Agenda

• Who’s DDN?
• S2A Architecture
• SFA Architecture
• WOS: Web Object Storage
The Worldwide Scalability Leader

Enable Organizations to Maximize the Value of All Information Everywhere

<table>
<thead>
<tr>
<th>The DDN Mission</th>
<th>The Worldwide Scalability Leader</th>
</tr>
</thead>
<tbody>
<tr>
<td>Established</td>
<td>1998</td>
</tr>
<tr>
<td>Ownership</td>
<td>Privately-Held, Self-Funded</td>
</tr>
<tr>
<td>Revenue</td>
<td>Over $200M Annually</td>
</tr>
<tr>
<td>Profitability</td>
<td>Consistently Profitable Since 2002</td>
</tr>
<tr>
<td>Growth</td>
<td>30% Annual Growth (’09-’10), about 400 employees</td>
</tr>
<tr>
<td>Presence</td>
<td>4 Continents, Located in 18 Countries</td>
</tr>
<tr>
<td>Markets</td>
<td>Content &amp; Cloud, HPC, BioTech, Intelligence, Surveillance</td>
</tr>
<tr>
<td>Recognition</td>
<td>Frost &amp; Sullivan Best Storage for Digital Media</td>
</tr>
<tr>
<td></td>
<td>World’s Largest Private Storage Company (IDC ‘11)</td>
</tr>
<tr>
<td></td>
<td>Deloitte Fast500 Technology Company (‘10)</td>
</tr>
<tr>
<td></td>
<td>Inc. Magazine 500</td>
</tr>
<tr>
<td></td>
<td>Frost &amp; Sullivan Best Practice for Video Surveillance</td>
</tr>
<tr>
<td></td>
<td>HPCWire Best HPC Storage Product (6 Yrs. Running)</td>
</tr>
</tbody>
</table>
DDN = HPC

- 6 out of Top10
- 15 out of Top20
- 56 out of Top100
- 122 out of Top500
- **13 Petaflops computing powered**
- **5 systems over 120 GB/s**
- DDN provides more bandwidth (> 2TB/s) to the top500 list than all other vendors combined!
DDN is the leading provider of affordable, high-availability storage for the next generation of particle physics research.

DDN Supplied Over 30PB of LHC Storage in the last 3 years
The Worldwide Scalability Leader

140,000
# of Supercomputer CPUs
World’s Fastest File System

23,000,000
Online Users Served
Xbox Live Community

5,000,000,000
Individual Photos
~35 PBs of Storage

Drawing From Leadership Development Experience To Scale Business Drivers
Sample HPC Partners & Customers
DDN has delivered solutions to over 600 of the world's largest media organizations.
S2A & SFA Architecture
Product Portfolio

Array Platforms

File Storage

Cloud Storage

Supporting SATA, SAS and SSD Disks

Featuring:
Leading Scalability • Highest Efficiency • Fastest ROI

© 2011 DataDirect Networks. All rights reserved.
S2A9900
Real-Time Content Storage
An Implementation of Parallelism w/ Double Parity RAID Protection

- Double Disk Failure Protection
- LUNs can span tiers
- All ports access all storage
- Reed-Solomon Code Implemented in a 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, SSD, SAS, SATA Disks
- Up to 1200 disks total
  - 960 formattable disks

8 FC-8 and/or 4 IB 4X Parallel Host Ports

RAID 6, 8+2 Byte Stripe

© 2011 DataDirect Networks. All rights reserved.
Data Corruption Error Handling

### Host Data Striping

- **First step isolates error**
- **Second step corrects error**

### FPGA

### Cache

### Protocol/PHY

### Disks

- SCSI
- FC - or - SAS
The data is flushed to the disk and the disks have now correct data on channel F.
The cache has been repaired by the FPGA using the parity information.

Host Data Striping

FPGA

Cache

Protocol/PHY

Disks

SCSI
FC - or - SAS
Supported Enclosures

60 x 3.5” drives in 4U
SSD, SAS, SATA

16 x 2.5” drives in 3U
SSD, SAS
Simple, Reliable Configuration

Direct Connection and RAID Striping Provides Maximum Data Availability

- Direct cabling avoids daisy chaining
- Data is striped across channels/enclosures
- Drive Channels are RAIDed 8+2
- Drive Enclosures are RAIDed 8+2

Only DDN Enclosure RAIDing can withstand the loss of 20% of system enclosures & drives while delivering full data availability!!
Scalability & Density

The World Scalability & Density Leader

- 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.

<table>
<thead>
<tr>
<th></th>
<th>5 Enclosures</th>
<th>10 Enclosures</th>
<th>20 Enclosures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24U: 1/2 Rack</td>
<td>44U: 1 Rack</td>
<td>84U: 2 Racks</td>
</tr>
<tr>
<td>Drives</td>
<td>Up to 300</td>
<td>Up to 600</td>
<td>Up to 1,200</td>
</tr>
<tr>
<td>Storage (TB)</td>
<td>Up to 900</td>
<td>Up to 1.8PB</td>
<td>Up to 3.6PB</td>
</tr>
</tbody>
</table>
Transition To SW Platforms: Complete

Previous Design
36-24 mos. spin

Custom HW for Accelerated Storage Processing

The New DDN
< 9 mos. product spin

Full Storage SW Portfolio = Maximum Design Flexibility
Embedded Virtualization to Natively Host Storage Apps
2010+ Petaflop Systems

- LLNL
  - 1TB/sec and 30PB (Lustre)
- Argonne
  - 500GB/sec and 60PB (GPFS, PVFS)
- ORNL
  - 800GB/sec and 30PB (Lustre)
- CEA
  - 500GB/sec (Lustre)
- HLRS
  - 150-300GB/sec
- LRZ
  - 200-400GB/sec
SFA10000

Highly Parallelized SFA Storage Processing Engine

- Active/Active Design
- 1 Million Burst IOPS from 16GB Mirrored, Non-Volatile Cache
- Up to 300K Sustained Random Read Disk IOPS with 1200 SAS 15K Drives
- Up to 600K Sustained Random Read IOPS from SSDs
- 13GB/s Raw Sequential Read & Write Speed
- RAID Levels 1, 5 and 6
- Intelligent Write-Through Striping
- SATAssure Data Protection
- GUI, SNMP, CLI
- 16 x FC-8 ports or 8 x QDR-IB ports

© 2011 DataDirect Networks. All rights reserved.
Sustained Bandwidth

IOR Writes on Exascalar 1.5.0.RC1
SFA10K 1.4.0.7347, 3TB SATA, 5x7000 enclosures, 12 clients
28 x 8+2 128k: W M Re Pools
System Bandwidth Results by Number of OSTs

© 2011 DataDirect Networks. All rights reserved.
SFA10000 Configurations

5 Enclosure System
Up to 300 Drives
2 BBUs, 28U

10 Enclosure System
Up to 600 Drives
2 BBUs, 48U

20 Enclosure System
Up to 1,200 Drives
2 BBUs, 88U

High Availability Drive Channel & Enclosure RAIDing

© 2011 DataDirect Networks. All rights reserved.
Dynamic Workload Arrays: Roadmap

SFA OS Performance Roadmap

- 6620 (Present)
- 10KT (Q3 2010)
- 12K (Q1 2011)
- 12KT (Q2 2011)
- High-End (Q3 2011)
- Midrange (Q4 2011)
- Q1 2012
- Q2 2012

© 2011 DataDirect Networks. All rights reserved.
Scaling Performance with the SFA12K

GridScaler & ExaScaler Clients
IB or 10Gig-E

9.6PB

SFA12K
300-600 3.5" Disk Drives

Add additional SFA Couplets to Linearly Scale Performance

Integrate multiple appliances to scale to over hundreds of GB/s and 10’s of Petabytes

25GB/s
50GB/s
75GB/s
100GB/s

2.4PB
4.8PB
7.2PB
9.6PB

© 2011 DataDirect Networks. All rights reserved.
SFA10000E Features

- Low Latency Embedded Storage Application Platform
- Active/Active Design
- 8 Application CPU Cores
- 90GB of Application RAM
- 16 x 10Gb Ethernet or 16 x QDR InfiniBand Host Ports
- 6.5 GB/s Read & Write Speed
- 500,000+ Burst IOPS
- 150K Random Disk IOPS
- 16GB Mirrored Cache
- RAID Levels 1, 5 and 6
- Intelligent Block Striping
- Up to 600 SAS, SATA or SSD Drives

© 2011 DataDirect Networks. All rights reserved.
Embedded Services Eliminate Communication Overhead

6KB

Communication per traditional I/O Transfer

- 4KB I/Os = 10KB of Communication
- 32KB I/Os Become 20% Less Efficient
- Accelerated Through Memory Copy, Eliminating SCSI Transfer
IO Path Acceleration

Storage Fusion Architecture shortens the IO path from the application to storage, reducing latency and increasing IOPS performance.
Embedded Applications

QDR IB/10GbE

Applications, File Systems, Database, etc.

Native PCI-e Drivers

High Speed I/O
Virtualization Hypervisor

DDN RAID Stack

Real-Time Storage OS

SFA Controller

Failover

QDR IB/10GbE

Applications, File Systems, Database, etc.

Native PCI-e Drivers

MMAP’d Hi-Speed Direct Disk I/O

High Speed I/O
Virtualization Hypervisor

DDN RAID Stack

Real-Time Storage OS

SFA Controller

Cache Coherency & Mirroring High-Speed Interconnect

MMAP’d Hi-Speed Direct Disk I/O

Native PCI-e Drivers

QDR IB/10GbE

Applications, File Systems, Database, etc.

Native PCI-e Drivers

MMAP’d Hi-Speed Direct Disk I/O

High Speed I/O
Virtualization Hypervisor

DDN RAID Stack

Real-Time Storage OS

SFA Controller
SFA10000E Appliances

- Reduce complexity and Cost
- Increase performance for latency sensitive applications
- SFA10000E initially available with DataDirect Networks’ parallel clustered file system solutions

ExaScaler
SFA10000E
6.5GB/s
Up To 900TB

GridScaler
SFA10000E
6.5GB/s
Up To 1.8PB
Multi-Platform Architecture

Block Storage Array

- SFA10K
- Block Storage Target

Clustered Filer

- SFA10KE
- DDN File Storage
- EXAScaler
- GridScaler

Open Appliance

- SFA10KE
- Customer Applications
- SFA10KE
- Embedded Storage Server
- SFA10K
- Block Storage Target

Product Evolution

Flexible Deployment Options: 3 System Modalities
Distributed Hyperscale
Collaborative Storage

Web Object Storage
The Big Data Reality

Information universe in 2009:
- 800 Exabytes

In 2020’s:
- 35 Zettabytes

A new type of data is driving this growth

- Structured data - Relational tables or arrays
- Unstructured data – All other human generated data
- Machine-Generated Data - growing as fast as Moore’s Law
A Paradigm Shift is Needed

File storage
- Millions of Files
- Point to Point, Local
- Fault-Tolerant
- Files, Extent Lists
- 75% on average

Object Storage
- Scalability
- Access
- Management
- Information
- Space Utilization

- 100’s of Billions of Objects
- Peer to Peer, Global
- Self-Healing, Autonomous
- Objects w/ Metadata
- Near 100%
What Big Data Needs

- Hyper-scale
  - World-wide single & simple namespace
  - Dense, efficient & green
  - High performance versatile on-ramp and off-ramp

- Geographically distributed
  - Process the data close to where it's generated vs. copying vast amount of data to processing
  - Cloud enabling
  - World-wide single & simple namespace

- Resiliency with extremely low TCO
  - No complexity
  - Near zero administration

- Ubiquitous Access
  - Legacy protocols
  - Web Access
Storage should improve collaboration

- ... *Not make it harder*

- Minutes to install, not hours
- Milliseconds to retrieve data, not seconds
- Replication built in, not added on
- Instantaneous recovery from disk failure, not days
- Built in data integrity, not silent data corruption
• Understand the data usage model in a collaborative environment where immutable data is shared and studied.

• A simplified data access system with minimal layers.

• Eliminate the concept of FAT and extent lists.

• Reduce the instruction set to PUT, GET, & DELETE.

• Add the concept of locality based on latency to data.
WOS Fundamentals

» No central metadata storage, distributed management
» Self-managed, online growth & balancing, replication
» Self-tuning, zero-intervention storage
» Self-healing to resolve all problems & failures with rapid recovery
» Single-Pane-of-Glass global, petabyte storage management
A file is uploaded to the application or web server.

The system then replicates the data according to the WOS policy, in this case the file is replicated to Zone 2.

A user needs to retrieve a file.

The WOS client returns a unique Object ID which the application stores in lieu of a file path. The application registers this OID with the content database.

The WOS client automatically determines what nodes have the requested object, retrieves the object from the lowest latency source, and rapidly returns it to the application.

Application makes a call to the WOS client to store (PUT) a new object.

OID = 5718a36143521602

Application returns file to user.

Application makes a call to the WOS client to read (GET) the object. The unique Object ID is passed to the WOS client.

A user needs to retrieve a file.

LAN/WAN

The WOS client automatically determines what nodes have the requested object, retrieves the object from the lowest latency source, and rapidly returns it to the application.

The WOS client returns a unique Object ID which the application stores in lieu of a file path. The application registers this OID with the content database.
Intelligent WOS Objects

Sample Object ID (OID): ACuoBkmWW3Uw1W2TmVYthA

- **Signature**: A random 64-bit key to prevent unauthorized access to WOS objects
- **Policy**: Eg. Replicate Twice; Zone 1 & 3
- **Checksum**: Robust 64 bit checksum to verify data integrity during every read.
- **User Metadata**: Object = Photo, Tag = Beach

© 2011 DataDirect Networks. All rights reserved
WOS Advantages
Simple Administration

• Designed with a simple, easy-to-use GUI
• “This feels like an Apple product”
  » Early customer quote
WOS Deployment & Provisioning

WOS building blocks are easy to deploy & provision – in 10 minutes or less

» Provide power & network for the WOS Node
» Assign IP address to WOS Node & specify cluster name (“Acme WOS 1”)
» Go to WOS Admin UI. WOS Node appears in “Pending Nodes” List for that cluster
» Drag & Drop the node into the desired zone
» Assign replication policy (if needed)

Congratulations! You have just added 180TB to your WOS cluster!
Data Protection: Drive and Node Failure Handling

This slide needs to be viewed in PowerPoint presentation mode. Static display such as editing mode or printed slides will not convey anything meaningful due to the interactive nature of this slide.
**WOS Accessibility**

**NAS Protocols** (CIFS, NFS, etc)
- CIFS/NFS protocols
- Scalable
- HA & DR Protected
- Migration from existing NAS

**Cloud Platform**
- S3 compatibility

**Native Object Store**
- C++, Python, Java, PHP, HTTP REST interfaces
- PUT, GET, DELETE object, RESERVE ObjectID, etc

**NAS Gateway**
- Scalable to multiple gateways
- DR protected & HA Failover
- Synchronized database across remote sites
- Local read & write cache
- LAN or WAN access to WOS
- Federates across WOS & NAS

**Cloud Storage Platform**
- Targeted at cloud service providers or private clouds
- Enables S3-enabled apps to use WOS storage at a fraction of the price
- Supports full multi-tenancy, bill-back, and per-tenant reporting
Get Operation – Corrupted with Repair

1. WOS-Lib selects replica with least latency & sends GET request
2. Node in Zone “San Fran” detects object corruption
3. WOS-Lib finds next nearest copy & retrieves it to the client app
4. In the background, good copy is used to replace corrupted object in San Fran zone
Geographic Replica Distribution

**PUT with Asynchronous Replication**

1. WOSLib selects “shortest-path” node
2. Node in Zone “San Fran” stores 2 copies of object to different disks (nodes)
3. San Fran node returns OID to application
4. Later (ASAP) Cluster asynchronously replicates to New York & London zones
5. Once ACKs are received from New York & London zones, extra copy in San Fran zone is removed

![Diagram of Geographic Replica Distribution]
WOS + IRODS is a simple solution for Cloud Collaboration

- iRODS, a rules oriented distributed data management application meets WOS, an object oriented content scale-out and global distribution system.

- WOS is a flat, addressable, low latency data structure.

- WOS creates a “trusted” environment with automated replication.

- WOS is not an extents based file system with layers of V-nodes and I-nodes.

- IRODS is the ideal complement to WOS allowing multiple client access and an incorporation of an efficient DB for metadata search activities.