



StorPool
DISTRIBUTED STORAGE

Performance test results

May 2015

Introduction

StorPool is distributed storage software. It is installed on each node in a cluster of servers and combines the capacity **and** performance of all of the drives attached to the nodes into one **shared pool of block storage**.

StorPool provides standard block devices and customers can create volumes through the sophisticated volume manager. Redundancy is provided by multiple copies (replicas) of the data written synchronously across the cluster, which provides high reliability, without the performance degradation, typical of erasure coded systems. The system is managed through a CLI and JSON API. More details about our technology are available [here](#).

StorPool replaces traditional storage arrays, all-flash arrays or other inferior storage software (“SDS 1.0” solutions). It provides exceptional levels of performance, reliability and flexibility and can greatly reduce storage (and infrastructure) costs.

This document presents results from performance tests run in StorPool's test lab. No test result can be considered as an exact replica of actual workloads, therefore these results should only be used as an indication of the expected performance of StorPool. Customers are advised to perform their own set of tests, for their typical application before running StorPool in production.

Executive Summary

StorPool is a distributed storage system running on standard server hardware. It uses minimal system resources to achieve outstanding performance.

Performance highlights of a small system comprising of 3 servers, 12 SSDs, 18 hard drives, 10GE network are as below:

Test	Hybrid , 3 copies*	All-SSD, 2 copies*
Sequential reads	4,200 MB/s	4,844 MB/s
Sequential writes	890 MB/s	2,251 MB/s
Random reads 4k block size	250,000 IOPS	230,000 IOPS
Random writes 4k block size	62,000 IOPS	112,000 IOPS

*Represents the number of copies of the data

Resources reserved for the storage system - total across 3 servers:

- 8GB RAM per node (24 GB RAM total)
- 3 CPU cores per node (9 CPU cores total)

The remaining resources can be used for running virtual machines in a converged Storage+Compute architecture. This means running both storage and computation (applications) on the same servers to significantly increase utilization and reduce TCO.

The total cost of the hardware used in the test is \$ 15,993 when using Micron 480GB SSDs and \$ 14,361 when using Intel 240GB SSDs and the usable space in this Hybrid configuration being 5.2 TB using the Micron SSDs and 2.6 TB when using the Intel SSDs.

A small production deployment with 13 TB equates to \$ 2.53/GB and \$ 0.13/IOPS.

The system is designed to scale linearly with number of drives and servers. For example, a system with 30 nodes will have roughly 10x the performance of the 3 node system which is benchmarked below.

Server Configuration

Name	CPU	RAM	RAID/HBA	Drives
s07	Xeon E3-1220V2 4 cores × 3.1GHz	16 GB	LSI SAS 2308 Fusion	4x SSD - Micron M500DC AHCI - Intel S3500 6x HDD - HGST Ultrastar 1TB
s08	Xeon E3-1220V2 4 cores × 3.1GHz	16 GB	LSI SAS 2308 Fusion	4x SSD - Micron M500DC AHCI - Intel S3500 6x HDD - HGST Ultrastar 1TB
s09	Xeon E3-1220V2 4 cores × 3.1GHz	16 GB	LSI SAS 2308 Fusion	4x SSD - Micron M500DC AHCI - Intel S3500 6x HDD - HGST Ultrastar 1TB

Hard drive model:

- HGST Ultrastar 1TB (HUA722010CLA330)

SSD models:

- Intel DC S3500 240GB (SSDSC2BB240G4)
- Micron M500DC 480GB (MTFDDAK480MBB)

Total:

- 3 servers
- 12 SSDs
- 18 hard drives

Network configuration

- 3x Mellanox Technologies MT27520 ConnectX-3 Pro
- Dual 10GE link per server
- Switch: Dual Dell S8024F – 24-port 10GE SFP+ switch
- 9000 bytes MTU (Jumbo frames)
- Flow control – enabled

Software configuration and testing methodology

The tests are performed on 3 servers in total. The 3 servers are used as storage nodes, and the same 3 servers are used as storage clients and run the benchmarking software (fio).

Each test run consists of:

1. Configuring and starting a StorPool cluster
2. Creating three 100GB volumes
3. Filling synchronously the volumes with random test data
4. Performing all test cases by running fio on three client machines in parallel, starting and finishing in sync

The following table summarises the test parameters common to all tests.

Operating system	Ubuntu 14.04 LTS
Performance-testing software	fio-2.1.7 and StorPool test runner scripts
Number of servers used for storage nodes	3
Number of servers used for storage clients	3 (shared with the above servers)
Number of volumes	3
Volume size	100 GB each
Replication level (number of copies)	
- All-HDD (18 HDDs)	3 copies (all on HDDs)
- Intel-Hybrid/Micron-Hybrid (12 SSDs and 18 HDDs)	3 copies (2 copies on HDDs, 1 copy on SSD)
- All-SSD Intel/Micron (12 SSDs)	2 copies (similar to RAID1/RAID10)
Number of drives used for tests	18 hard drives, 12 SSDs
Kernel version	3.13.0-49-generic
Distributed storage software version	StorPool 15.02.23
Integrity provided by system	End-to-end data integrity. Protects data throughout its lifetime.
Caching, Buffering, Data consistency	Read caching with 4GB cache per storage node Write-through caching shared with read cache Write-Back cache enabled with up to 100MB per hard drive, shared with read cache

Test cases

Test name	Read/write	Block size	Queue depth	Duration
IOPS tests				
Random write	Random Writes	4 KB	4, 16, 64, 256	1min
Random readwrite	Random Reads + Random Writes 50/50	4 KB	4, 16, 64, 256	1min
Random read	Random Reads	4 KB	4, 16, 64, 256	10sec
Latency tests				
Write latency	Random Writes	4 KB	1, 2, 4, 8, 16, 32	1min
Read latency	Random Reads	4 KB	1, 2, 4, 8, 16, 32	10sec
Sequential tests				
Sequential write	Sequential Writes	1 MB	4, 16, 64, 256	1min
Sequential read	Sequential Read	1 MB	4, 16, 64, 256	1min

All tests use fio with the following options:

- ioengine=libaio
- direct=1
- sync=1
- norandommap
- randrepeat=0.

Test runs

The following scenarios were tested.

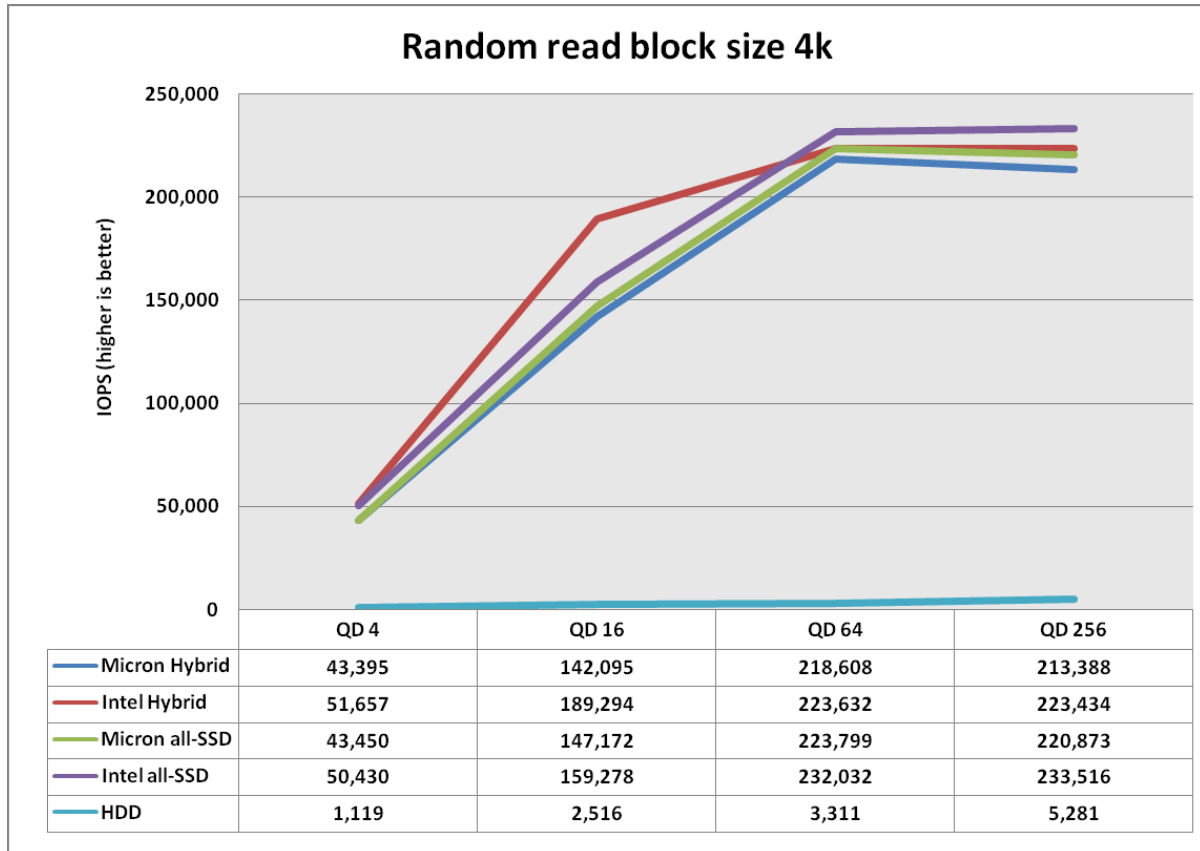
- Triple replication
 - Hard drives – 18 HDDs
 - Hybrid (Micron) – 18 HDDs and 12 Micron SSDs
 - Hybrid (Intel) – 18 HDDs and 12 Intel SSDs
- Dual replication
 - All-SSD (Micron) – 12 Micron SSDs
 - All-SSD (Intel) – 12 Intel SSDs

For each scenario all tests were run in sequence, after filling the 100GB volumes with random data. All workloads are limited in time and started synchronously.

The queue depth of the workloads is one of the key factors that affect storage performance. Queue depth marks the number of outstanding operations waiting to execute to the underlying drives. For example, the typical database workload has a relatively low number of outstanding waiting operations, as opposed to a typical web server, which will have a much higher number of outstanding operations.

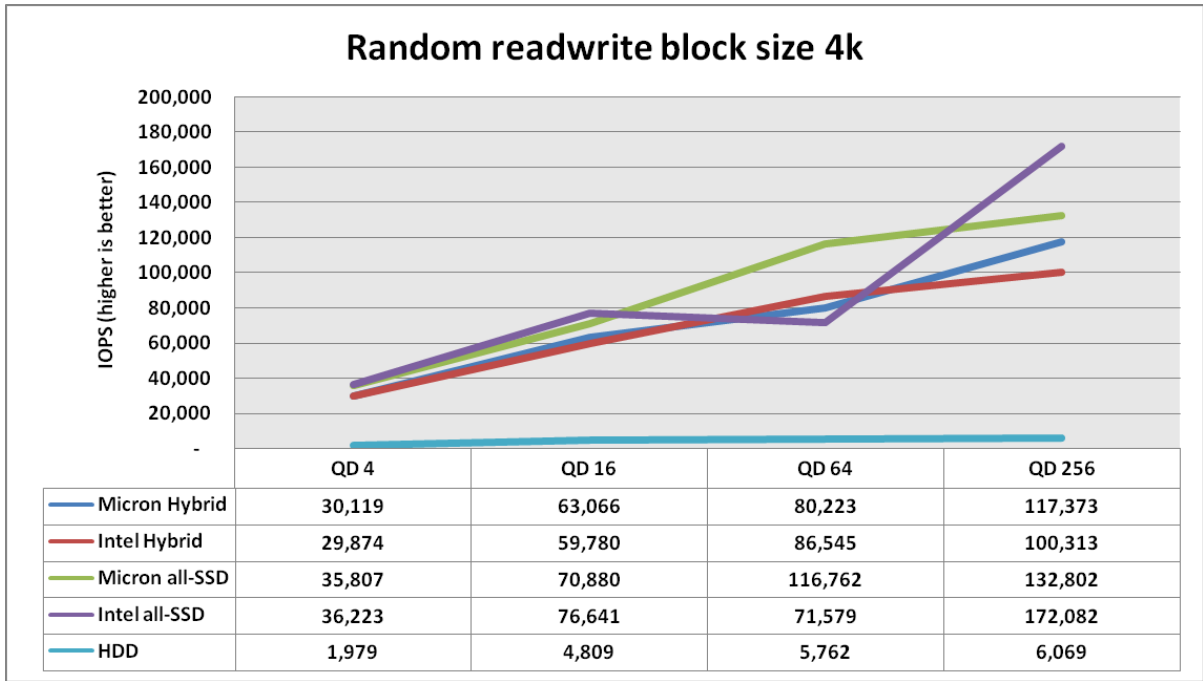
Test results

Random ops (IOPS)



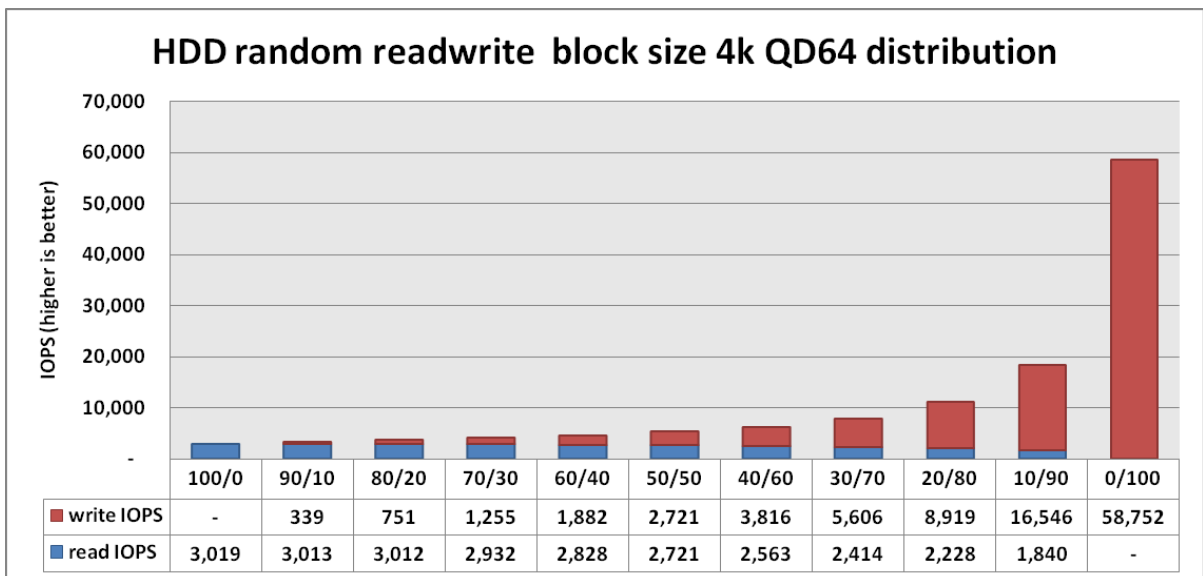
In terms of random read performance it can be seen that there is not a significant difference between all-SSD and hybrid volumes. The reason for this is due to the way StorPool handles read operations. In hybrid pools all reads are carried out by the SSD part of the pool, this is because a full copy of the data is stored on the SSDs.

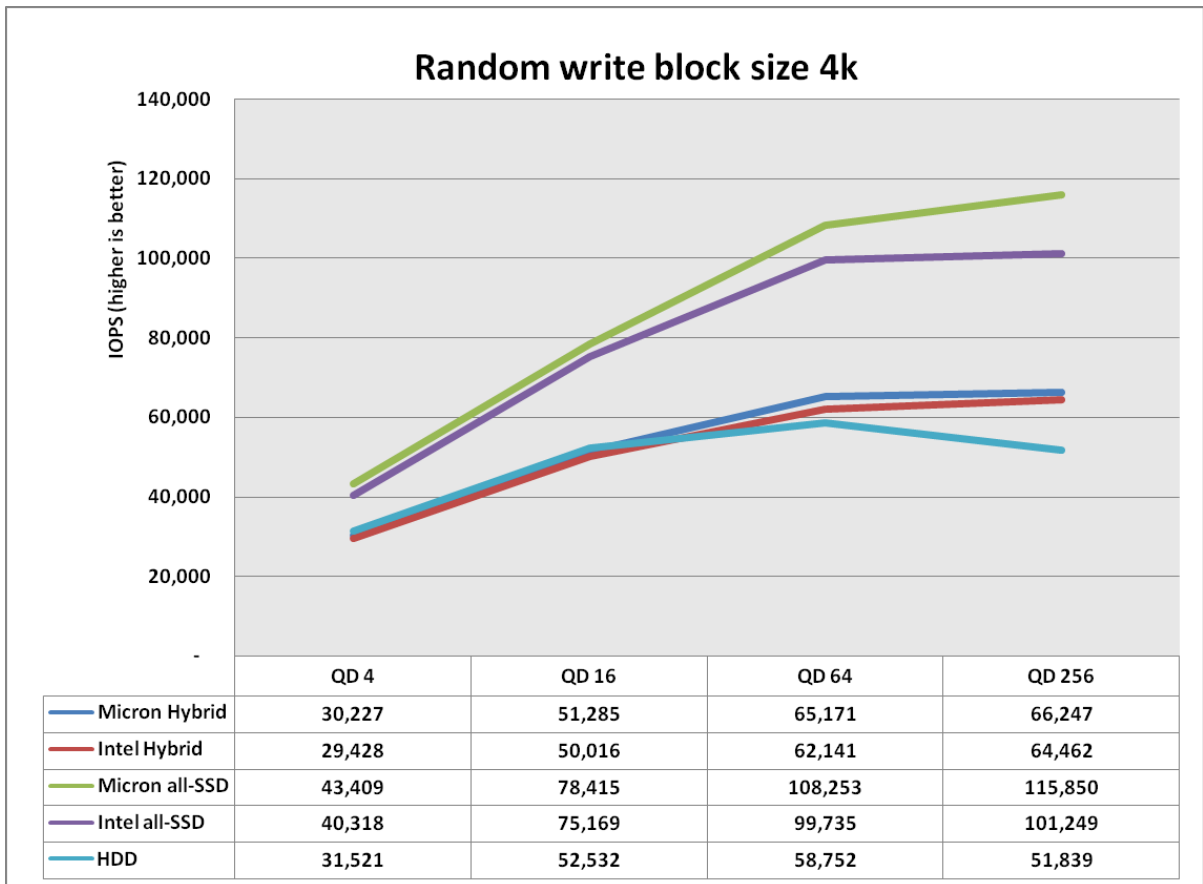
It should be noted that although the IOPS results achieved by the all HDD pool, are much lower than the results achieved by both the all SSD and Hybrid pools, in fact the IOPS results attained by the all HDD pool, would be far lower, if it was not running utilising the StorPool software.



This workload provides some insight on expected mixed peak performance in a 50/50 random read/random write distribution. Depending on the distribution of the mix between random reads and writes the expected result would trend closer to either pure random read or random write results (random write results are on the next page).

There is a notable exception with HDD pools, whereby in cases of mixed random workloads, heavily inclined towards writes, the overall IOPS result will be much higher, since all write operations are write-backed as seen below.



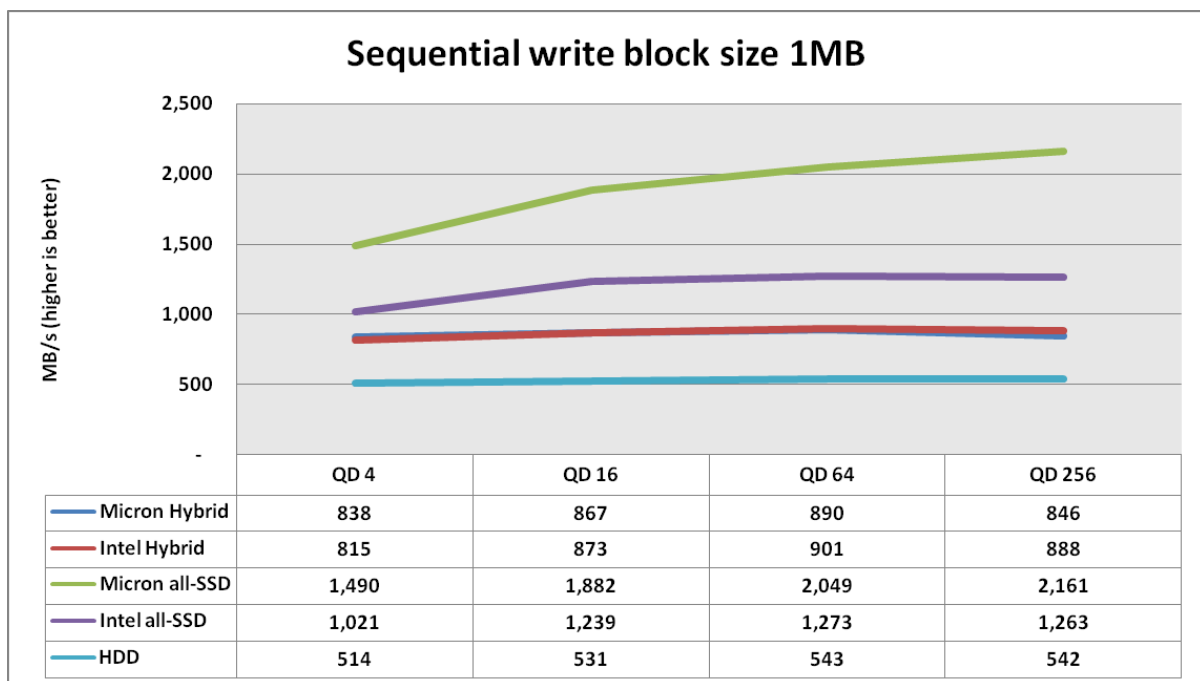
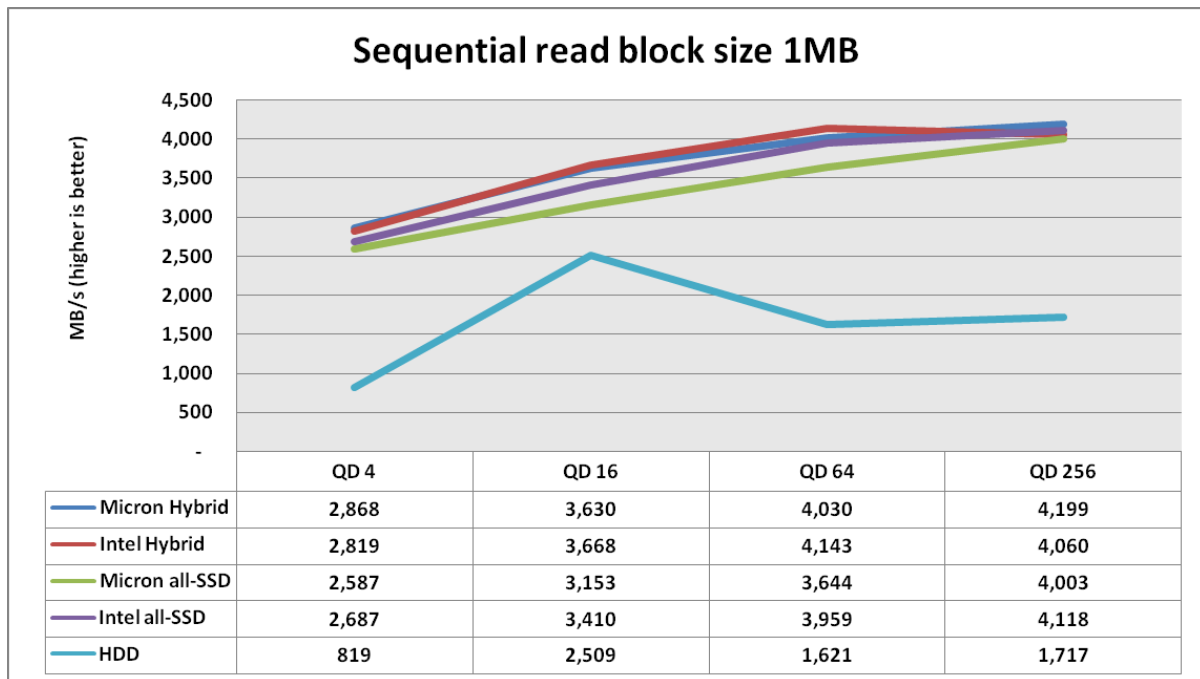


There is a visible difference in performance between all-SSD and hybrid/HDD pools. A part of the explanation for this, lies again in the number of copies of the data that has to be written on the underlying disks/SSDs, for every write operation.

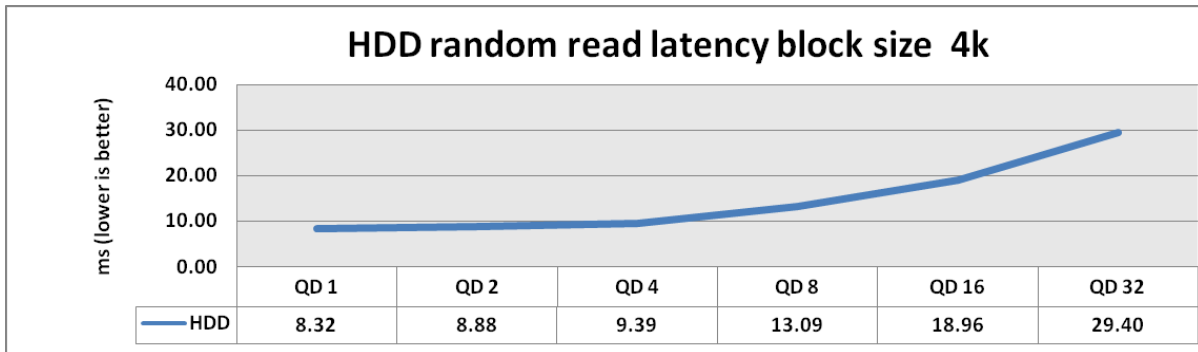
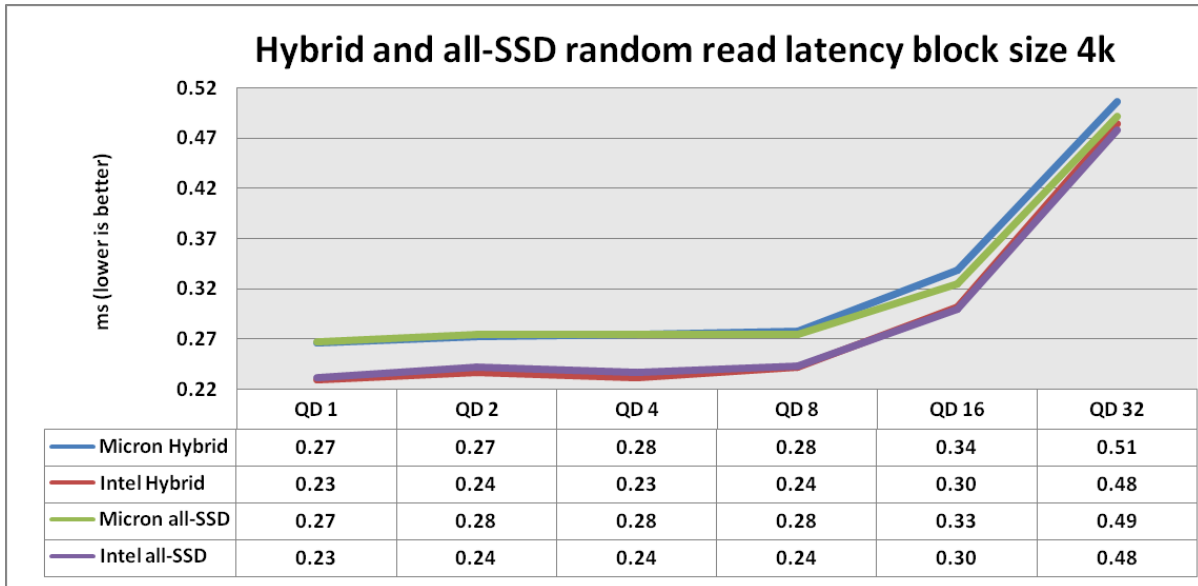
As shown by the chart, creating 3 copies of the data in the write backed disk drives provides close results compared to hybrid types (two copies on disks and one on SSDs). In heavier workloads with queue depths higher than 64 the HDD volume's performance is seen to decrease, mainly due to the small number of drives in the HDD pool.

In this example, the 18 drives in the HDD pool still have to do 3 copies of the data for every write operation. Every HDD has up to 100 MB of write back cache, so in this example there is 1.8 GB throughout the cluster.

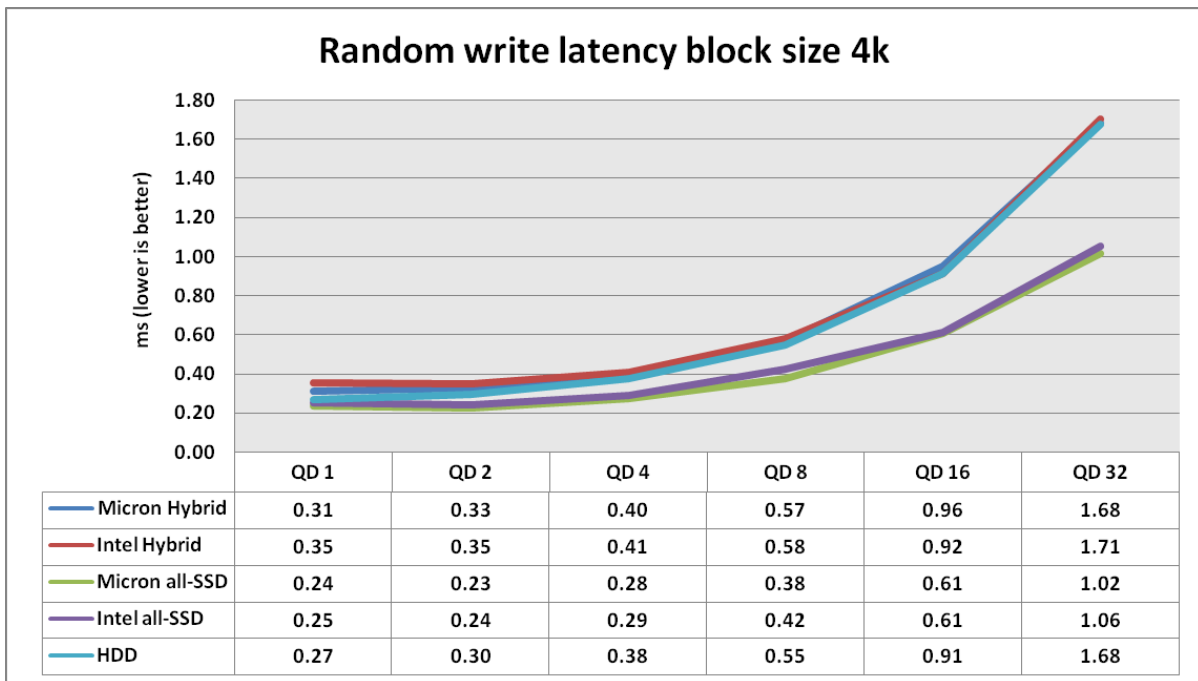
Sequential (MB/s)



Latency random reads (ms)

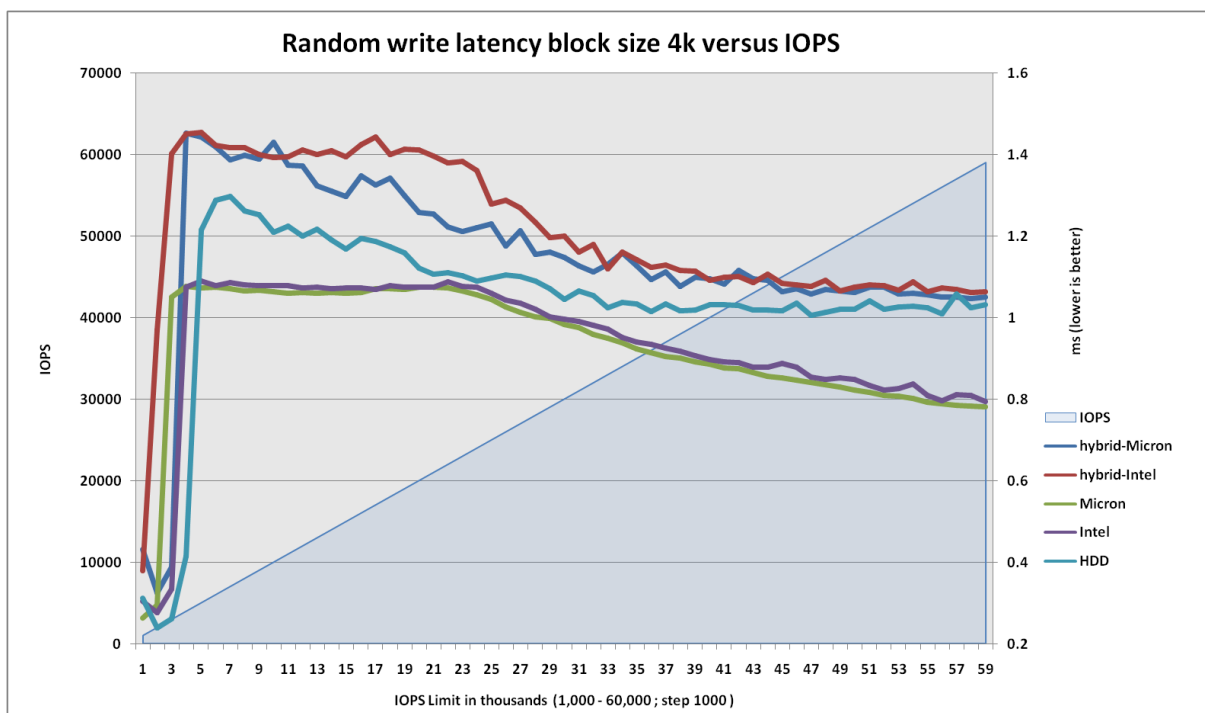
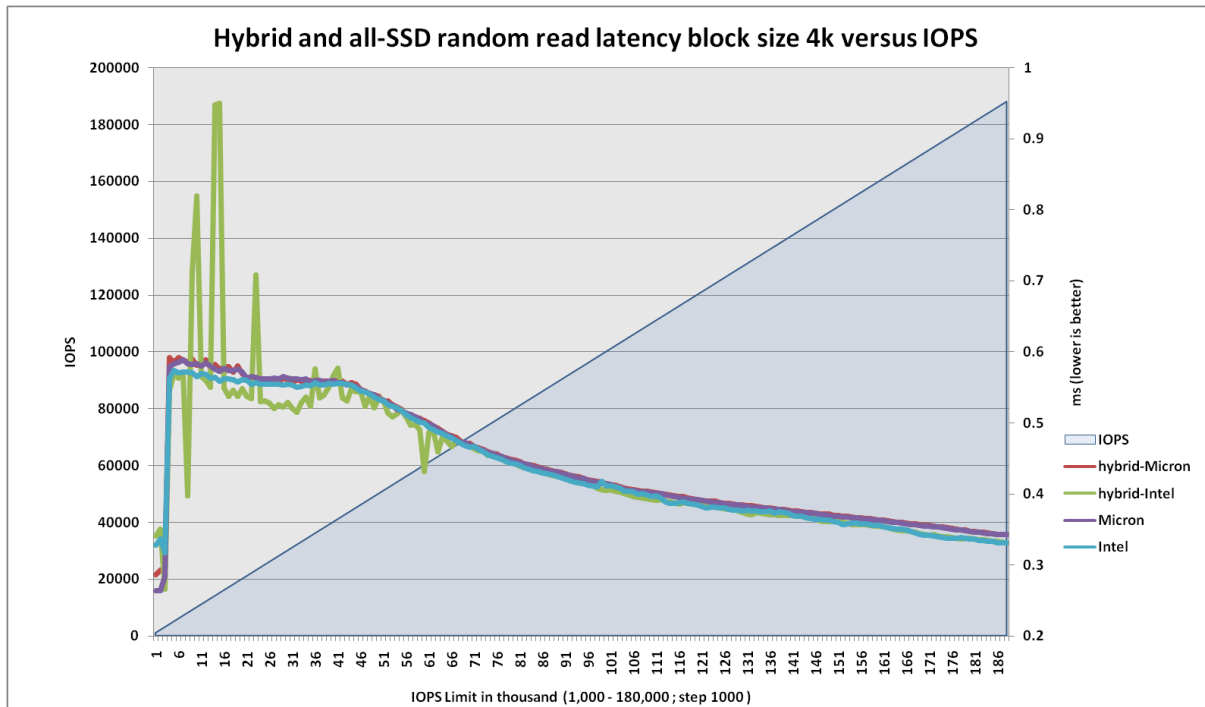


Latency random writes (ms)



Latency vs IOPS

This test was performed using a single client and provides more detailed results on latency distribution within a limited range of IOPS, starting from 1000 right up to the upper limit that the storage system provides with this setup (186 thousand IOPS for random reads, 60 thousand IOPS for random writes). The workloads represent a burst of random reads or writes started one after another, each with 10 seconds duration. The charts represent the average latency measured for every burst.



Conclusion

StorPool provides exceptional performance, whilst at the same time providing end-to-end data integrity and shared storage capabilities. In addition, StorPool maintains high performance and maintains unrivalled efficiency, even whilst running and managing many competing workloads.

In the last two major releases StorPool introduced Write Back Cache and support for standard ATA over Ethernet (AoE) protocol which enabled sub-millisecond write latencies and expanded our compatibility.

Other key features are StorPool's integration with OpenNebula (ONE), official OpenStack support and compatibility with OnApp hypervisors.

Commercially StorPool unlocks potentially huge reductions in TCO, offering exceptional value for money, when compared to the costs of existing traditional shared block storage solutions, all flash arrays and other alternative software products.

Contacts

If you would like to learn more or wish to test the product, please feel welcome to contact us:

StorPool

+1 415 670 9320

info@storpool.com

www.storpool.com

@storpool

StorPool UK

+ 44 (0) 207 097 8536

ukoffice@storpool.com