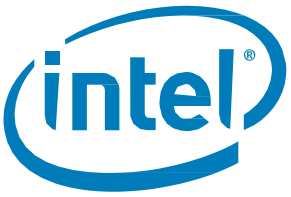


Write Pressure Simulation with Intel® Optane™ SSDs and NAND-based SSDs

Application Note

January 2018
Revision 001

Document Number: 336789-001US

Revision History

Revision	Description	Date
001	• Initial release.	January 2018

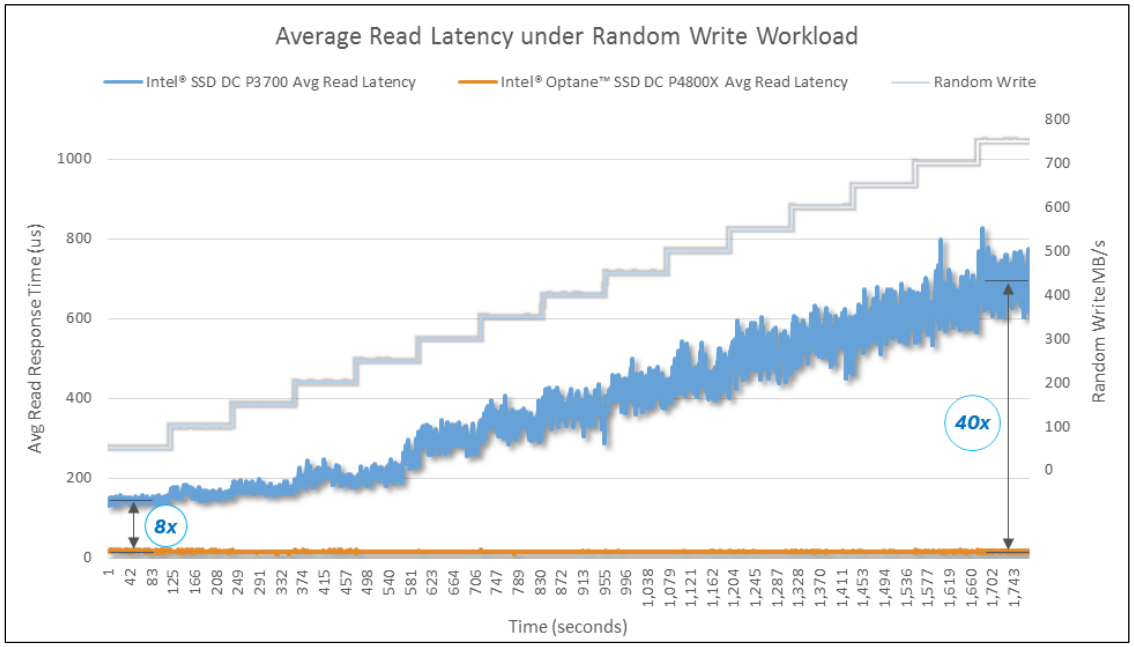
1 Introduction

Imagine moving a large volume of new data into a currently active environment; for example, spinning up a number of new VMs concurrently, while the existing VM is still managing user requests against an existing database. This happens regularly in real production environments, especially as IT environments move to Cloud-like multi-tenant models. This example illustrates typical conditions of "responsiveness under load" which is also called "write pressure".

This application note examines the simulated results of such an example. It analyzes the difference between a class of Solid State Drives, and clearly demonstrates the significant write workload benefits of Intel® Optane™ SSDs.

In Figure 1, the chart's right axis indicates the "offending write workload" burden being placed on the SSD; represented by the top grey stair step line. In this test, the SSDs were presented with a gradually increasing, random write workload—starting at 100MB/s and increasing to more than 700MB/s.

Figure 1: Read Latency Comparison



1. Responsiveness defined as average read latency measured at queue depth 1 during 4k random write workload. Measured using FIO 2.15*. Common Configuration - Intel 2U Server System, OS CentOS 7.2, kernel 3.10.0-327.el7.x86_64, CPU 2 x Intel® Xeon® E5-2699 v4 @ 2.20GHz (22 cores), RAM 396GB DDR @ 2133MHz. Configuration –375GB Intel® Optane™ SSD DC P4800X Series and 1600GB Intel® SSD DC P3700 Series. Latency – Average read latency measured at QD1 during 4K Random Write operations using FIO 2.15.

The chart's left axis shows the read response time from a single-thread random read workload.

The blue line represents the performance of the NAND-based 1.6TB Intel® SSD DC P3700, as the offending "write pressure" increases. As you can see, as the write pressure increases, the average read response time of the NAND-based DC P3700 also increases, significantly. Because of the many background operations of these workloads, NAND-based SSDs cannot efficiently manage complex random write operations, therefore latency response is significantly impacted.

The orange line represents the performance results of the same workload using our new 3D XPoint™ technology-based 375GB Intel® Optane™ SSD DC P4800X. The lower number indicates faster, therefore better, response time.

Unlike the decreased efficiency of the NAND-based SSD, the Intel® Optane™ SSD DC P4800X Series maintains consistent read response time regardless of the increased write pressure applied to the SSD. The Intel® Optane™ SSD provides more than 40 times improvement in read responsiveness as compared to the NAND-based SSD.

2 Simulating the Workload with FIO*

The following method of utilizing FIO* enables write pressure testing, and produces results similar to those presented in Figure 1. Each write pressure test step requires an independent FIO script; each step must be run independently, in sequence, and then the overall data merged. The script is designed to run eight parallel 8K random write workloads, and one 8K random read workload. Then the latency output (write_lat_log) is only analyzed for random read workload. The percentile determines maximum latency and Quality of Service (QoS) range.

FIO script:

```
[global]
thread
direct=1
group_reporting
norandommap=1
randrepeat=0
refill_buffers
ioengine=libaio
gtod_reduce=0
time_based
blocksize=8k
runtime=60
log_avg_msec=500
[Random_Write]
rw=randwrite
numjobs=8
iodepth=8
write_bw_log=nvme0n1_WPT_500MBs
rate=,62500k
filename=/dev/nvme0n1

[Random_Read]
rw=randread
numjobs=1
iodepth=1
write_lat_log=nvme0n1_WPT_500MBs
percentile_list=99:99.9:99.99:99.999:99.9999:99.99999:99.999999:99
.9999999
filename=/dev/nvme0n1
```

In the example above, 500MB write pressure is simulated with FIO parameter “rate=” which is specified per job. Because the script is running 8 write jobs, the number of MB must be divided by 8, resulting in “rate=,62500k”.

Download the remaining FIO scripts at:

https://github.com/intel/fiovisualizer/tree/master/Workloads