

---

# KALISTA IO

Get ready for a data storage revolution

# Phalanx Ceph OSD and Rados benchmarks

# Test system

---

# Test system configuration

Processor Intel(R) Core(TM) i7-4771 CPU @ 3.50GHz

---

Memory 16GiB DDR3 Synchronous 2400 MHz

---

Storage interface SATA 3.2, 6.0 Gb/s

---

OS device 128GB Samsung 840 PRO Series SSD (MZ-7PD128)

---

Metadata device 480GB Samsung SM843T (MZ-7WD4800/003)

# Storage devices

---

## Device info



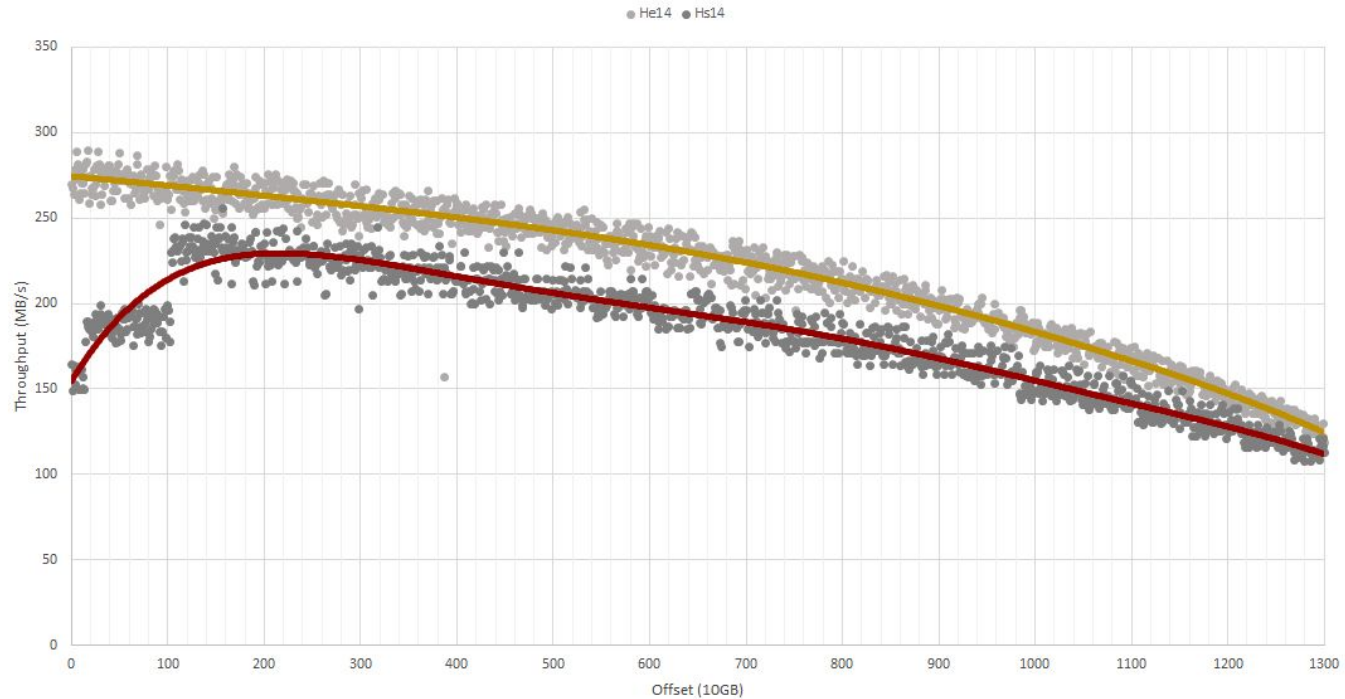
---

Model number	HSH721414ALN6M0 (Hs14)	WUH721414ALE6L4 (He14)
Firmware revision	L4GMT240	LDGNW07G
Drive type	Host managed SMR	CMR

---

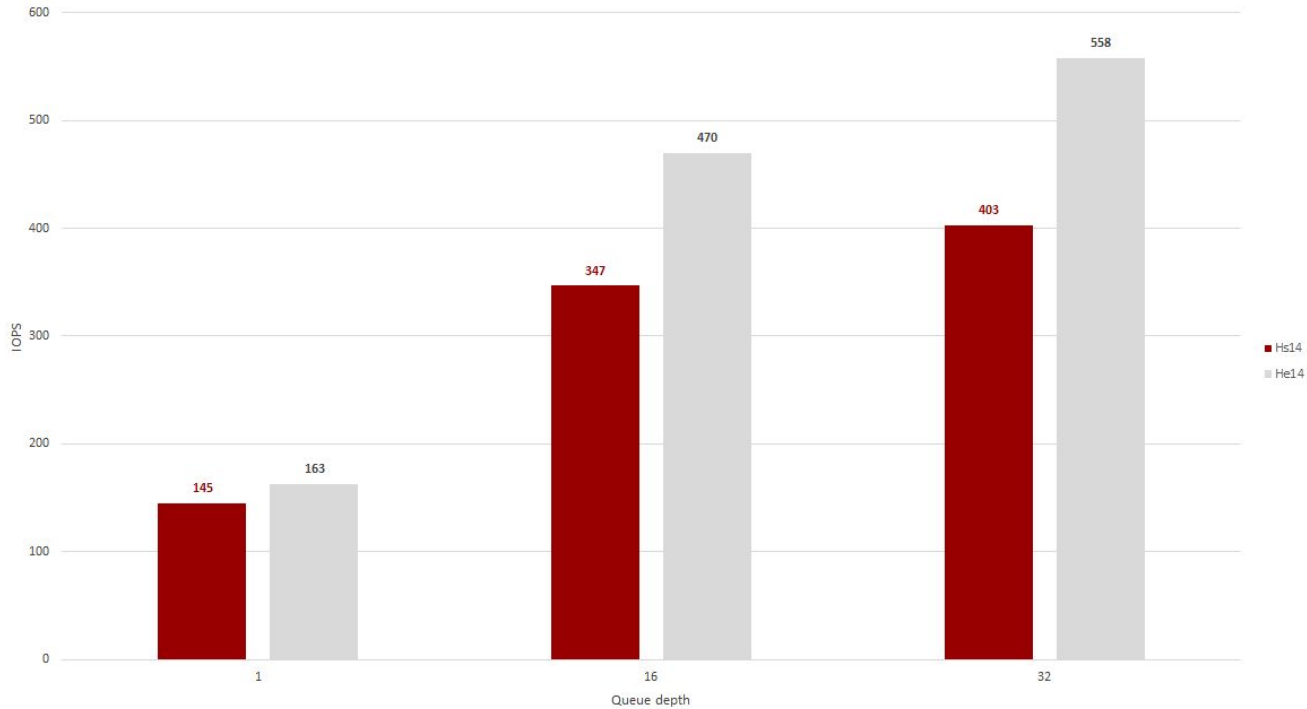
---

# Disk throughput at different LBA offsets Hs14 vs He14



---

## 4KB random read IOPS Hs14 vs He14 (0 to 128GB from OD)





Ceph is a free-software storage platform, implements object storage on a single distributed computer cluster, and provides interfaces for object-, block- and file-level storage. Ceph aims primarily for completely distributed operation without a single point of failure, scalable to the exabyte level, and freely available.

[https://en.wikipedia.org/wiki/Ceph\\_\(software\)](https://en.wikipedia.org/wiki/Ceph_(software))



# Methodology and SW versions

---

# Methodology

OSD bench executed 3 times to capture average and standard deviation values

Rados bench executed once for each thread count, object and operation size combination

XFS and ext4 initialized and benchmarked with He14

Phalanx initialized and benchmarked with Hs14 (data) and Samsung SM843T (metadata)

Single device configuration with disk read & write cache enabled

---

## Methodology (continued)

OSD bench: localized 1GB write in 4MB increments

Rados bench (write): write objects to test pool, 30 mins run time

Rados bench (sequential read): read objects sequentially (in order it was written) from dataset created by write test, 30 mins run time (less if all data have been read)

Rados bench (random read): read objects randomly from dataset created by write test, 30 mins run time

---

# Software

Version

---

Ceph

13.2.6 (7b695f835b03642f85998b2ae7b6dd093d9fbce4) mimic (stable)

---

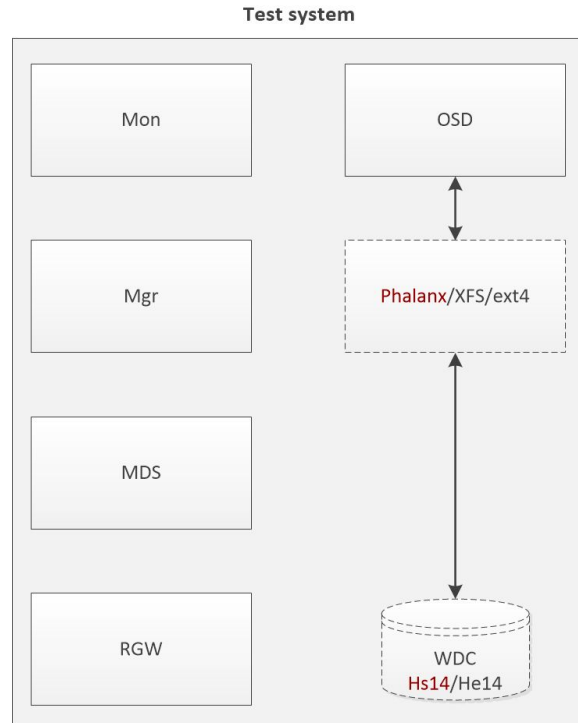
Operating system

Ubuntu 18.04.2 LTS (5.0.0-25-generic)

# Topology and configuration

---

# Ceph cluster topology (single node)



---

# Ceph configuration

	Value	Description
pg-num	100	Number of placement groups for test pool.
pgp-num	100	Number of placement groups for placement.
osd pool default size	1	Number object replicas in test pool.
osd crush chooseleaf type	0	CRUSH leaf type (0 = OSD).



# Configuration references

## Links

---

<https://docs.ceph.com/docs/mimic/rados>

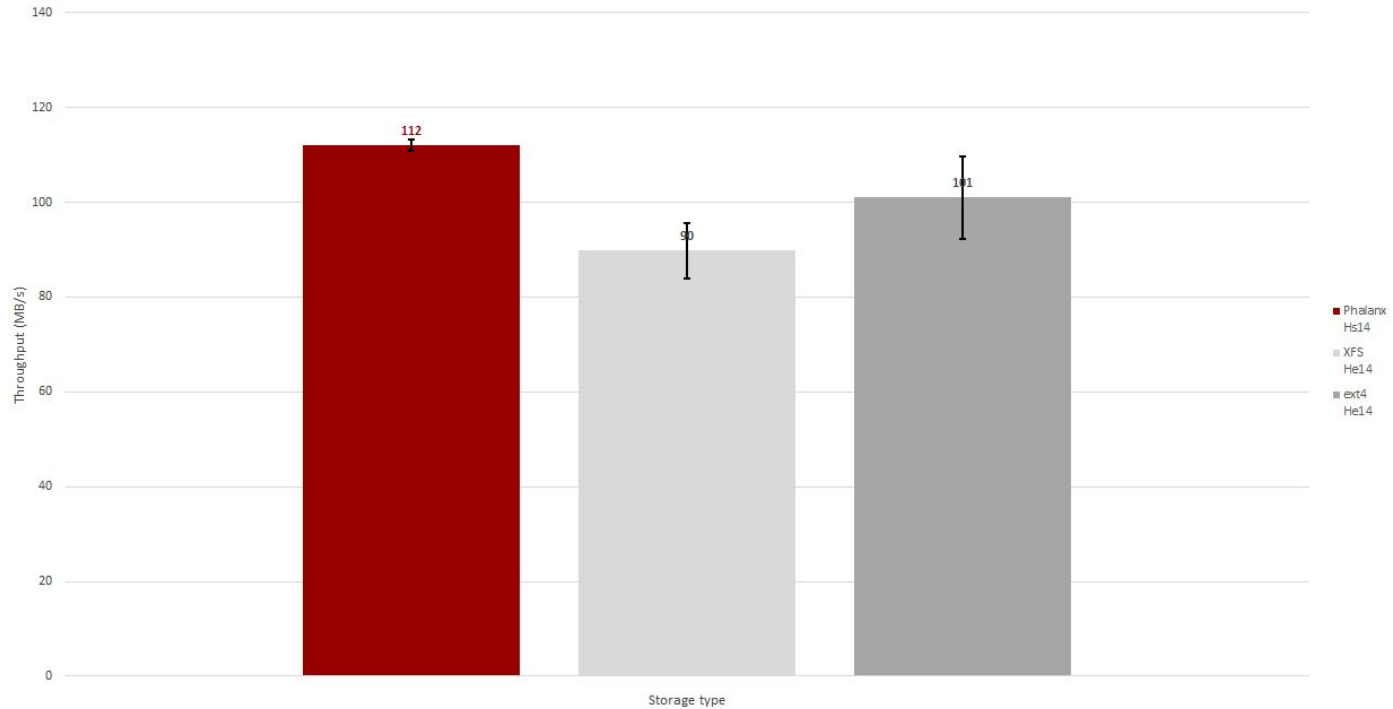
<https://docs.ceph.com/docs/mimic/rados/operations/pools>

<https://docs.ceph.com/docs/mimic/rados/configuration/pool-pg-config-ref>



# OSD bench

# OSD bench results (1GB write in 4MB increments)

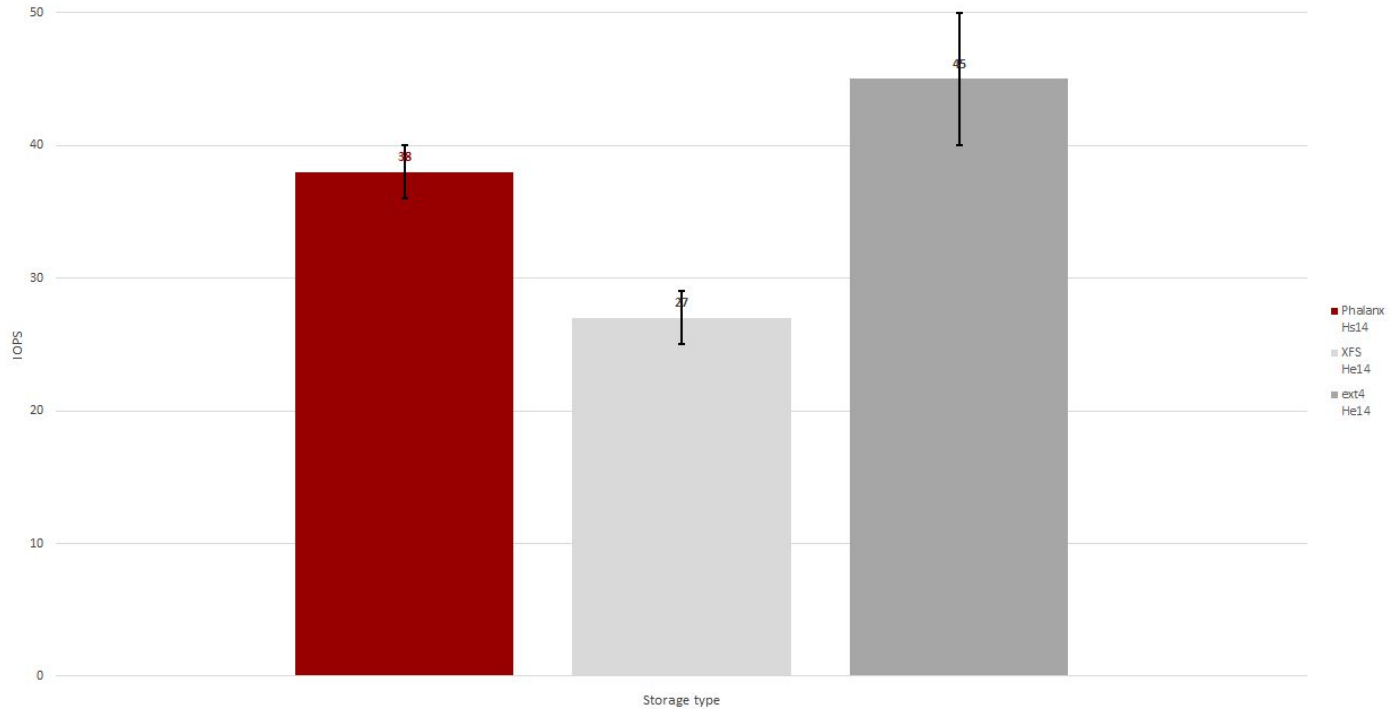


# Rados bench read

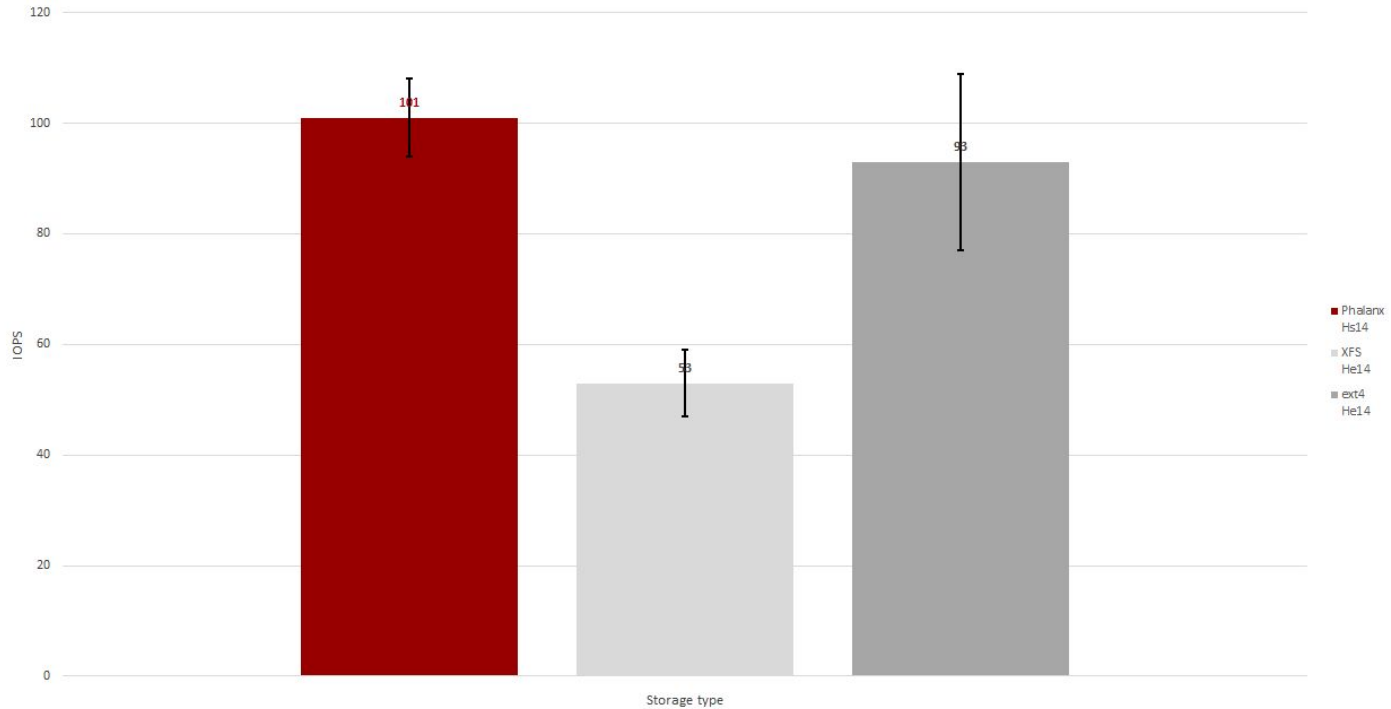
# Rados bench read configurations

Threads	Access type	Object size	Operation size
16	Sequential	4MB	4MB
64	Sequential	1MB	1MB
16	Random	4MB	4MB
64	Random	1MB	1MB

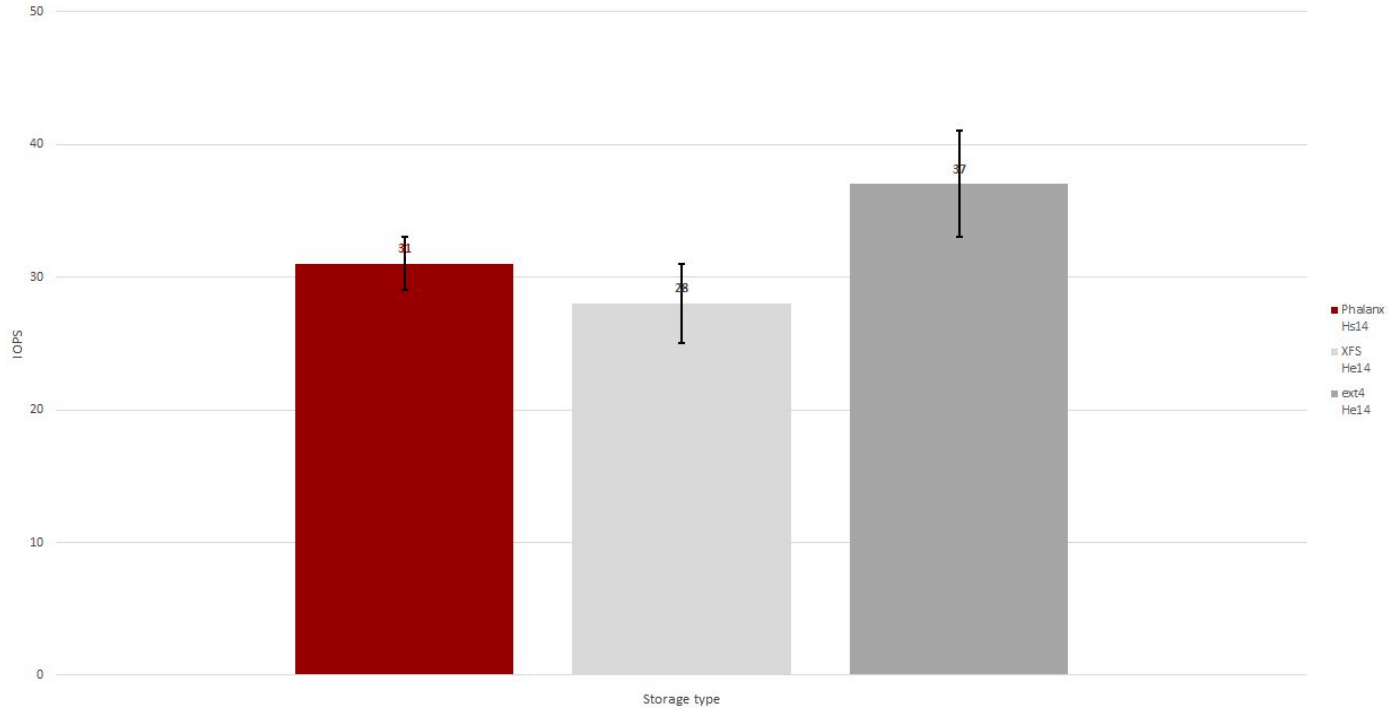
## Rados seq read results (4MB obj size, 4MB op size, 16 threads)



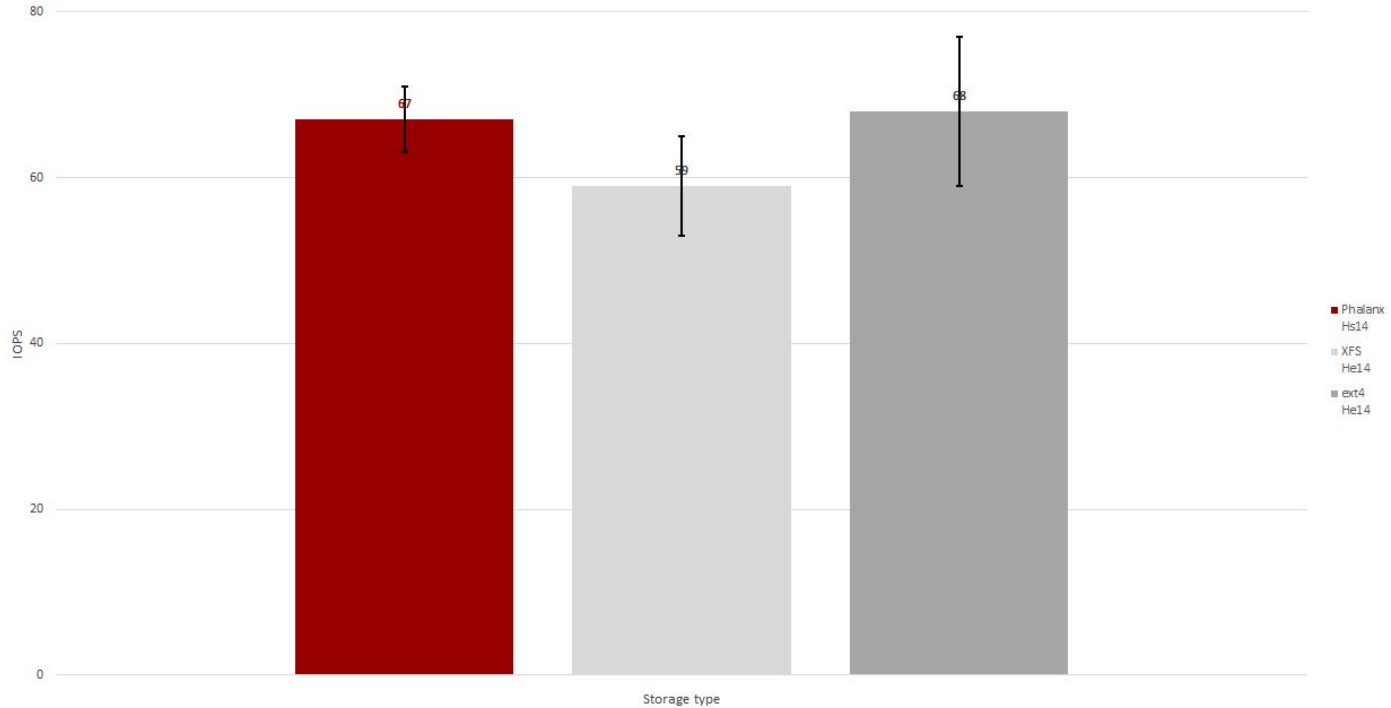
## Rados seq read results (1MB obj size, 1MB op size, 64 threads)



# Rados rnd read results (4MB obj size, 4MB op size, 16 threads)



# Rados rnd read results (1MB obj size, 1MB op size, 64 threads)





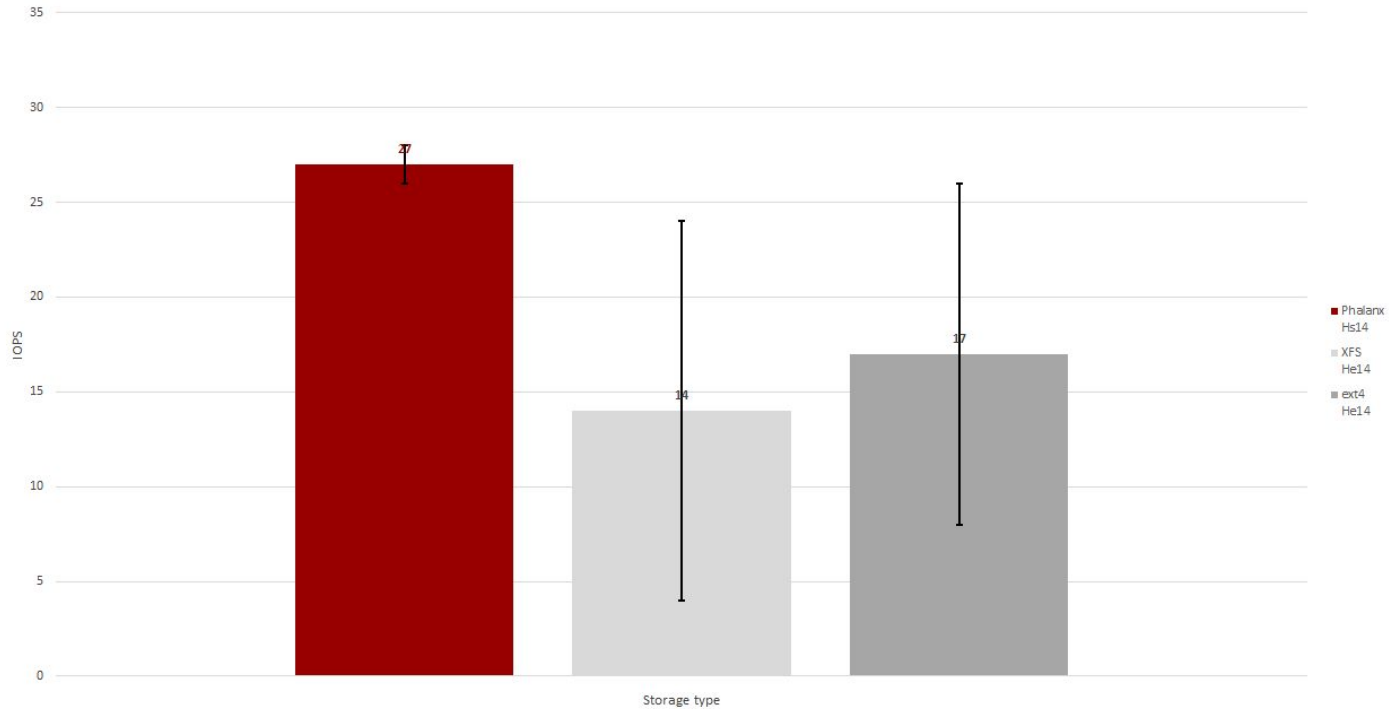
# Rados bench write



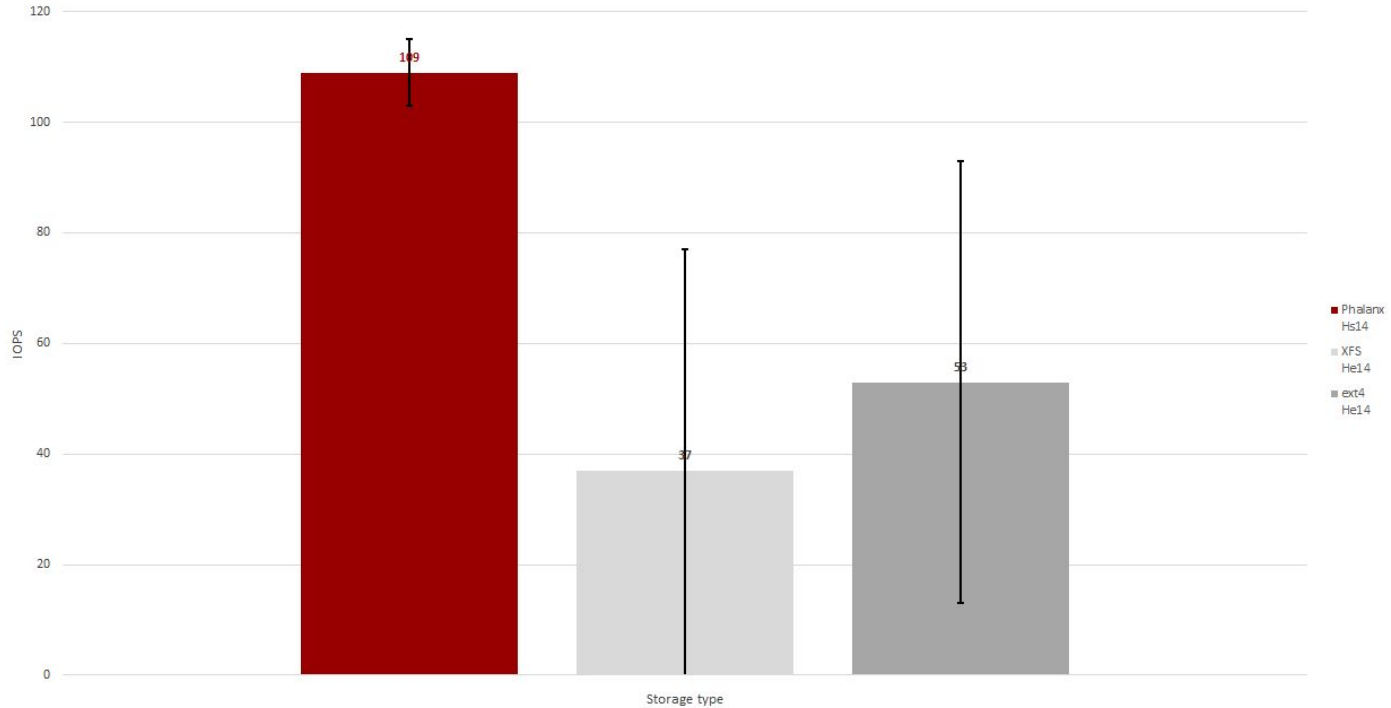
# Rados bench write configurations

Threads	Object size	Operation size
16	4MB	4MB
64	1MB	1MB

## Rados write results (4MB obj size, 4MB op size, 16 threads)



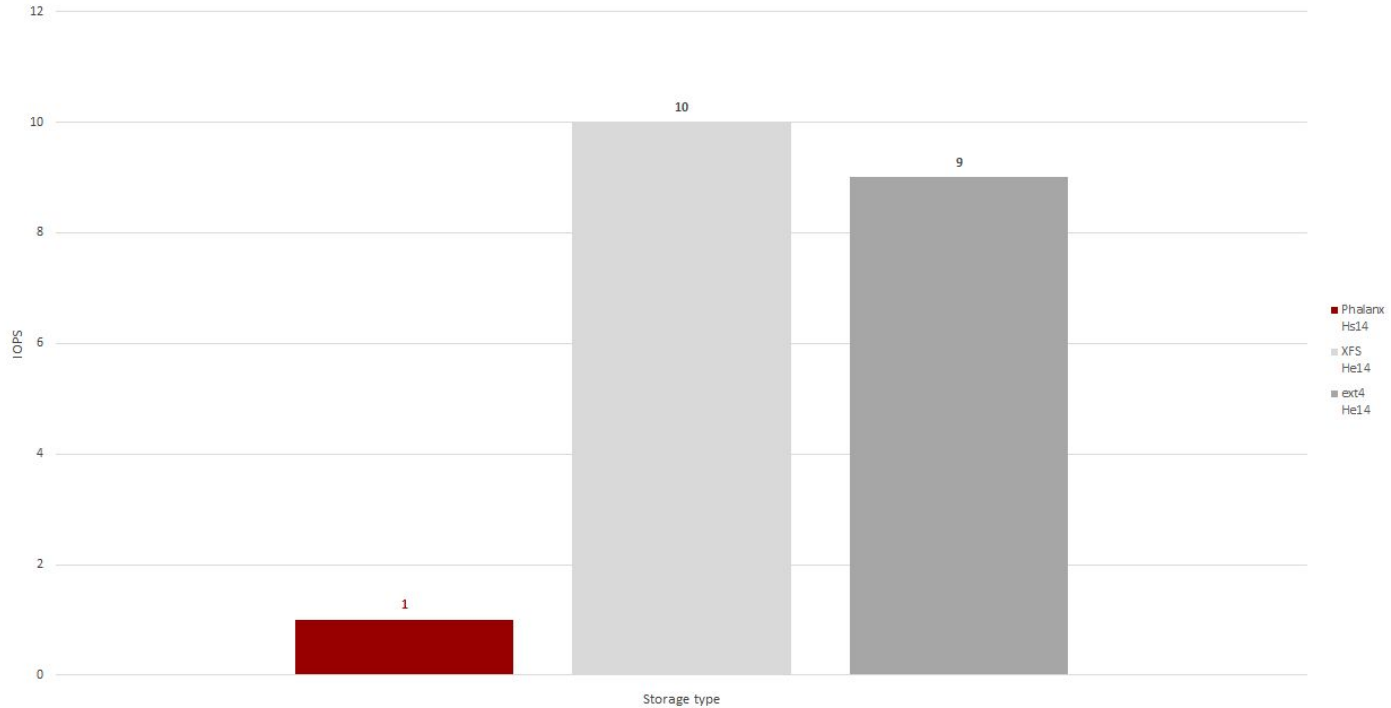
## Rados write results (1MB obj size, 1MB op size, 64 threads)



**Write performance variability  
(lower is better)**

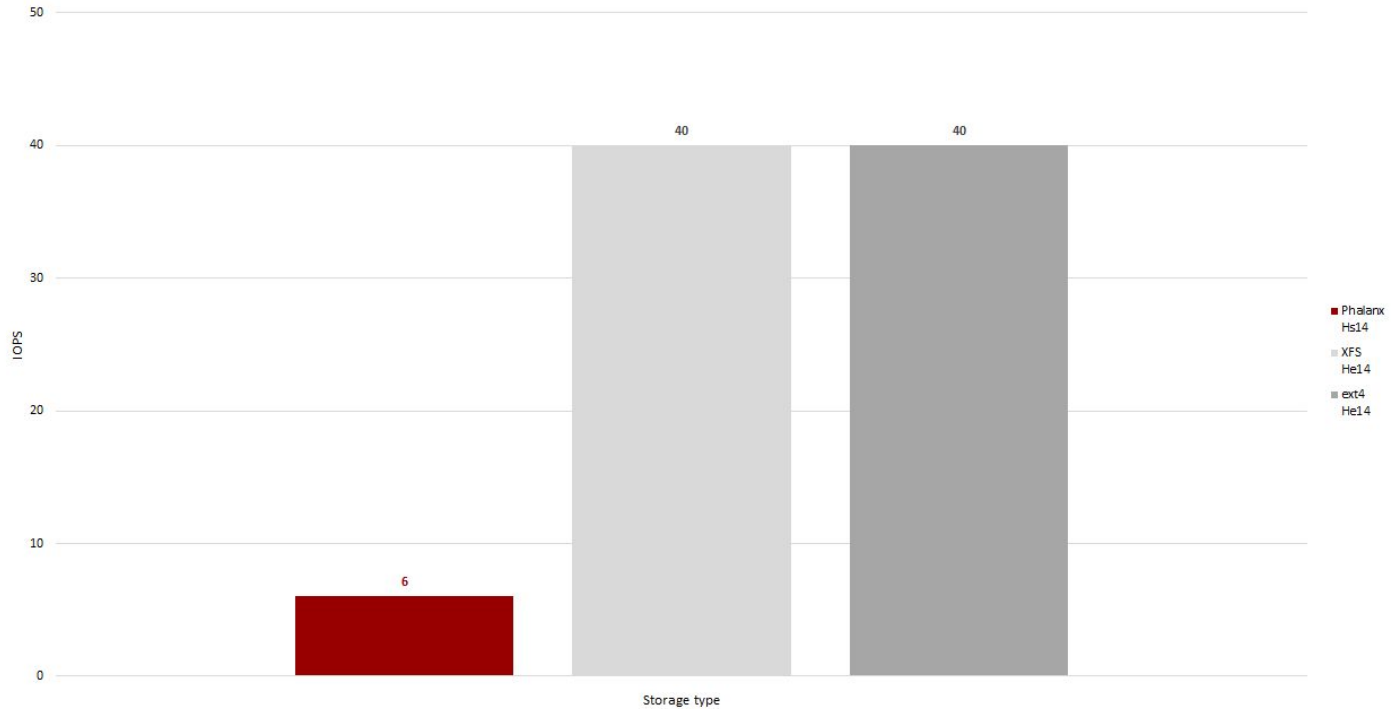
---

## Std dev (Rados write, 4MB obj size, 4MB op size, 16 threads)



---

## Std dev (Rados write, 1MB obj size, 1MB op size, 64 threads)



---

# Contact

<http://www.kalista.io>

@kalista.io

info@kalista.io