Accelerate Ceph By SPDK on AArch64

Jun He, jun.he@arm.com

Tone Zhang, tone.zhang@arm.com

2018/3/9

© 2018 Arm Limited

SPDK

© 2018 Arm Limited

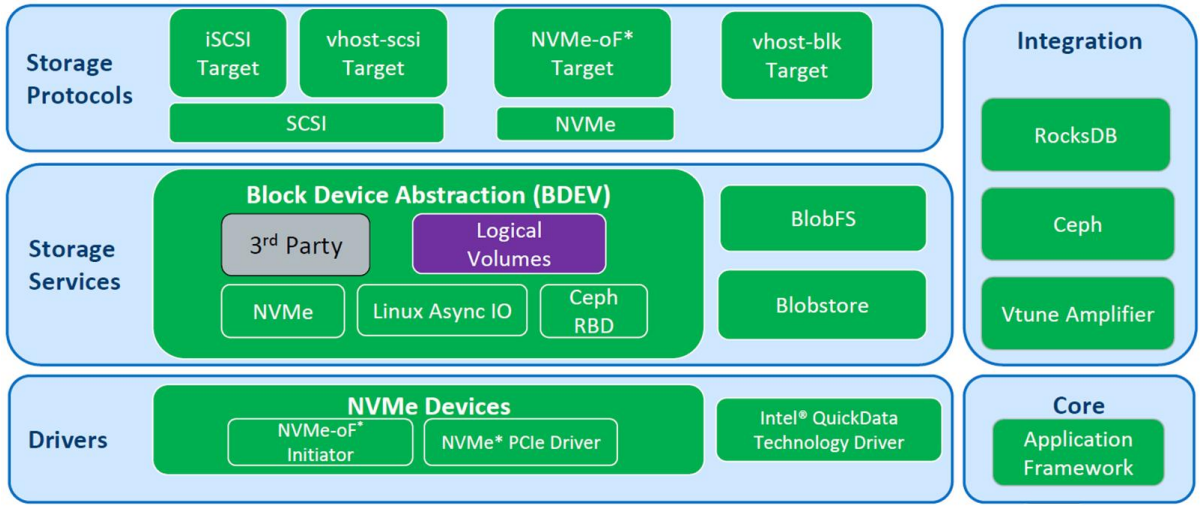
arm

SPDK

What's SPDK?

- Storage Performance Development Kit
- A set of tools and libraries to create high performance, scalable, user mode storage applications
- Designed for new storage HW devices (NVMe). Can achieve millions of IOPS per core. Better tail latency.

Architecture diagram



© 2018 Arm Limited

SPDK on AArch64

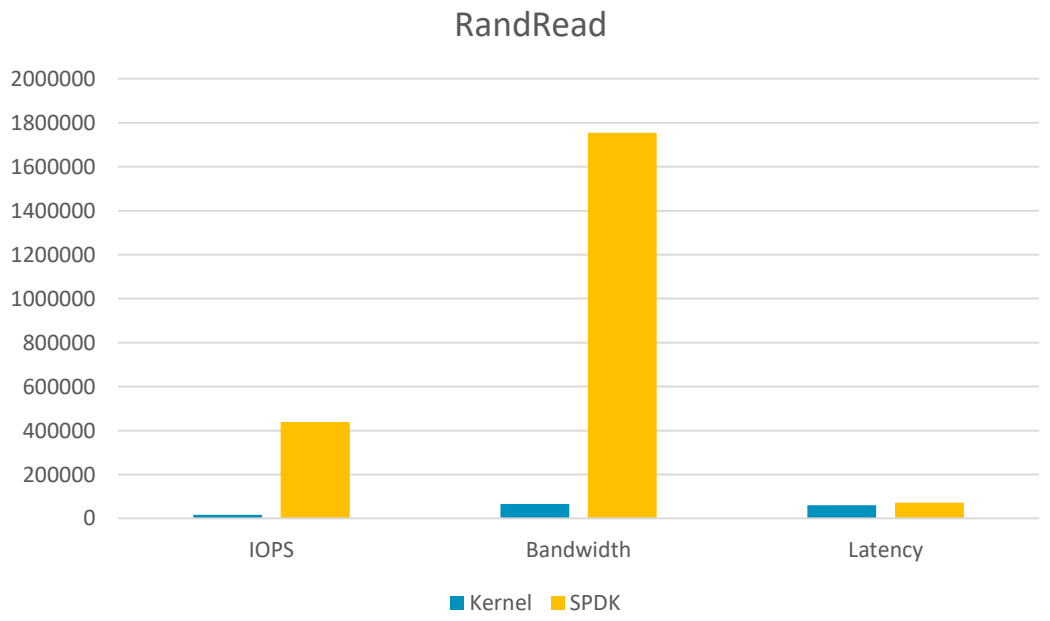
- Several ARM related patches are merged
  - Memory_barrier
  - VA address space
- 17.10 release verified
  - Kernel: 4.11, 48bit/42bit VA, 4KB pagesize
  - UIO/VFIO

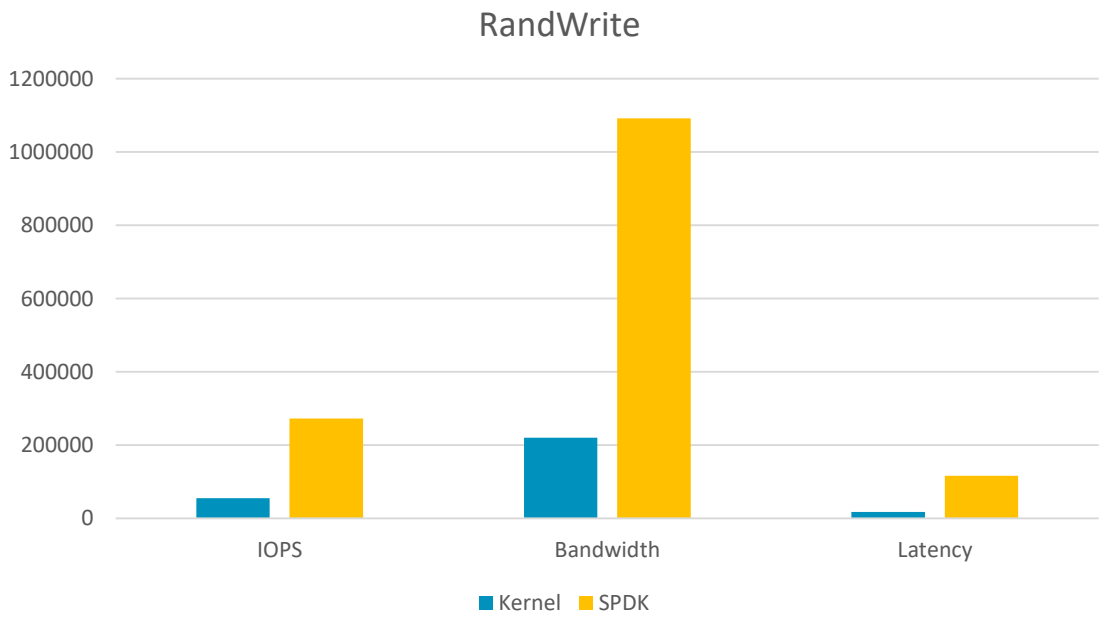
© 2018 Arm Limited

SPDK Performance on AArch64

- SPDK perf
  - UIO/4K pagesize
- FIO

```
Starting thread on core 0
========================================================
                                                                                Latency(us)
Device Information                                  :       IOPS       MB/s    Average        min        max
INTEL SSDPEDMW012T4  (CVCQ6510001Z1P2JGN  ) from core 0:  291310.34    1137.93     439.35       5.75    8326.70
========================================================
Total                                               :  291310.34    1137.93     439.35       5.75    8326.70
```



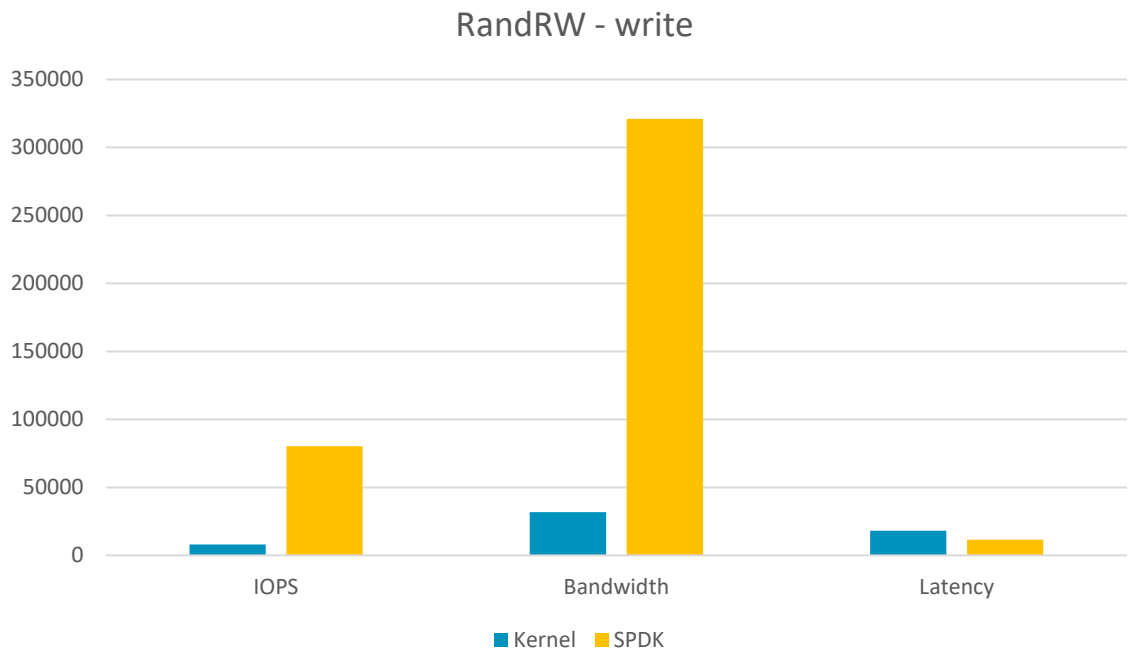


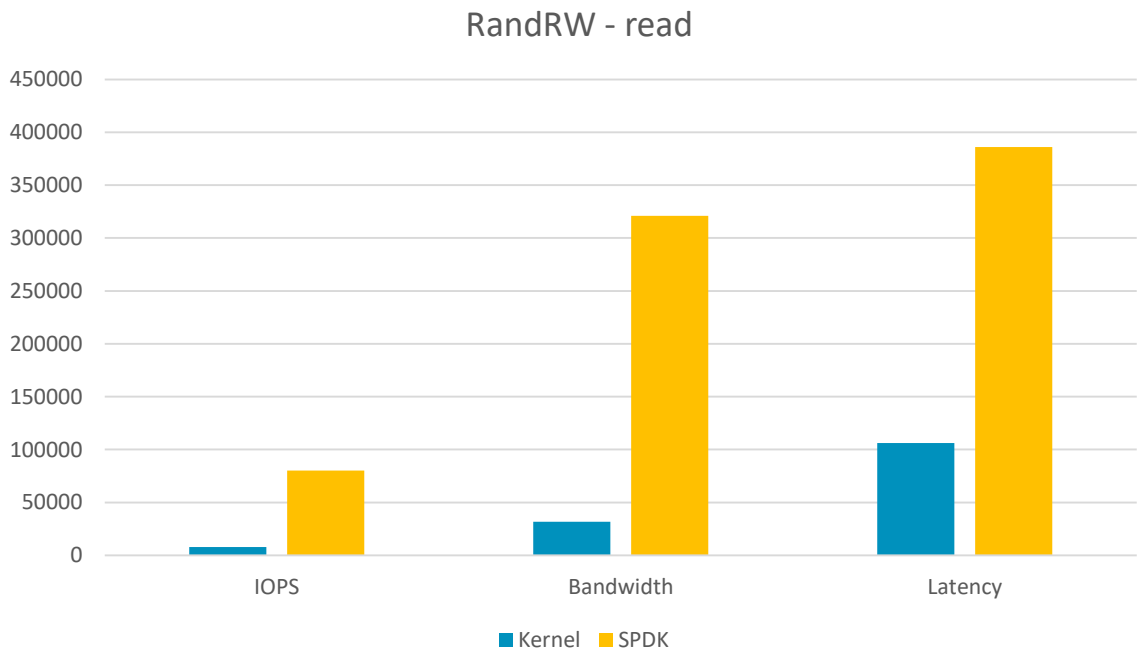
FIO configuration: direct=1, bs=4096, rwmixread=50, iodepth=32, ramp=30s, run_time=180s, jobs=1

© 2018 Arm Limited

SPDK Performance on AArch64

- FIO





© 2018 Arm Limited

FIO configuration: direct=1, bs=4096, rwmixread=50, iodepth=32, ramp=30s, run_time=180s, jobs=1

What’s the next?

- Optimization with ASIMD and Crypto extensions
- Tuning with different page-size(16KB/64KB)
- Cache strategy improvement for better read/write performance

© 2018 Arm Limited

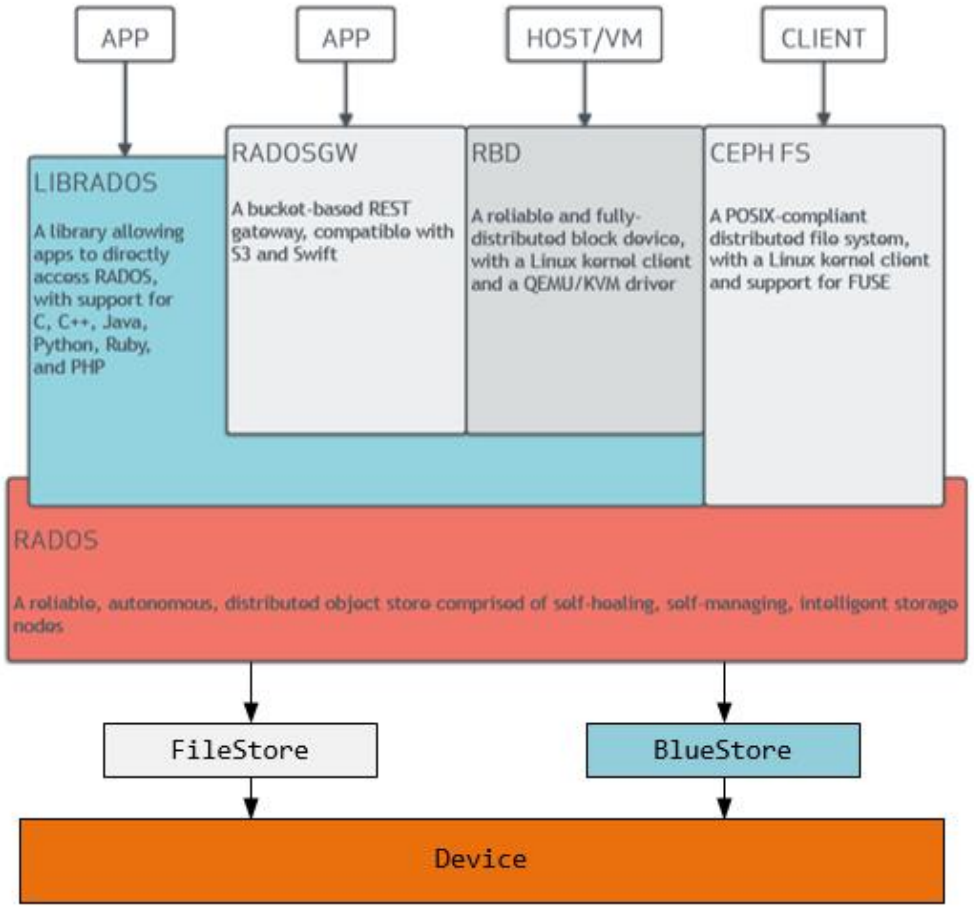
Ceph

© 2018 Arm Limited

Ceph

What's Ceph?

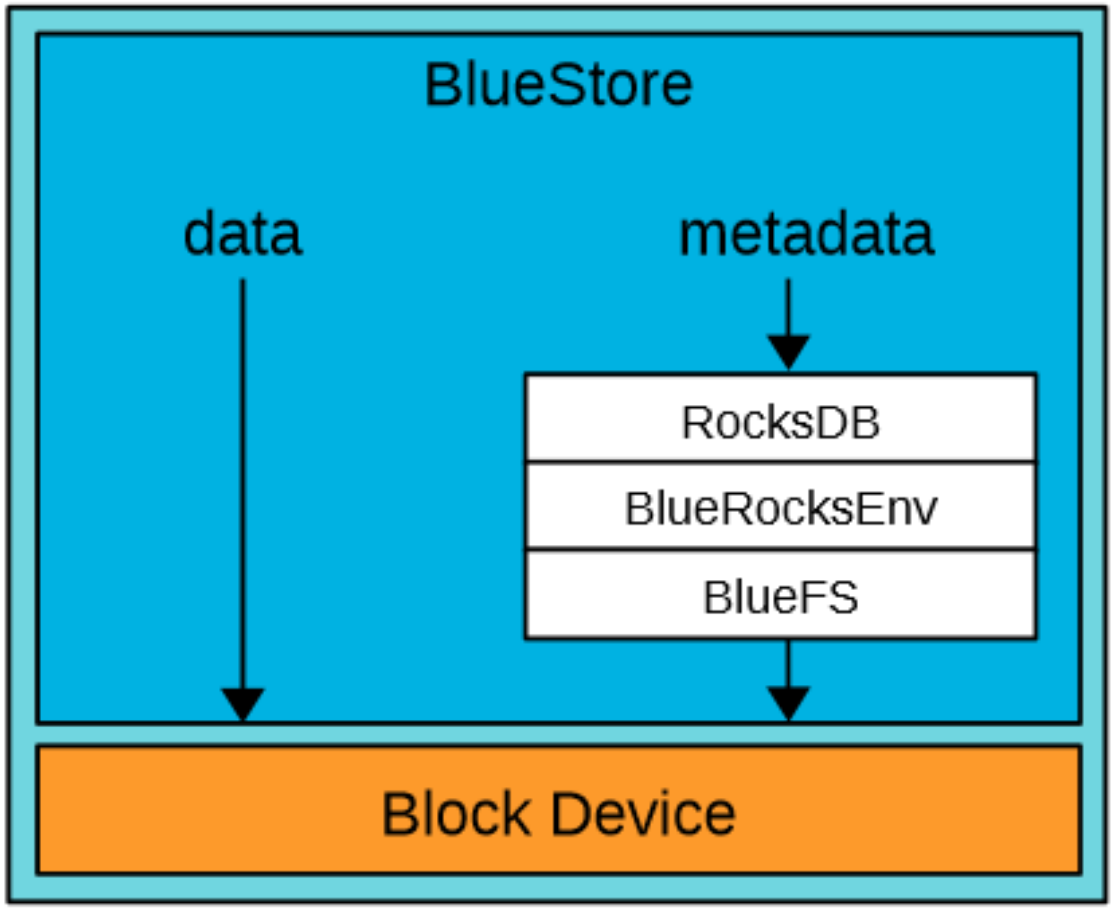
- Ceph is a unified, distributed storage system designed for excellent performance, reliability and scalability
- Ceph can supply following services
  - Object storage
  - Block storage
  - File system
- The backend storage types
  - FileStore
  - BlueStore



© 2018 Arm Limited

BlueStore

BlueStore is a new storage backend for Ceph.



- Full data built-in compression
- Full data checksum
- Boasts better performance
  - Get rid of file system, and write all data to RAW device via asynchronous libaio infrastructure

© 2018 Arm Limited

Ceph on AArch64

- Has already been integrated with OpenStack
- Has been validated and released by Linaro SDI team
- Has committed many patches to fix the functional faults and improve the performance
- Has validated “Ceph + SPDK” on top of NVMe devices
- Tuned Ceph performance on AArch64

© 2018 Arm Limited

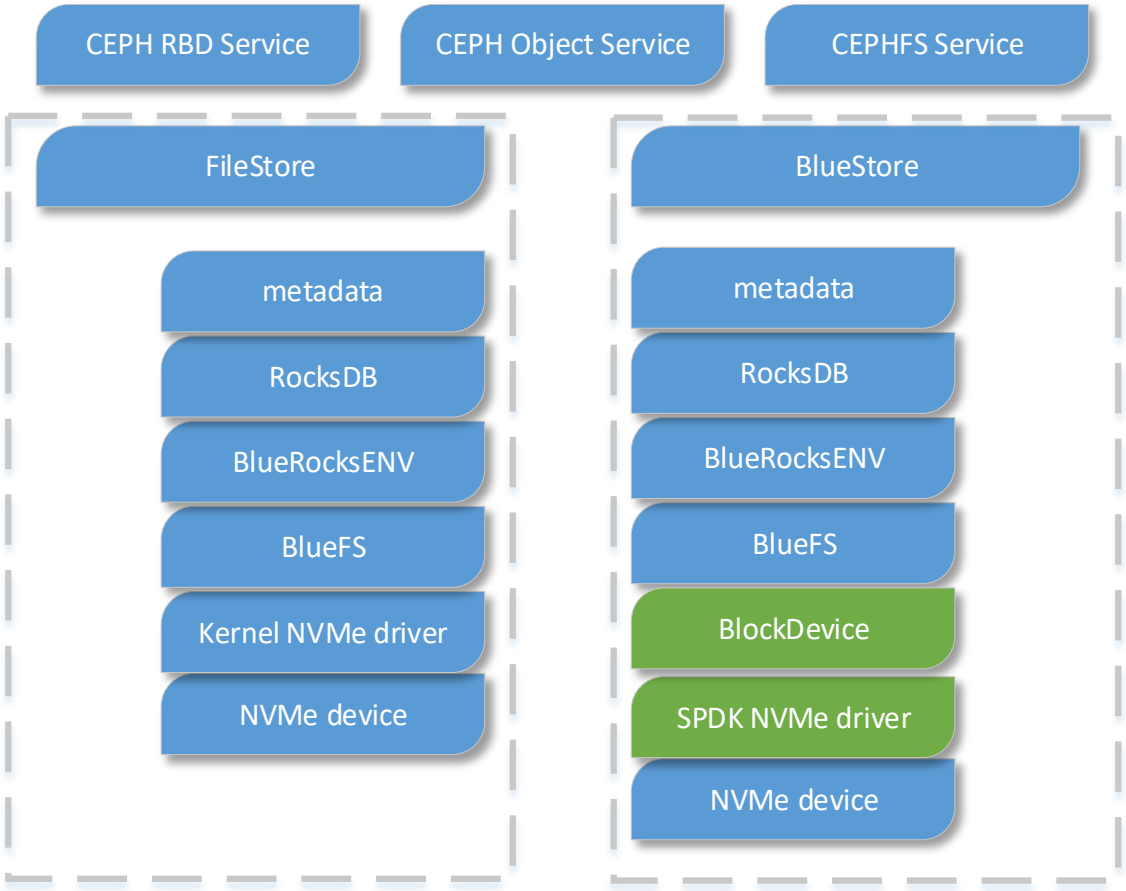
Ceph + SPDK on AArch64

- Dependencies
  - NVMe device
  - SPDK/DPDK
  - BlueStore
- Enabled SPDK in Ceph on AArch64
- Extended virtual address map bits from 47 to 48 bits in DPDK

© 2018 Arm Limited

Ceph + SPDK on AArch64

BlueStore is a new storage backend for Ceph.



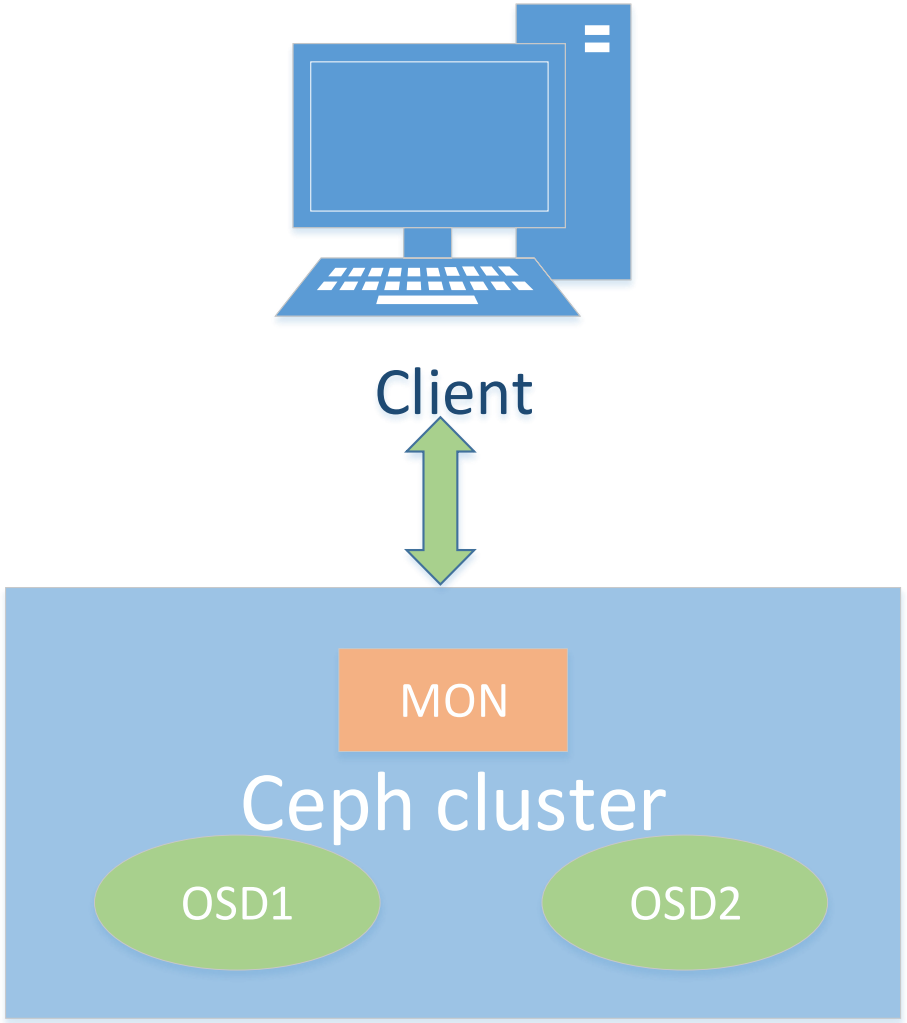
- BlueStore can utilize SPDK
- Replace kernel driver with SPDK user space NVMe driver
- Abstract BlockDevice on top of SPDK NVMe driver

© 2018 Arm Limited

Ceph + SPDK Performance test on AArch64

Test case

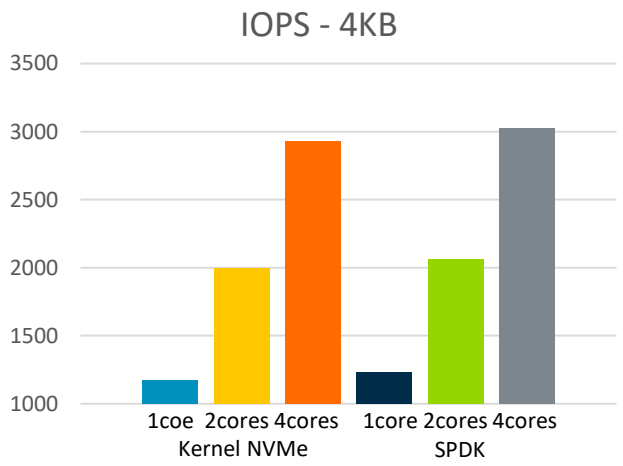
- Ceph cluster
  - Two OSD, one MON, no MDS and RGW
  - One NVMe card per OSD
  - CPU: 2.4GHz multi-core
- Client
  - CPU: 2.0GHz multi-core
- Test tool
  - Fio (v2.2.10)
- Test case:
  - Sequential write with different block_size (4KB, 8KB and 16KB)
  - 1 and 2 fio streams

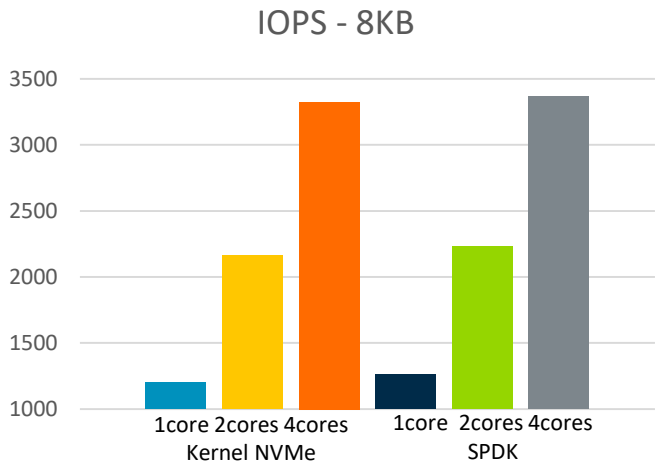


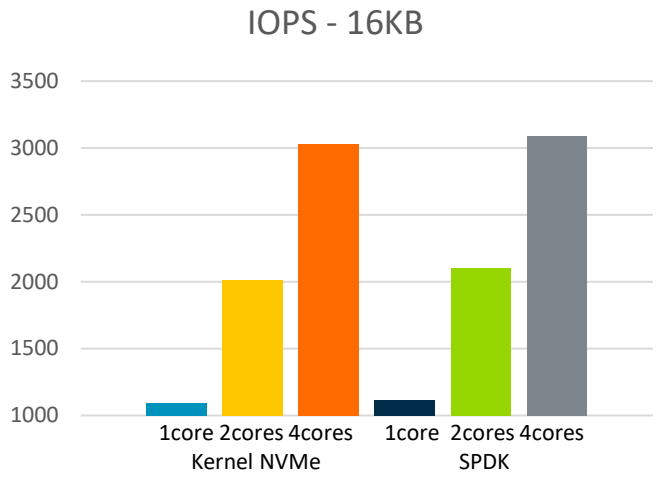
© 2018 Arm Limited

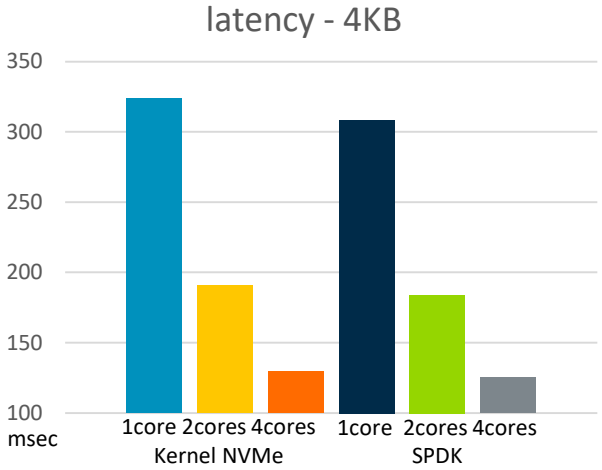
Write performance result

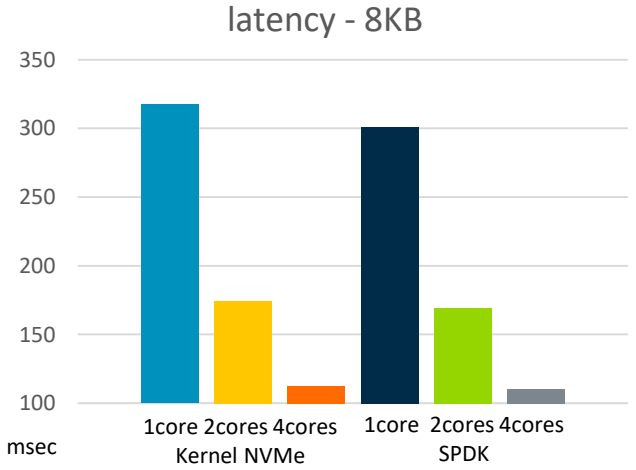
1 stream

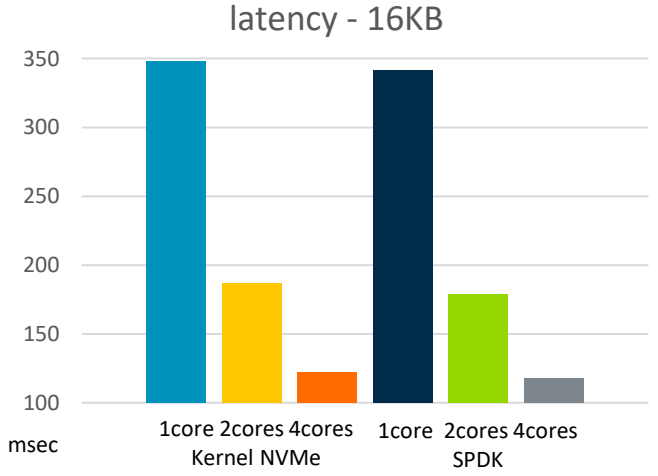










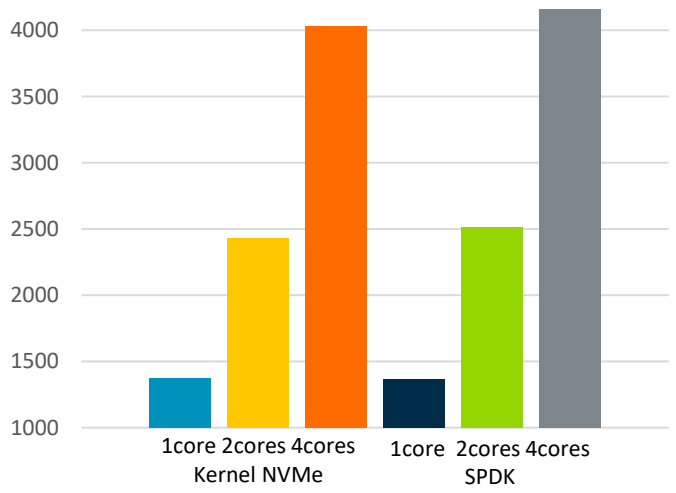


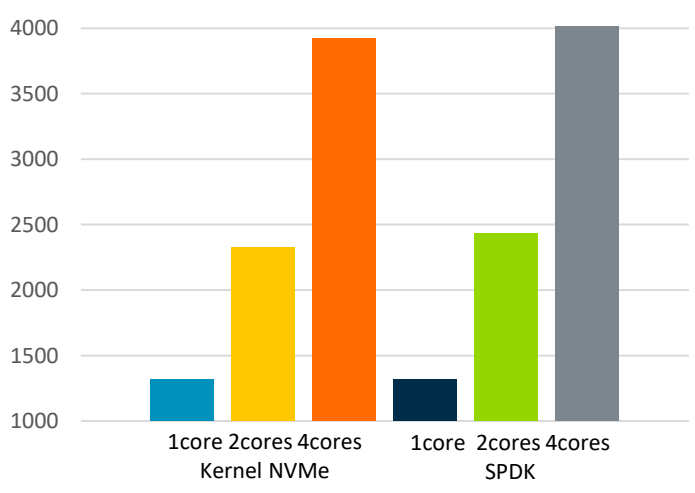
1 fio stream, FIO configuration: bs=4K/8K/16K, rw=write, iodepth=384, run_time=40s, jobs=1, ioengine=rbd

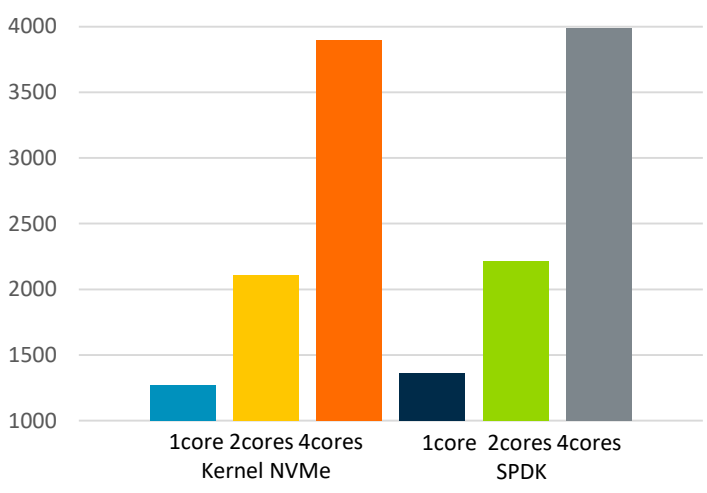
© 2018 Arm Limited

Write performance result

2 streams







2 fio streams, FIO configuration: bs=4K/8K/16K, rw=write, iodepth=384, run_time=40s, jobs=1, ioengine=rbd

© 2018 Arm Limited

Performance improvement

SPDK accelerated Ceph in below:

- More IOPS
- Lower latency
- Linear scaling associate with the number of CPU cores

© 2018 Arm Limited

What's the next?

- Continue improving Ceph performance on top of SPDK
  - Enable NVMe-OF and RDMA
  - Enable zero-copy in Ceph
  - Simplify the locking in Ceph to improve the OSD daemon performance
  - Switch PAGE_SIZE to 16KB and 64KB to improve the memory performance
  - Modify NVMEDEVICE to improve its performance associate with different PAGE_SIZE

© 2018 Arm Limited

Thank You
Danke
Merci
谢谢
ありがとう
Gracias
Kiitos
감사합니다
धन्यवाद
תודה

arm

© 2018 Arm Limited