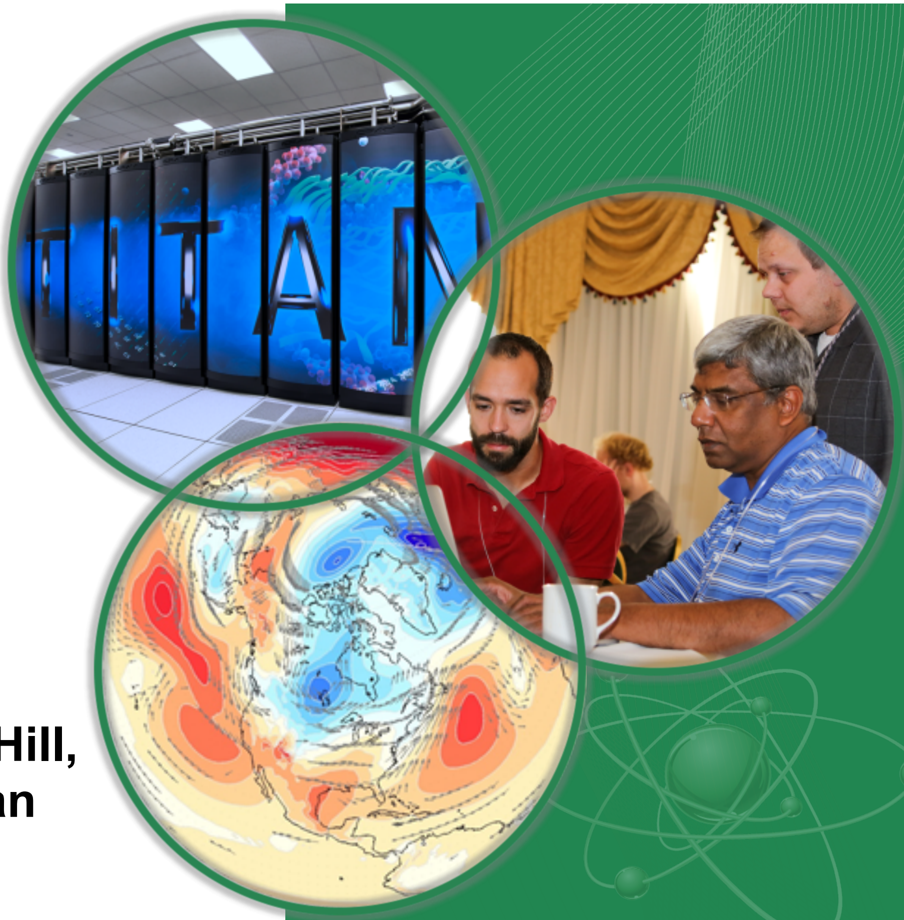
Comparative I/O Workload Characterization of Two Leadership Class Storage Clusters

Presented by Sarp Oral



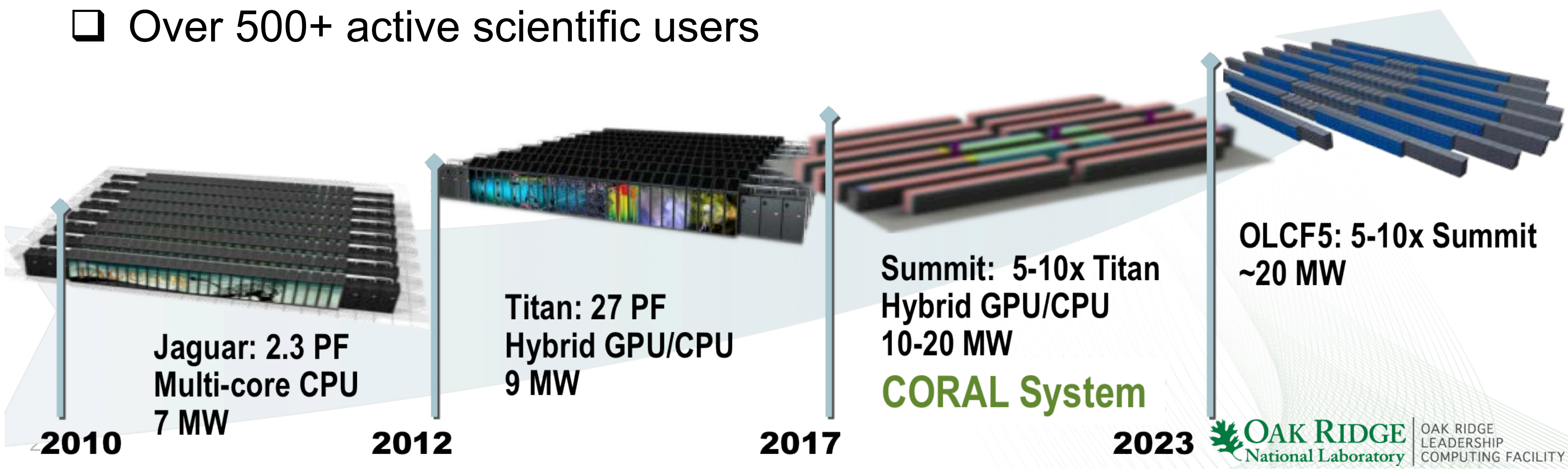
Raghul Gunasekaran, Sarp Oral, Jason Hill, Ross Miller, Feiyi Wang, Dustin Leverman
Oak Ridge Leadership Computing Facility.

ORNL is managed by UT-Battelle
for the US Department of Energy

OAK RIDGE National Laboratory | OAK RIDGE LEADERSHIP COMPUTING FACILITY

Oak Ridge Leadership Computing Facility

- Design and operate compute and data resources for the most computationally challenging science problems.
- Deliver science and transforming discoveries in materials, biology, climate, energy technologies, and basic sciences.
- 250+ research organizations, university and industry participants.
- Over 500+ active scientific users



Oak Ridge Leadership Computing Facility

- Compute resources
  - TITAN, primary compute platform, 18688 compute clients
  - EOS, CRAY XC30 compute platform, 736 compute node
  - Rhea, data analysis cluster, 512 node commodity cluster
  - Everest, visualization cluster
- Spider Storage System
  - 32PB, +1 TB/s - data resource for OLCF computational needs
  - Lustre parallel file system
  - Center-wide shared storage resource, for all OLCF resources
  - Resilient to system failures, both internal to the storage system as well as computational resources

OLCF System Architecture

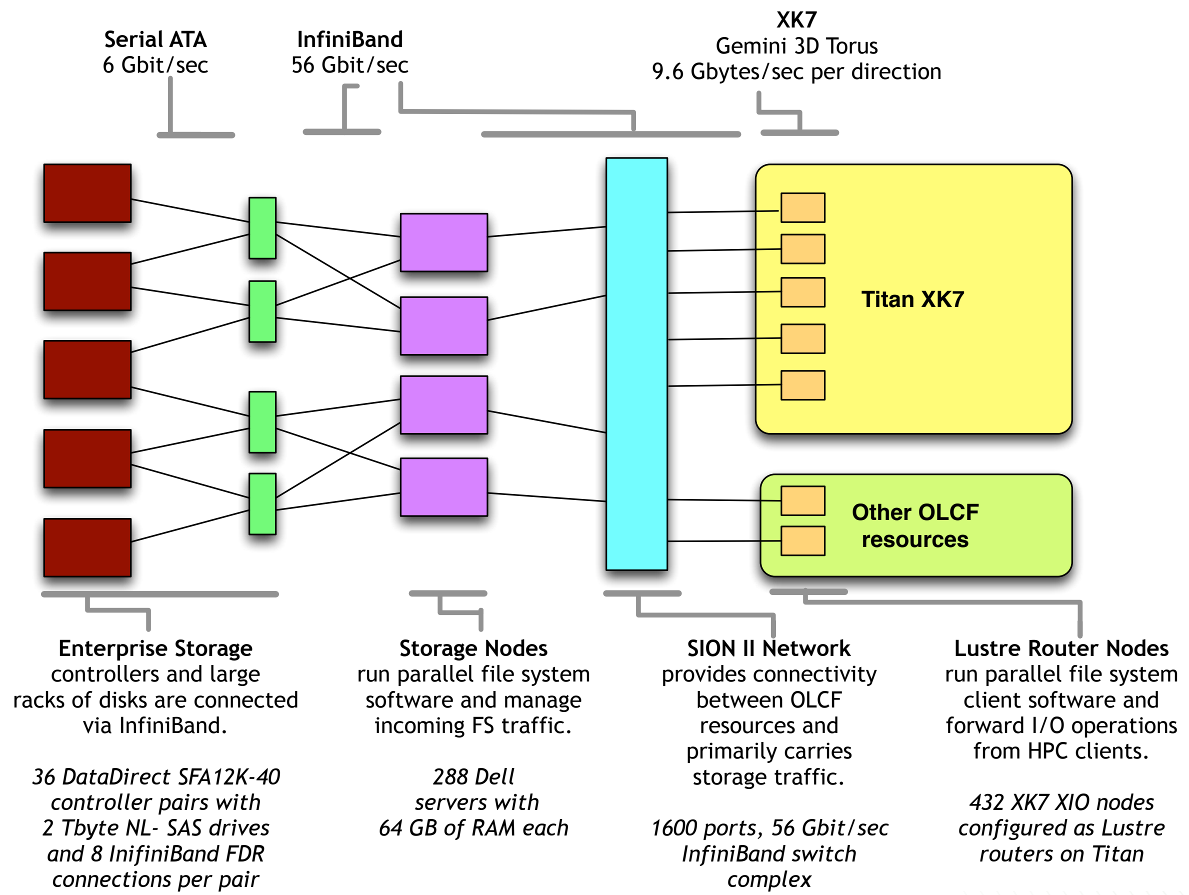
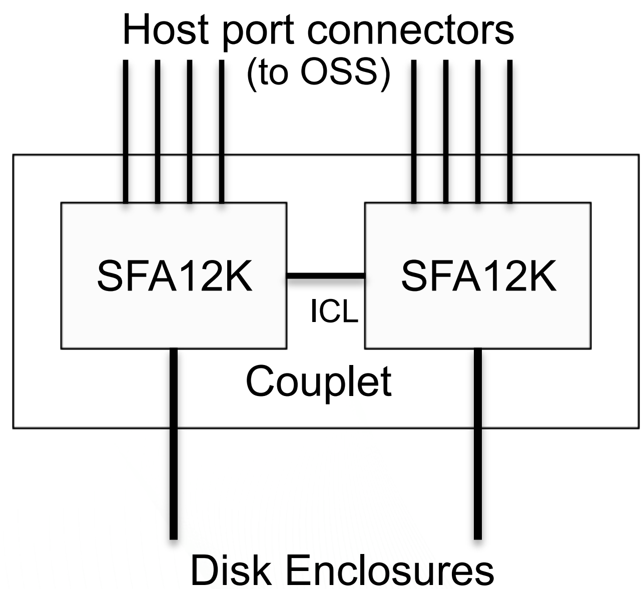


Figure reference: S. Oral, et al. OLCF's 1 TB/s, next-generation lustre file system. In the proceedings of the Cray User Group Conference, 2013

Spider 2 System

- Deployed 2014
- Max bandwidth:1.4 TB/s read and 1.2 TB/s write
- 36 DDN SFA12K couplets
- Two namespaces: Atlas1 and Atlas2
  - 18 Couplets each, no shared hardware
  - Purpose: Load balancing and capacity management
- Why a couplet
  - Failover configuration
  - Bottleneck: ICL (Inter Controller Link)
- Designed for mixed random I/O workload
  - Non-sequential read and write I/O patterns

Spider Couplet Setup



Spider File System - Comparison

	Spider 1	Spider 2
Years	2008 – 2014	2014 onwards
Bandwidth	240 GB/s	+1 TB/s
Capacity	10 PB	32 PB
RAID Controller	DDN S2A9000	DDN SFA12KX
Disk Type	SATA	Near-line SAS
Number of disks	13,440	20,160
Connectivity	IB DDR	IB FDR
Number of OSTs	1,344	2,016
Number of OSSs	192	288
Lustre version	1.8	2.5
Disk Redundancy	RAID 6 (8 + 2)	

Workload Comparison: Spider 1 vs. Spider 2

Primary Compute Platform: What changed ?

- 2.3 Petaflop Jaguar → 27 Petaflop Titan
- CPU → CPU + GPU
- Memory: 300 → 710 TeraBytes
- 3D Torus Interconnect bandwidth: 3.2GB/s → 10.4 GB/s
- I/O router nodes: 192 → 440

What did not change ?

- # of compute clients: 18688
- Spider architecture (just scaled up)

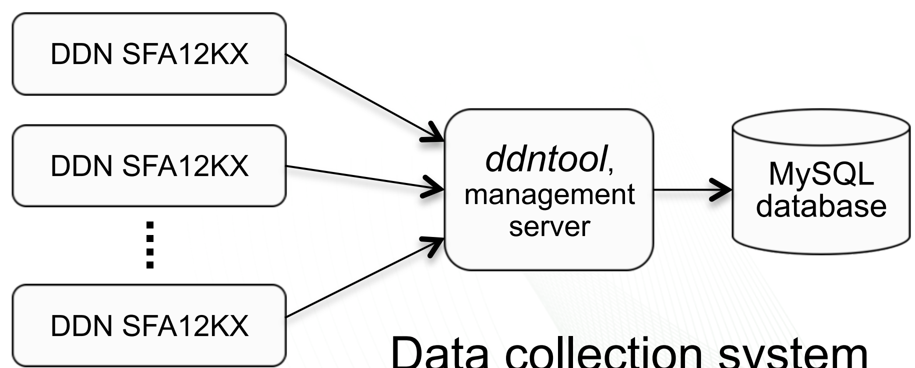
Workload Characterization

Workload Data

- From the DDN RAID controllers; using ddntool, a custom tool developed at ORNL
- Periodic polling: read/write bandwidth and IOPS, request size and latency data.
- Spider 1 data from 2010 (Jan – June); Spider 2 data from 2015 (April – August)

Characterization Metrics

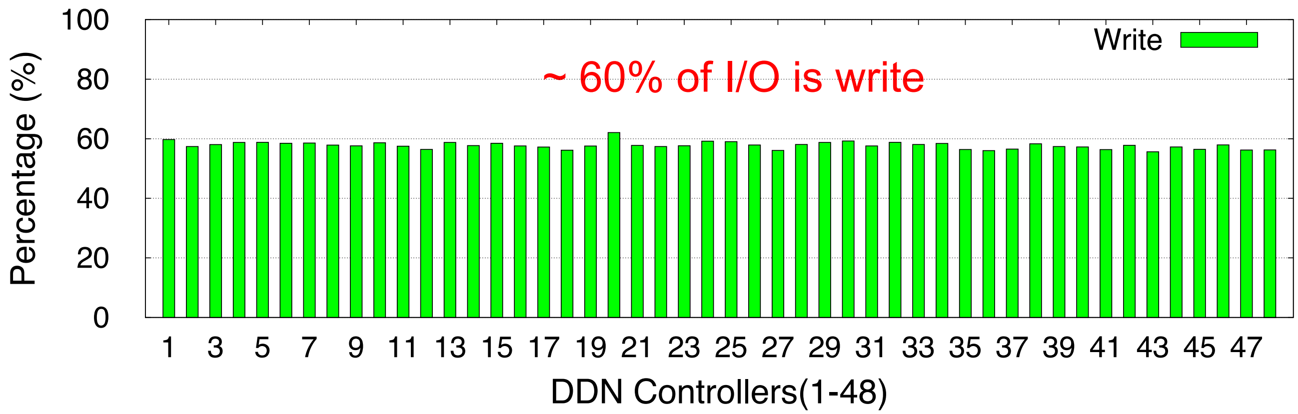
- I/O Access (Read vs Write)
- Peak bandwidth utilization
- I/O Bandwidth usage trends
- Request size distribution
- Service latency distribution



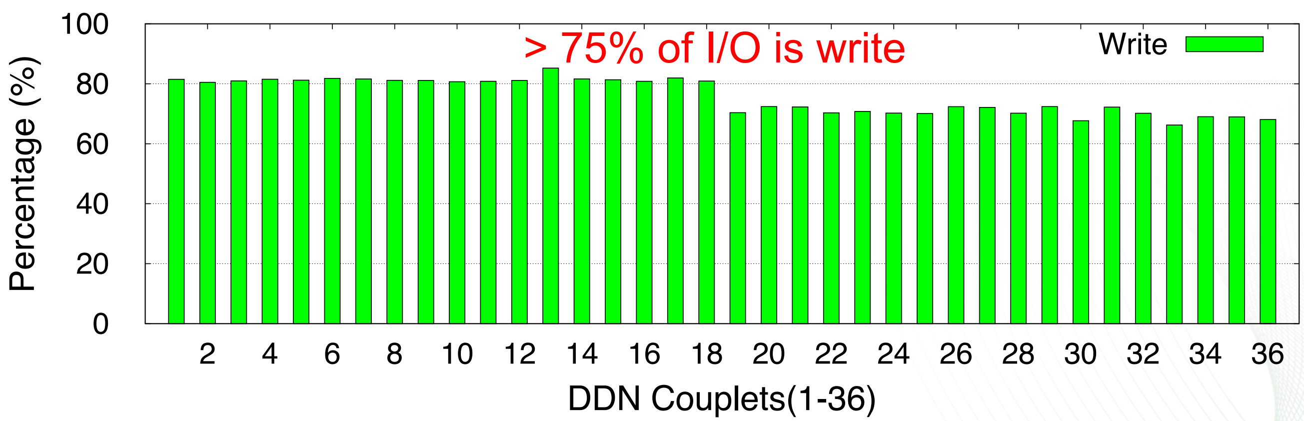
Data collection system

Read vs Write

Spider 1

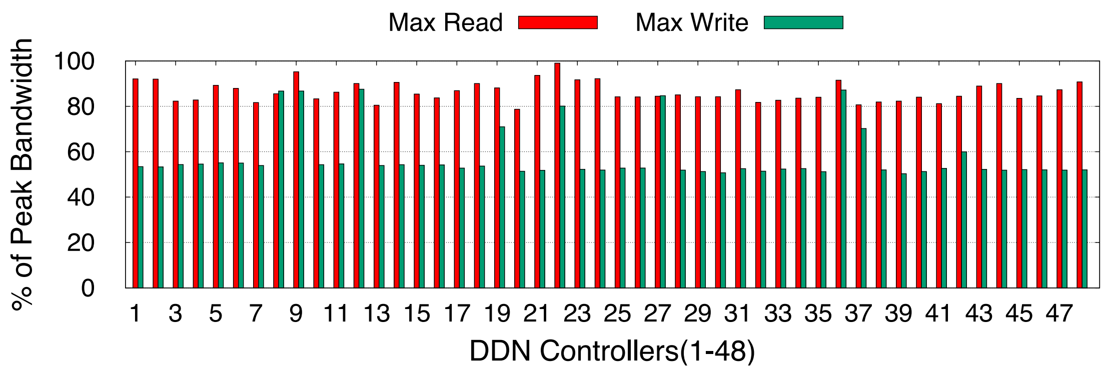


Spider 2



Peak Bandwidth Utilization

Spider 1



Spider 2

Peak Bandwidth
Spider 1

- ~ 90% for read
- Only 50% for write

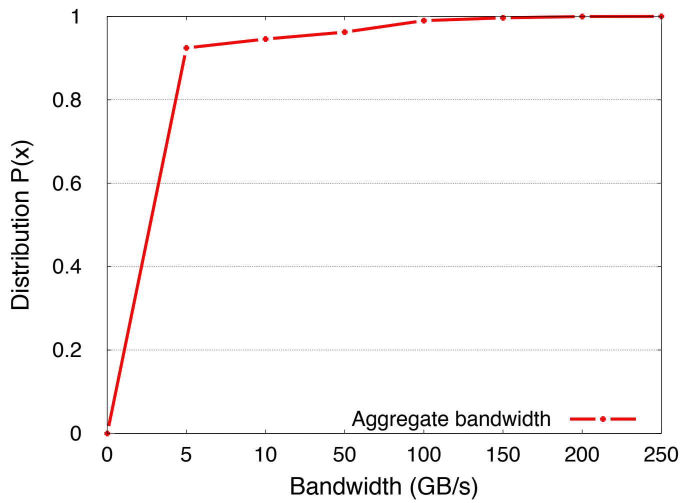
Spider 2

- ~ 80% for read
- ~ 75% for write

Reasons:

- Larger request sizes
- Write-back cache enabled

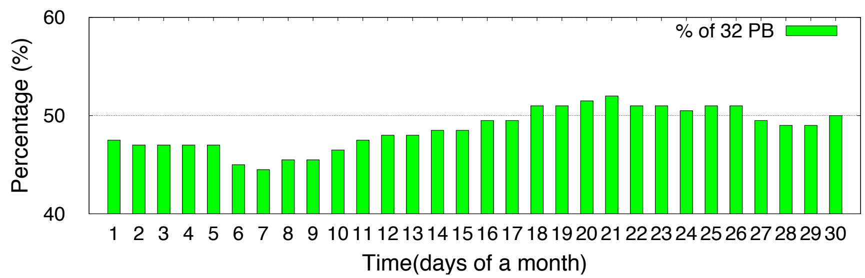
Spider 2 - Bandwidth Usage Trends



Cumulative Distribution Function (CDF)

- ~92% time usage is less than < 5 GB/s
- This is expected
  - Most applications are compute-intensive
  - < 5% of runtime is spent on I/O
  - Scientific application's I/O are bursty

BURST BUFFER !!!!



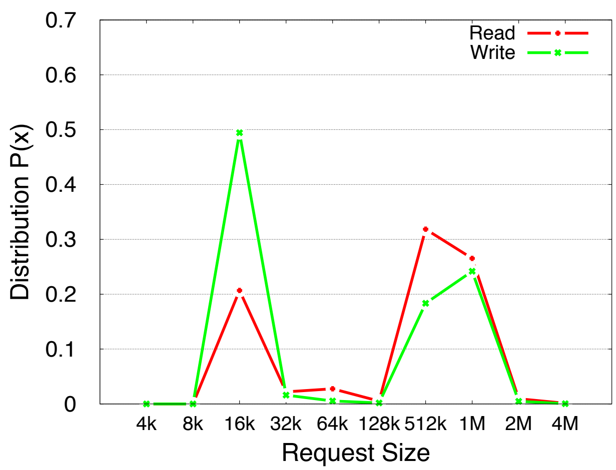
Storage system usage over a month

~50% of our storage space is utilized on an average with

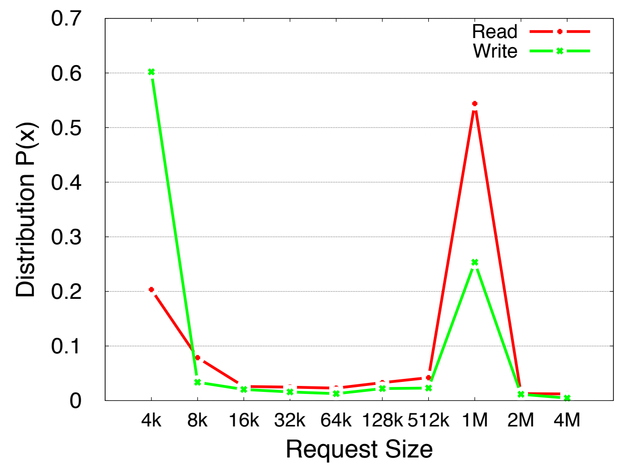
- Data being purged periodically
- Large file system idle time (<5GB/s)

Request Size Distribution

Probability Distribution Function (PDF)



Spider 1

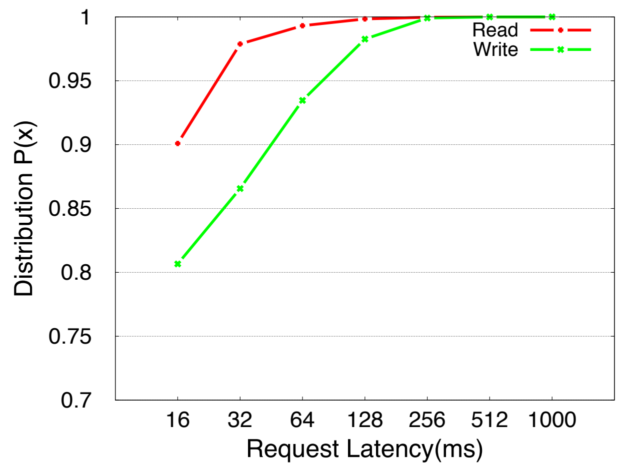


Spider 2

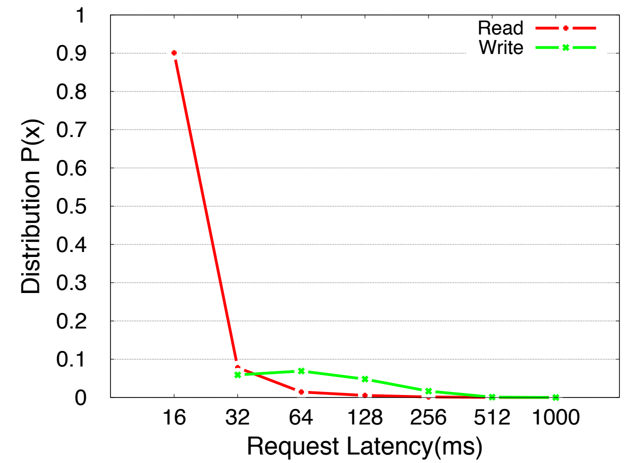
- Smallest measurable unit on Spider 1 is 16 KB, Spider 2 is 4KB
- Large 512 KB requests on Spider 1
  - *dm-multipath* issue, breaks1MB requests to 2, 512 KB requests
  - *deadline* I/O request scheduler, in 2011 migrated to *noop* scheduler

Request Service Latency Distribution

Spider 2



Cumulative Distribution Function (CDF)



Probability Distribution Function (PDF)

- Service Latency = Queue time + Disk I/O time
- 90% of read requests, and 80% of write requests served in less than 16ms
- 16ms is the smallest measurable unit on the DDN controllers

Request Service Latency Distribution

- Read-ahead cache disabled
  - Mixed aggregate read workload is non-sequential
  - Prefetching read blocks impacts performance (cache trashing)
- Write-back cache enabled
  - ReACT (Real-time Adaptive Cache Technology)
  - 1MB data blocks written to disk directly, no caching on peer controller
  - <1MB data blocks
    - Cached and mirrored on either controllers
    - Grouped for single large block write

Conclusion

- What is our next storage system (for Summit 100+petaflop) ?
  - Simply scale up Spider 2 ? Not very likely !!!!
  - But we will need a center-wide shared storage system like Spider
  - Explore: Burst Buffer or an intermediate fast I/O cache layer
- Expected I/O workload trends
  - Increased write I/O
  - Bursty, with identical or increased file system idle times
  - Support for larger request sizes
- Open Questions
  - How does the next generation of compute platform affect storage system design
  - Summit: 20+ → 100+ Petaflops but scaling down from 18k to 4k compute nodes