Neue Entwicklungen in Lustre profitieren von aktuellen Netzwerkstandards
ZKI –AK Supercomputing 22-23. September Kaiserslautern

Auke Kuiper
DDN – ZKI Autumn presentation
DDN | Agenda

- Introduction
- Solutions for various vertical markets:
  - Lustre benefits from new network standards
- Summary
- Q&A
DDN | Who We Are

We Solve Big Data Lifecycle Management Challenges at Large Scale

- **Markets:** Enterprise Big Data, Cloud, HPC
- **Solutions:** Platforms, File Systems, Object Storage
- **Customers:** 1,000+ in 50 Countries
- **Go-To-Market:** Direct, Partner Assist
- **Employees:** 550 in 20 Countries
- **Headquarters:** Santa Clara, CA
- **History:** Founded in 1998 and Profitable

The World’s Largest Private Storage Company
DDN | Storage Appliance Portfolio

Exa/Grid Scaler or block storage

**SFA12KX**
Bandwidth: 48GB/s
IOPS: 1.7M
Scales to 1680 Drives
In-Storage Processing

**SFA7700**
Bandwidth: 10GB/s
IOPS: 450K
Scales to 396* Drives

**WOS7000**
HD or Archive
4U, 60-Drive System
8 x GbE per Node
2.6PB/Rack, 22B Objects/Rack
30PB/Cluster; 32 Clusters/Namespace

Object storage

* Other names and brands may be claimed as the property of others.
SFA Feature Breakdown

**SFX Storage Cache**
Automated and API-driven caching system to enable hybrid SSD & HDD pools

**QoS Engine**
Real-Time Quality of Service for Read Operations

**DirectProtect™**
Real-Time Data Integrity Verification for every I/O Drive Power Cycling to Eliminate False Positives

**Storage Fusion Fabric**
Highly-Over-Provisioned Backend – Redundant Fabric Withstands Failures of Drives, Enclosures, Cables, etc

**SFA Embedded (Optional)**
Low-Latency KVM Environment with Custom DDN DMA Memory Drivers to Host File Servers and Pre/Post Process Analytics

**Real-Time, Multi-CPU RAID Engine**
Full Implementation of Real-Time Storage OS, Interrupt-Free; Massively Parallel I/O Delivery System

**ReACT Cache Management**
Real-Time I/O Data-Aware Cache Optimizer, Small I/Os Go to Cache; Large & Streaming I/Os Bypass Cache

**High Density Array**
Up to 840 HDDs in a single rack. 84 HDDs and SSDs in 4U

©2014 DataDirect Networks. All Rights Reserved. Any representations around future events are subject to change. * Other names and brands may be claimed as the property of others.
DDN End-to-End Big Data Portfolio

DDN optimizes your HPC storage appliances into most performing and stable solutions.
Lustre developments
(Bulk RPC 4 MB, 16 MB, DNE)
DNE + SSD’s for internal metadata
Multicore strategy in SFA12K + 84 HDD bay enclosures
FDR for IB

DNE = Distributed Name Space to provide better MDS performance and HA
Lustre setup with 12KX40 ExaScaler Scalable Storage Unit (SSU)

4 OSS Servers

2 or 4 MDS Servers

400 x NL-SAS

Up to 20 SSDs

©2014 DataDirect Networks. All Rights Reserved. Any representations around future events are subject to change. * Other names and brands may be claimed as the property of others.
ExaScaler 12K40X Single Rack Scalable Storage Unit (SSU)

- **ExaScaler/Lustre Configuration**
  - 4 x OSS
  - 2 or 4 x MDS

- **Storage Configuration**
  - 12K40X with 5 x SS8460
  - 400 x 4 TB NL-SAS, Up to 20 x SSD for Metadata

- **Software Stack**
  - Server OS: RHEL6.5 or CentOS 6.5
  - OFED: Mellanox OFED-2.1-1.0.6
  - Server Lustre: Lustre-2.5..x_DDN branch
  - Client Lustre: Lustre-2.5..x_DDN branch
  - Lustre 2.6 available for non production systems
12KX40 ExaScaler
Scalable Storage Unit (SSU)

▸ 1,25 PB usable capacity (1,6 PB raw)
  (maximum 1,88 PB usable 2,4 PB raw with 6TB HDD)

▸ 30 - 32GB/s bandwidth on Lustre

▸ > 1 Billion files (inodes)

▸ 1 Rack 34 U + 1U IB Switch

These specifications do not include the rack and the servers.

Power (nominal)

- Power: 4,7 kW (6TB HDD)
- Power: 5,3 kW (4TB HDD)
- Power servers: 3,2 kW (4 OSS + 2 MDS)
- Cooling (nominal): 29500 BTU/hr
- Weight: 800 kg

©2014 DataDirect Networks. All Rights Reserved. Any representations around future events are subject to change.
* Other names and brands may be claimed as the property of others.
Lustre Performance

Write

Throughput (GB/sec)

400 x NLSAS

Read

Throughput (GB/sec)

400 x NLSAS

1MB RPC
Write(Buffered)
1MB RPC
Write(O_DIRECT)
4MB RPC
Write(Buffered)
4MB RPC
Write(O_DIRECT)
16MB RPC
Write(Buffered)
16MB RPC
Write(O_DIRECT)
1MB RPC
Read(Buffered)
1MB RPC
Read(O_DIRECT)
4MB RPC
Read(Buffered)
4MB RPC
Read(O_DIRECT)
16MB RPC
Read(Buffered)
16MB RPC
Read(O_DIRECT)
Summary
DDN | Summary

- Network innovation results in smaller footprint with more performance
- Smaller persistent FS with burst buffer cache implementation
- Smaller FS with near line archive (object store)
- TCO optimization and initial investment optimization
Thank You