SEAGATE

ClusterStor Nytro Intelligent IO Manager

ClusterStor Product Line Overview

Vertically Integrated Like No other: From the RAW media the fastest systems in the world

Lustre Solutions
›Up to 360 GB/s per rack
›Lustre 2.5 / 2.7

lustre™

Lustre Secure
›Up to 60 GB/s per rack
›Lustre 2.5 on SE-Linux

lustre™

Spectrum Scale
›Up to 100 GB/s per rack
›IBM Spectrum Scale 4.x

IBM Spectrum Scale

A200 - Object Store
›Tiered Archive
›More than 5 PB per rack

ISV Solutions
Scality
Cleversafe
OpenStack Swift

CP-3584
›Up to 84 x 10 TB drives
›Dual Controllers

SP-3424
›24 x 8 TB drives
›Dual Controllers

SP-3224
›24 x 2.5’ drives or SSDs
›Dual Controllers

SAS
›NL SAS
›10 TB
›7.2K RPM

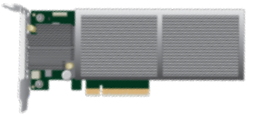
›HPC Drive
›4TB
›10K RPM

SATA


›SMR Drive
›10 TB

SSD
›SAS SSD
›1.3 TB
›Up to 60 TB

›NVMe
›1.3 TB

Flash accelerators


›PCIe x 16
›NVMe
›10 GB/s

* File system performance (GB/s) per [HDD, RU, Enclosure, Rack ….]

Problem Statement

Workloads are becoming increasingly unpredictable for the storage system

1. **Many compute applications on the same storage platform**
   - Mixed I/O patterns - Random, unaligned, strided, small, large, IOPS, GB/sec)
2. **HPC storage needs to be more than "Scratch"**
   - Home / Work directories
   - Databases that support mixed analytics workloads
3. **I/O Optimization System Administration is difficult**
   - Results is low application productivity

Ways to solve this mixed workload problem

- Over provision with additional storage servers & disk drives
- Buy a lot of expensive all Flash Arrays
- Deploy the new ClusterStor Nytro Intelligent IO Manager as Workload Accelerator on the Storage Platform

NytroXD – Overview

Summary:

- NytroXD – Core software component in NXD that implements Write Back Caching

I/O Architecture:

- Kernel I/O Filter Driver implemented as a Linux Device Mapper (DM) Target Driver that intercepts I/O and routes Small Blocks for Caching. Large Blocks bypass caching, and are directly handed off to underlying driver (GridRAID).
- Caching is implemented as a Cache Management Library with well-defined API's, deployed as a Linux kernel module
- Works at the kernel block layer - transparent to file system and applications
- Hardware agnostic, can work with any block device

Nytro Intelligent IO Manager

… uses intelligent caching algorithms to accelerate IO performance of applications

- ✓ **Applications tend to access some data more frequently than the other data**
- ✓ **NXD solution dynamically tracks this frequently accessed data and caches them using Intelligent caching algorithms**
- ✓ **Application performance is accelerated by servicing the IO request from the cache device without the need for a block/ RAID device to fetch from the backend media.**

Using Write Back cache & Read Persistence

- Small Block Random Acceleration
- Optimized for Large Block Sequential Advanced Seagate SSD management
- Dynamic tunables in cache software
- Dynamic Cache Flush Logic (Automatic, Manual) - Staged HPC application workloads
- Device IO error handling
- Cache validation, cache states with Read Persistance
- ClusterStor GridRAID & HA support
- Command Line Interface - Management
- Performance Monitor - cache hits, cache performance etc. &

Many more…

ClusterStor L/G 300 Key Components – Nytro-Option

Seagate Nytro XD Cache Management Software

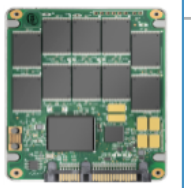
- Linux Filter Driver per OSS
- Monitors Writes Block Stripe Size
  - Admin Definable Threshold
    - *Eg; 32kb Block Stipe Size or less to SSD*
  - Small Blocks Write to SSDs
    - Data Flush/Writes to HDDs
  - Large Blocks Write to HDDs

ClusterStor Scalable Storage Unit

SSD Disk Pools are Configured as 1+1 / RAID 10 w/OSS High Availability

Object Storage *Server #1*

Object Storage *Server #2*



Small Block Stripe Sizes are Cached to a SDD Disk Pool

› *Small Block Sizes are Written to the GridRaid HDD storage pool*

› *The Last Accessed Small Block Stripe is Written to the HDD OST in a Continuous "Cache Flush" Cycle*

Large Block Stripe Sizes are written to HDD

Small Block Stripe Sizes are Cached to a SDD Disk Pool

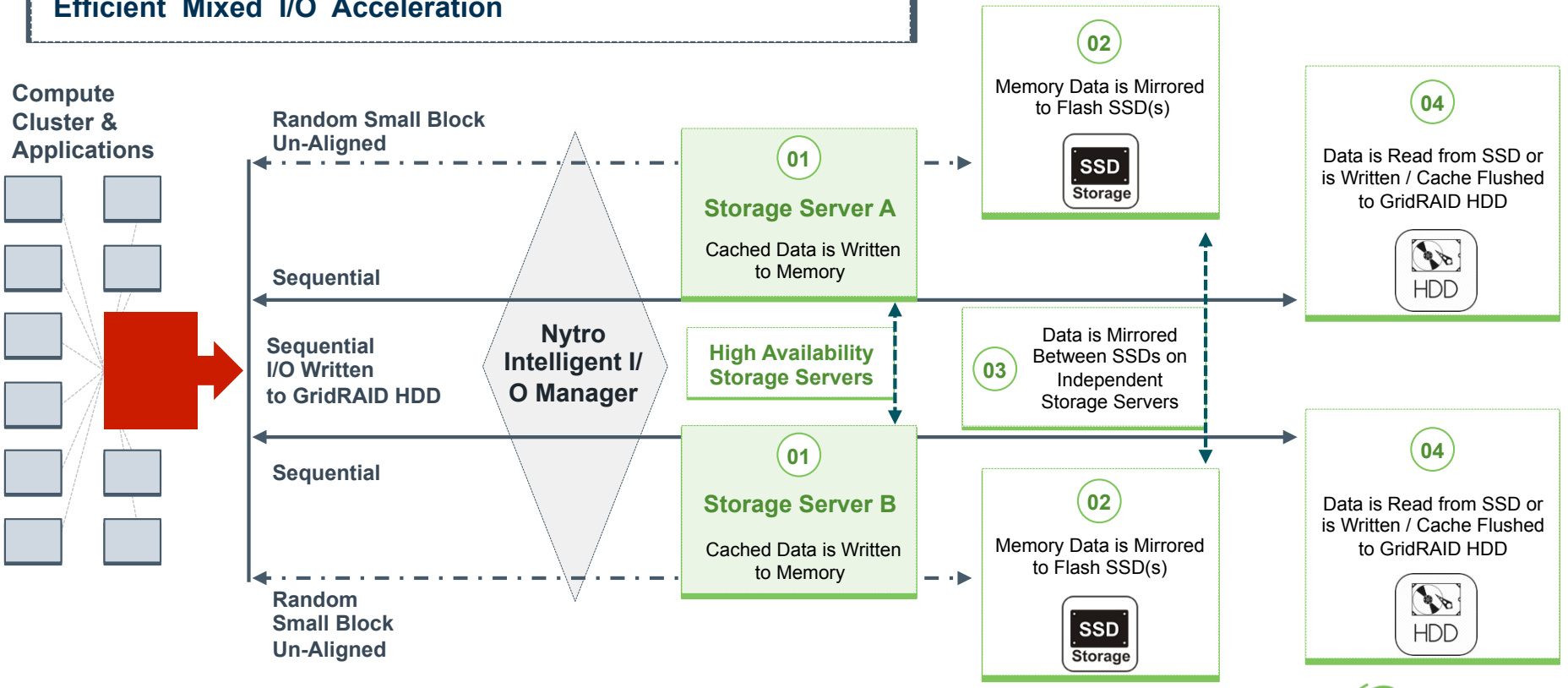
› *Small Block Sizes are Written to the GridRaid HDD storage pool*

› *The Last Accessed Small Block Stripe is Written to the HDD OST in a Continuous "Cache Flush" Cycle*

Large Block Stripe Sizes are written to HDD

Seagate ClusterStor: Any Workload, Any Time

Automated High Availability Architecture
Efficient Mixed I/O Acceleration



Seagate ClusterStor 300N - Nytro

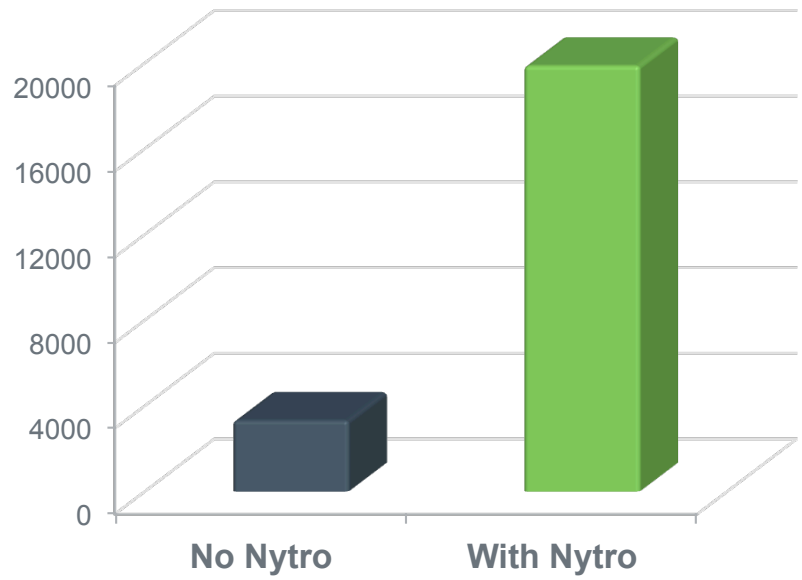
Transparent IOPS acceleration

- Introducing enhanced Seagate HPC Storage systems – 300N Nytro Platform
- New Hybrid SSD + HDD storage system – Lustre & IBM Spectrum Scale (GPFS)
- Nytro Intelligent I/O Manager transparently manage SSD and HDD
- Improves performance of small, random or unaligned IOs
- Uses Next-Gen Seagate SSDs, HDDs and upgraded Embedded Controllers
- Shipped now - Nytro will be enabled with system update in Q2'17

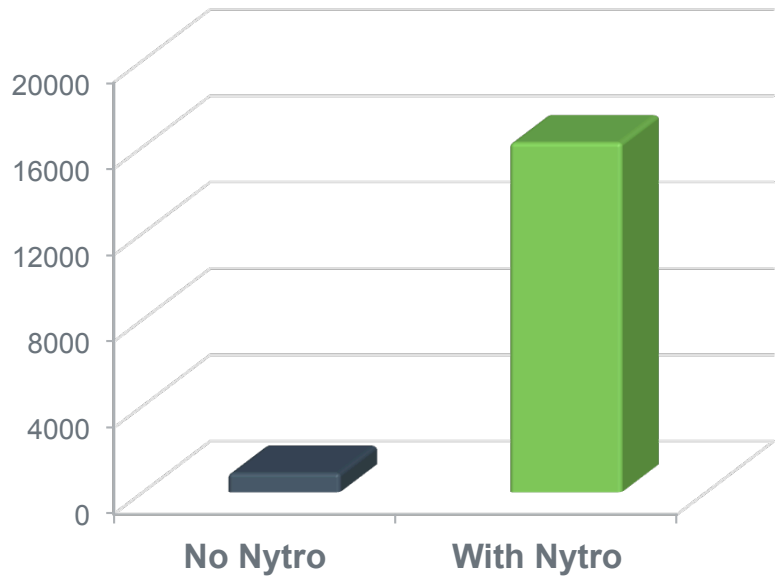
Seagate ClusterStor: Any Workload, Any Time

Lustre IOPS Acceleration

Lustre
4kb Rewrite IOPS
600% Performance Improvement



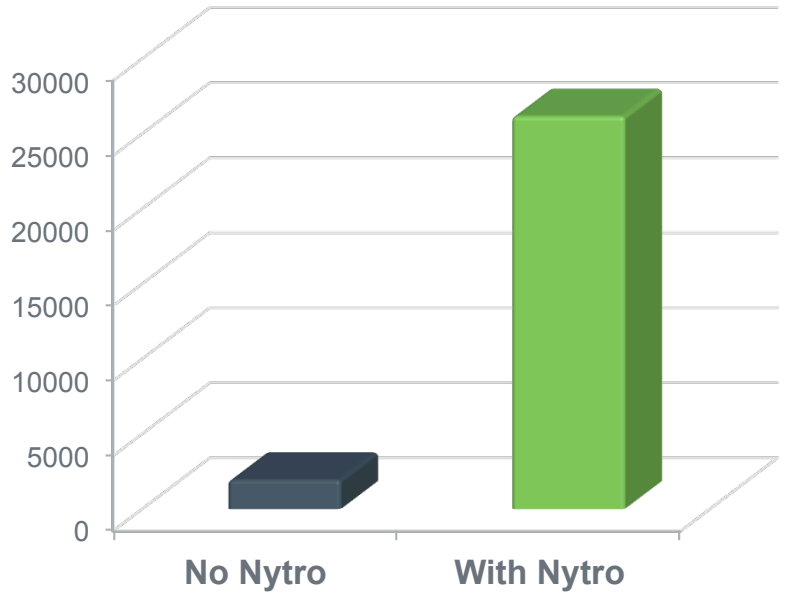
Lustre
16kb Rewrite IOPS
1840% Performance Improvement



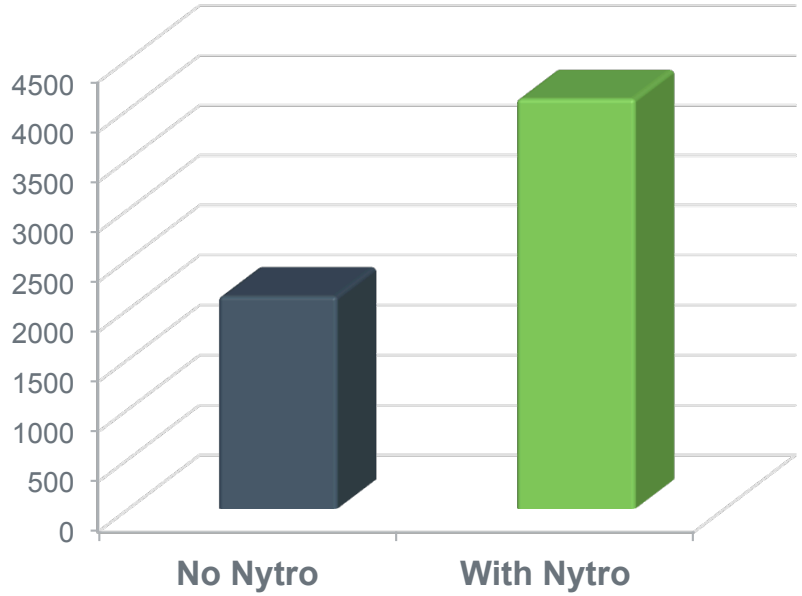
Seagate ClusterStor: Any Workload, Any Time

GPFS / Spectrum Scale IOPS Acceleration

Spectrum Scale / GPFS
4kb Rewrite IOPS
1335% Performance Improvement



Spectrum Scale / GPFS
16kb Rewrite IOPS
193% Performance Improvement



Questions?

SEAGATE

תודה
Dankie
Gracias
Спасибо
شكراً
Merci
Takk
Köszönjük
Terima kasih
Grazie
Dziękujemy
Dėkojame
Ďakujeme
Vielen Dank
Paldies
Kiitos
Täname teid
谢谢
Thank You
Tak
感謝您
Obrigado
Teşekkür Ederiz
감사합니다
Σας ευχαριστούμε
ขอบคุณ
Bedankt
Děkujeme vám
ありがとうございます
Tack