WHITEPAPER

NVMe User-space Driver with Memblaze® PBlaze SSD

Exclusive Summary

User-space NVMe drive has been receiving a lot of attention and integrated in applications to accelerate performance. This whitepaper demonstrates how to obtain millions of IOPS with user-space NVMe driver by using a few processor cores.

NVMe Driver Mode Introduction

Kernel-space driver: uses a MSI-X interrupt to wake the OS thread up when there are completion entries in completion queue occurs.

User-space driver: It also open source drive, integrate in Storage Performance Developer Kit (SPDK). Some of the advantages to the Kernel-space driver as bellow:

- Running at user level and kernel bypass, which avoids the context switching associated with jumping from kernel-space to user-space.
- Wording in Polled Mode, which CPU is constantly reading the tail of the completion queue rather than using MSI-X interrupts.

Test Configuration

Perf in SPDK package, it use to be benchmark SPDK read and write performance. The test environment as following:

CPU:	Dell PowerEdge R720x 2 socket Intel XeonE5-2630(6 cores)
Memory:	128GB
SSD:	3 x Memblaze 1.6 T PBlaze4
Linux:	CentOS 7.0
Benchmark Tool:	User-space SPDK perf, Kernel-space fio

Benchmark Procedure

1. Follow https://github.com/spdk/spdk to setup SPDK environment.
2. Pre-conditioning SSD to make sure SSD in stable state.

```
examples/nvme/perf/perf  -q 128 -w randwrite -s 4096 -t 14400
```

3. Perf Random Write benchmark test, remember to bind perf to near CPU core with iodepth 1, 4 … 128

```
examples/nvme/perf/perf  -q $iodep -w randwrite -s 4096 -t 1800 –c 2
```

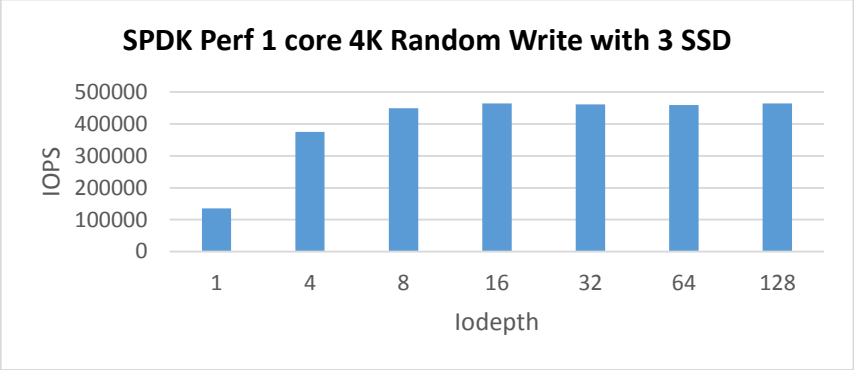

Figure 1, Perf Random Write Throughput

As Figure 1 illustrates, 1 CPU core achieves 123,565 IOPS with iodepth 1, and 447,067 IOPS with 128 iodepth.

4. Perf Random Read benchmark test, remember bind perf to near CPU core with iodepth 1, 4, ... 128

```
examples/nvme/perf/perf  -q $iodep -w randread -s 4096 -t 1800 –c 2
```

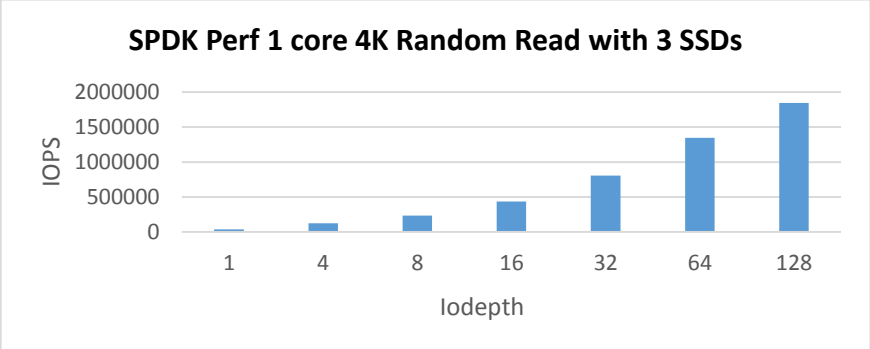

Figure 2, Perf Random Read Throughput

As shown in Figure 2, 1 core achieves 33,518 IOPS with 1 iodepth, and 172,3306 IOPS with 128 iodepth.

5. Fio and SPDK perf Random Read 4k benchmark test with 1 numjob and 128 iodepth.

```
fio --rw=randwrite --iodepth=128 --numjobs=1 --bs=4k --ioengine=libaio  --name= randwrite --
filename=/dev/nvme0n1   --filename=/dev/nvme1n1  --filename=/dev/nvme2n1 --direct=1 --runtime=1800 --
```

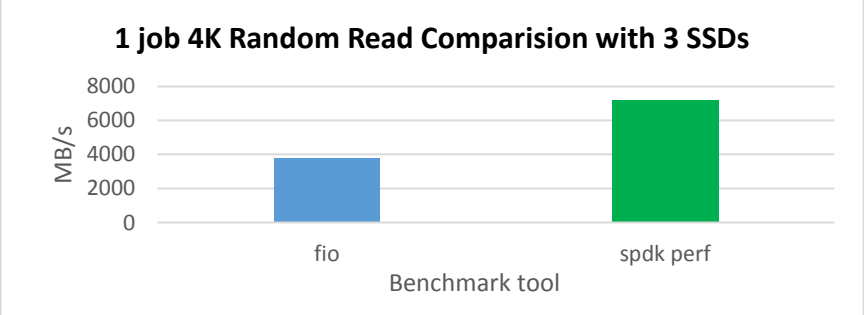

Figure 3, Fio and SPDK Perf Throughput Comparison

As shown in Figure 3, 1 job fio throughput is 3776MB/s, while SPDK perf throughput reaches up to 7208 MB/s.

6. Fio Random Read 4k benchmark test with 1, 6 numjob and 128 iodepth, test case example as below:

```
fio --rw=randwrite --iodepth=128 --numjobs=1 --bs=4k --ioengine=libaio  --name= randwrite --
filename=/dev/nvme0n1   --filename=/dev/nvme1n1  --filename=/dev/nvme2n1 --direct=1 --runtime=1800 --
group_reporting
```

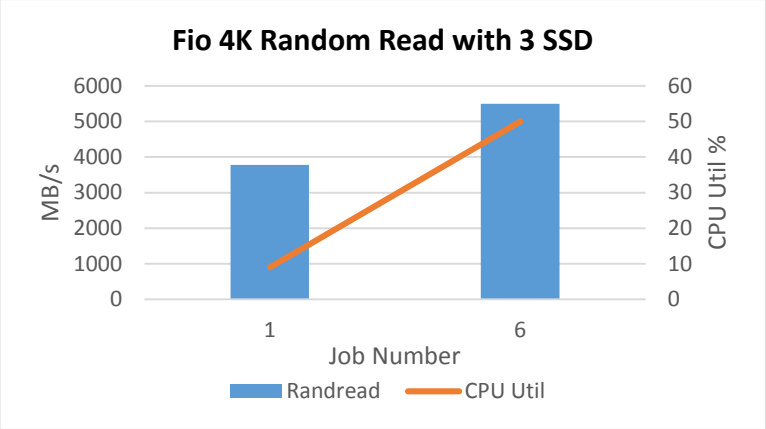

Figure 4, Fio Throughput and CPU Util

As shown in Figure 4, fio throughput linearly growth with job number, and CPU utilization also linearly growth with job number. Fio 6 numjobs throughput is 5498 MB/s, CPU utilization is 50%. However, kernel-space driver, SPDK perf throughput can reache up to 7208 MB/s with 1 numjob 1 core, and CPU utilization is only 1core / 12 cores = 8.3%.

Conclusions

Comparing with Kernel-space driver with 1 numjobs, User-space driver achieves preferable performance. User-space driver can obtain millions of IOPS with one core by utilizing Memblaze PBlaze SSD. In addition, User-space driver occupies less CPU resource than Kernel-space driver at the same throughput.

DISCLAIMER

Information in this document is provided in connection with Memblaze products. Memblaze provides this document “as is”, without warranty of any kind, neither expressed nor implied, including, but not limited to, the particular purpose. Memblaze may make improvements and/or changes in this document or in the product described in this document at any time without notice. The products described in this document may contain design defects or errors known as anomalies or errata which may cause the products functions to deviate from published specifications.

COPYRIGHT

©2015 Memblaze Corp. All rights reserved. No part of this document may be reproduced, transmitted, transcribed, stored in a retrieval system, or translated into any language in any form or by any means without the written permission of Memblaze Corp.

TRADEMARKS

Memblaze is a trademark of Memblaze Corporation. Other names mentioned in this document are trademarks/registered trademarks of their respective owners.

USING THIS DOCUMENT

Though Memblaze has reviewed this document and very effort has been made to ensure that this document is current and accurate, more information may have become available subsequent to the production of this guide. In that event, please contact your local Memblaze sales office or your distributor for latest specifications before placing your product order.