per Couplet

- **Consolidation**
  - Reduced infrastructure to manage and lower administrative overhead
  - Lower power, space and cooling requirements
    - Up to 5 times reduction
  - Density: Up to 2.4PB in two racks using 2TB drives
    - Consolidate multiple arrays into one
  - Lower TCO

The SFA10000

DataDirect NETWORKS

	SFA10K
General Availability	Q3 2009
Hosted Applications & Application Resources	8 Cores 16GB FS Cache
File Storage Ports	QDR IB, 10GbE
Host Port Options	16 x FC8 8 x QDR IB
Throughput (block)	10GB/s
IOPs (block)	1M (cache) 300,000 (disk)
Max Spindles	1,200 (600/rack)

DataDirect NETWORKS

S2A TECHNOLOGY ADVANTAGE

Future Requirements

Future Requirements

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 Technology

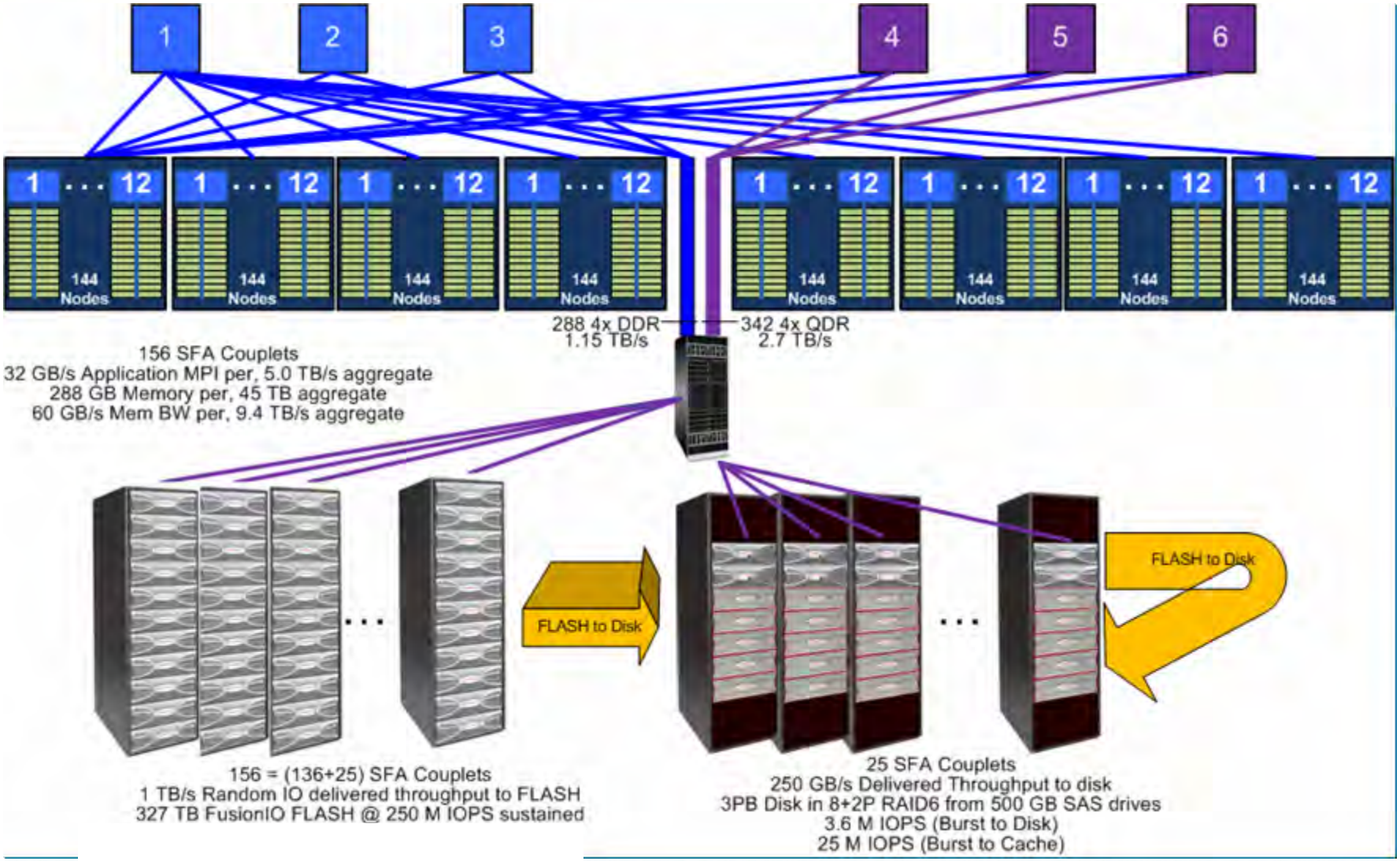
Evolving Technology

- **Faster physical transfer architectures such as IB 32x**
- **File systems with better transfer aggregation**
  - Lustre and GPFS @ 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

Future Solutions

- Systems must be kept as small and power efficient as possible
- SCSI Layers must be minimized
- SSD technology must be utilized in conjunction with rotating media to execute a short term HSM
- AI must be used to simplify management
- File system service must be a part of the storage system
- The storage system must be capable of data analysis (reduction, mining, runtime analysis)

Implementation Example



DataDirect NETWORKS

S2A TECHNOLOGY ADVANTAGE

Thank You

Dave Fellinger
dfellinger@ddn.com