

Hands-on Lab: How to Unleash Your Storage Performance by Using NVM Express™ Based PCI Express® Solid-State Drives

Andrey Kudryavtsev – SSD Solution Architect, Intel Corporation
Zhdan Bybin – Application Engineer, Intel Corporation

SSDL001

Agenda

- What is NVM Express™ (NVMe) ?
- How do I know my NVM Express drive is working?
- How do I measure performance of my NVM Express drive?
- Wow! I have 2 NVMe drives in my system! Now how do I RAID those?
- How can I monitor NVMe performance and debug application performance issues?
- Alright, I will get a NVM Express drive. How can I make sure it will have enough endurance?
- Bonus: How to over-provision the drive?
- Bonus #2: I only use Windows®! Which tool to use in this OS?

What is the name of new exciting memory technology which Intel® and Micron® have recently introduced?

- 1) 4D MegaNAND™
- 2) 3D Crossbow™
- 3) 3D XPoint™

What is NVM Express™ (NVMe)?

NVM Express™ is Transforming the Data Center



Efficiency

Developed to be lean, NVM Express™ (NVMe) delivers performance with less resources, increasing storage density and lowering TCO.



Performance

Scalable performance and low latency of PCI Express® (PCIe) with NVMe optimizing the storage stack.



Industry
Standard

NVMe is bringing PCIe SSDs into the mainstream with industry standard software and drivers and management for the future software defined data center.



Infrastructure for PCI Express® (PCIe):
Cables, connectors,
backplanes, switches,
retimers.



NVM Express™
(NVMe) Software,
drivers, and
management.



RAID for NVMe
with Intel®
Rapid Storage
Technology
enterprise.



IHV and ISV
partners.

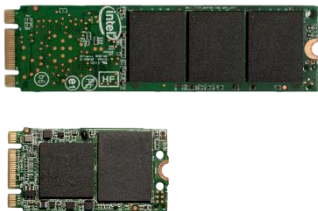


PCIe and NVMe
compatibility
programs.

Data Center Form Factors for



M.2



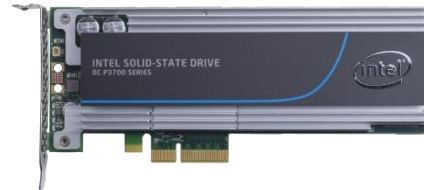
42, 80, and 110mm lengths,
Smallest footprint of PCI
Express®, use for boot or for
max storage density.

U.2 2.5in (SFF-8639)



2.5in makes up the
majority of SSDs sold
today because of ease of
deployment, hotplug,
serviceability, and small
form factor.

Add-in-card



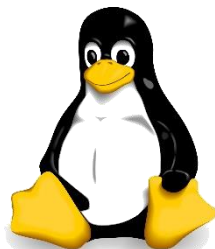
Add-in-card (AIC) has maximum
system compatibility with
existing servers and most
reliable compliance program.
High power envelope, and
options for height and length.

How do I Know My NVM Express™ Drive is Working?

NVM Express™ Driver Ecosystem



Windows 8.1



Linux* NVMe Express™ (NVMe) driver is
[open source](#)



Native / in-box

Install NVMe driver



vmware®
ESXi 5.5

I Do Not See the Drive in the System!

Step 1: First Simple Failure Analysis Steps

1. Check if Intel PCI Express® SSD is listed among devices:

a) `'lspci -nn | grep 0953'` or `'lspci | grep memory'`

2. Check if driver is loaded and which version it is:

b) `'lsmod | grep nvme'`

c) `'modinfo nvme'`

3. Check if controller and namespaces are visible in devices list:

c) `'ls /dev/nvm*'`

`nvmeX`

`nvmeXnY`

Log in: root

Password: password

```
root@localhost:~/Desktop
File Edit View Search Terminal Help
[root@localhost Desktop]# lspci -nn | grep 0953
02:00.0 Non-Volatile memory controller [0108]: Intel Corporation PCIe Data Center SSD [8086:0953] (rev 01)
07:00.0 Non-Volatile memory controller [0108]: Intel Corporation PCIe Data Center SSD [8086:0953] (rev 01)
[root@localhost Desktop]#
[root@localhost Desktop]# lsmod | grep nvme
nvme                    57344  0
[root@localhost Desktop]# ls /dev/nvm*
/dev/nvme0 /dev/nvme0n1 /dev/nvme1 /dev/nvme1n1
[root@localhost Desktop]#
```

Step 2: Check LEDs

LED Placement & Functionality



- Four LED indicators are available on the AIC, visible through the IO bracket

- The LEDs will indicate:

- Drive activity (yellow, *blink rate proportional to workload*)
- Drive failed (red)
- Drive operating in a defective state (yellow)
 - Tied to critical warnings in NVMe Express™ Health Log:
 1. Available spare memory has fallen below specified threshold
 2. Temperature has exceeded critical temperature
 3. Reliability of the device has degraded
 4. Volatile memory backup circuit test has failed
 5. Device has become read-only
- Drive healthy (green)

From top to bottom:

LED1. Activity (yellow)

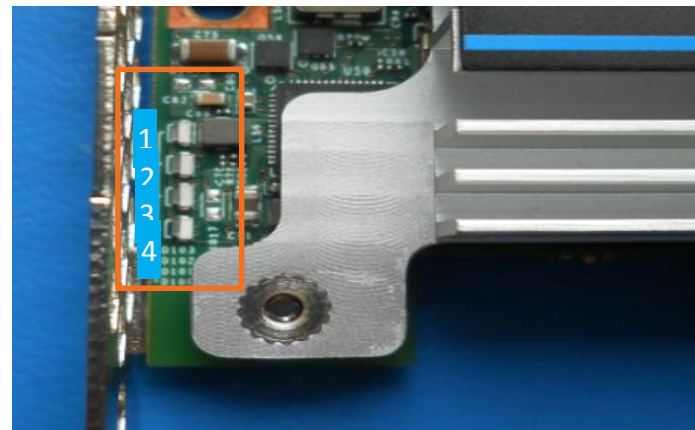
LED 2. Failed (red)

LED 3. Defect (yellow)

LED 4. Healthy (green)

2.5" SFF (U.2) - Pin 11 Activity Functionality

1. When idle or no IO activity, logic level on this pin is low (LED ON)
2. When there is IO activity, pin puts out a 250ms High, 250ms low signal
3. When drive is formatting, pin toggles at 250ms High, 250ms low signal



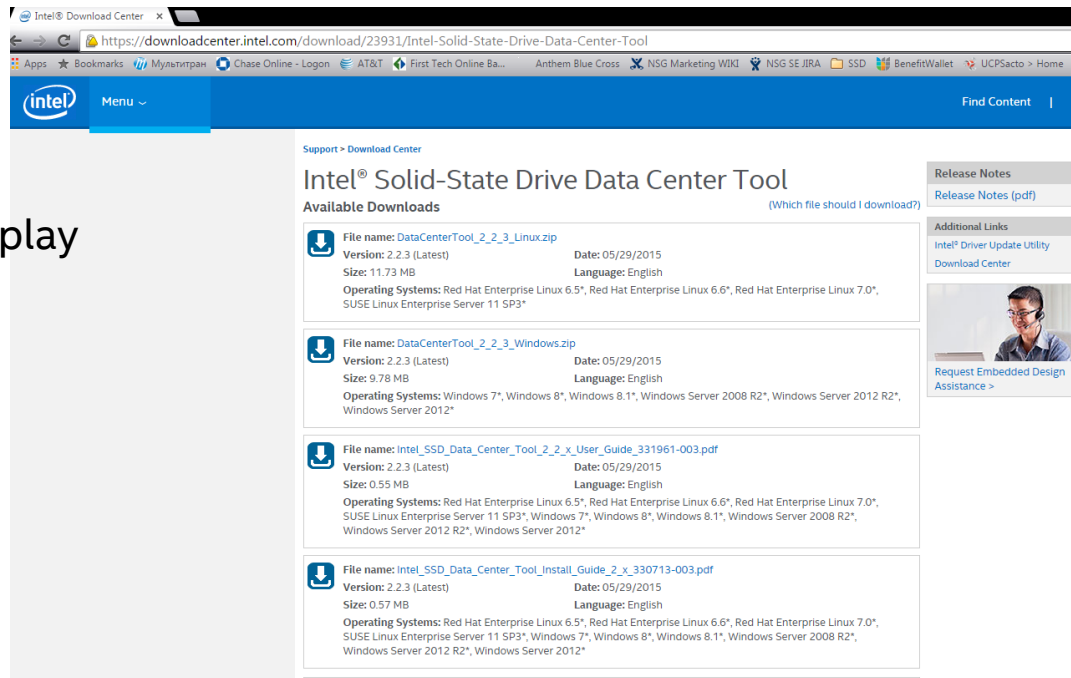
Step 3: Download and Install Intel® SSD Data Center Tool

Free tool to manage your Intel® SSD Data Center drives including PCI Express® family. Available for both Linux* and Windows® x86/x64 versions.

Intel SSD Data Center Tool (Intel® SSD DCT)

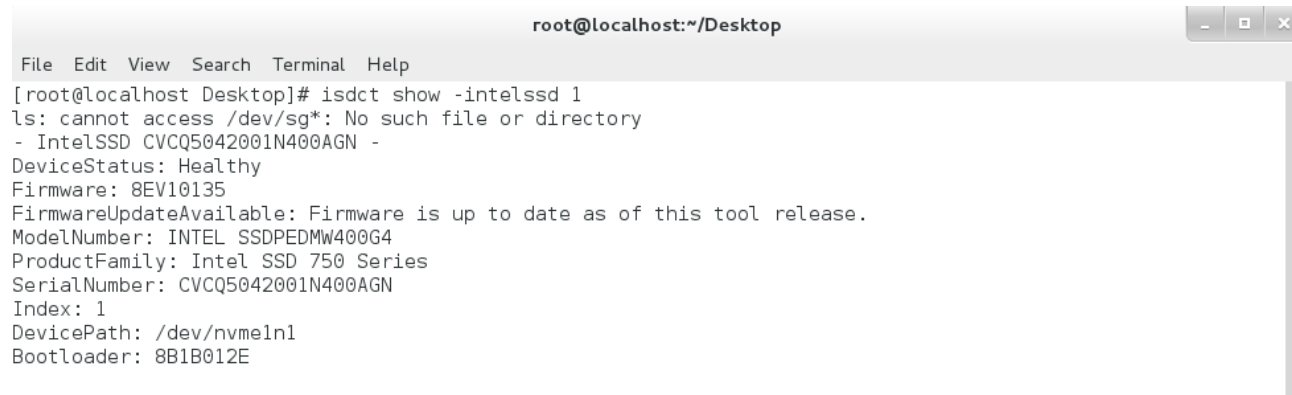
Capabilities include:

- Firmware update to latest version
- Secure Erase
- SMART/NVM Express™ health log display
- Settings change
- Over-Provisioning
- Endurance Analyzer
- Power governor settings
- Attributes display



Step 4: Check Device Status with Intel® SSD Data Center Tool

Show all supported Intel® SSDs in the system:

A terminal window titled 'root@localhost:~/Desktop' with a menu bar (File, Edit, View, Search, Terminal, Help). The command '[root@localhost Desktop]# isdct show -intelssd 1' has been executed. The output shows an error for the first device, followed by detailed information for the second device (Intel SSD 750 Series).

```
root@localhost:~/Desktop
File Edit View Search Terminal Help
[root@localhost Desktop]# isdct show -intelssd 1
ls: cannot access /dev/sg*: No such file or directory
- IntelSSD CVCQ5042001N400AGN -
DeviceStatus: Healthy
Firmware: 8EV10135
FirmwareUpdateAvailable: Firmware is up to date as of this tool release.
ModelNumber: INTEL SSDPEDMW400G4
ProductFamily: Intel SSD 750 Series
SerialNumber: CVCQ5042001N400AGN
Index: 1
DevicePath: /dev/nvme1n1
Bootloader: 8B1B012E
```

Check attributes available for your NVM Express™ (NVMe) based SSD

```
# isdct show -all -intelssd
```

Check NVMe Health Log (SMART)

```
# isdct show -sensor -intelssd 1
```

Start endurance analyzer (reset Timed Workload Indicator counters)

```
# isdct set -intelssd 1 EnduranceAnalyzer=reset
```

Step 4 (cont.): Checking Important Attributes

```
root@localhost:~/Desktop

File Edit View Search Terminal Help

AggregationThreshold: 0
AggregationTime: 0
ArbitrationBurst: 0
AsynchronousEventConfiguration: 0
Bootloader: 8B1B012E
DevicePath: /dev/nvme0n1
DeviceStatus: Healthy
EnduranceAnalyzer: Media Workload Indicators have reset values. Run 60+ minute workload endurance analyzer.
ErrorString:
Firmware: 8EV10135
FirmwareUpdateAvailable: Firmware is up to date as of this tool release.
HighPriorityWeightArbitration: 0
Index: 0
IOCompletionQueuesRequested: 30
IOSubmissionQueuesRequested: 30
LBAFormat: 0
LowPriorityWeightArbitration: 0
ProductFamily: Intel SSD 750 Series
MaximumLBA: 781422767
MediumPriorityWeightArbitration: 0
MetadataSetting: 0
ModelNumber: INTEL SSDPEDMW400G4
NativeMaxLBA: 781422767
NumErrorLogPageEntries: 63
NumLBAFormats: 6
NVMePowerState: 0
PCILinkGenSpeed: 3
PCILinkWidth: 4
PhysicalSize: 400088457216
PowerGovernorMode: 0 (25W)
ProtectionInformation: 0
ProtectionInformationLocation: 0
RAIDMember: False
SectorSize: 512
SerialNumber: CVCQ5042001S400AGN
SystemTrimEnabled:
TempThreshold: 85 degree C
TimeLimitedErrorRecovery: 0
TrimSupported: True
WriteAtomicityDisableNormal: 0
```

Bootloader version

Firmware version

Is drive over-provisioned?

Is PCI Express® link trained on Gen3 speed and x4 width?

Is drive's power limited by firmware?

How do I Measure Performance of My NVM Express™ Drive?

Solid-State Storage Drive States

Drive states

- **Fresh out of box**
 - All space on the drive is free and has not been written to
 - A Secure Erase will return a drive to FOB
 - SSDs will perform best in FOB since new data being written goes into next available free space
- **Transition**
 - Amount of time in transition state is dependent on available free space and write history, previous workloads (pre-conditioning), and current workload pattern
- **Steady State**
 - Desired state to properly measure and report solid state storage performance! This is typically worst case performance which is important for the data center.

Methodology

- **Drive Prep Process**

- A basic method would be to secure erase and sequentially write the entire drive. Steady State is achieved much faster as opposed to writing 4K random data to an empty drive, and the transition stage would look different.

- **Testing Order**

- The order you run tests can heavily influence the final results. For example reading sequentially over random written data can drop performance. The impact will be different depending on what SSD is being tested.

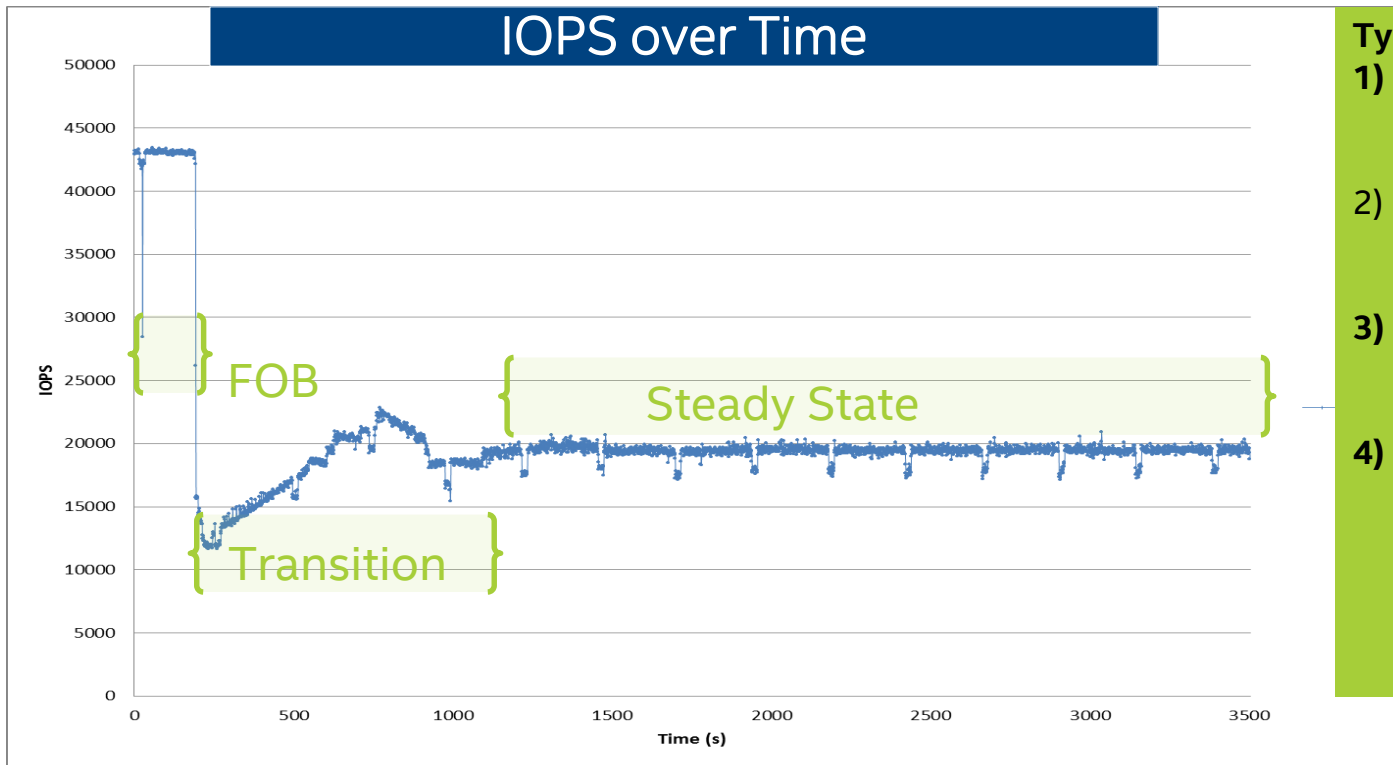
- **Test Time / Average Time**

- Depending on the architecture random writes (or writes in general) can take much longer to reach stability than other results

- **Test Span**

- A full drive test will generally perform worse than a test over a small region. Most SSDs use unwritten LBAs as spare area until they are written to. Intel® SSD Data Center Series are tested on 100% LBA span.

Solid-State Storage Performance



Typical test flow

- 1) **Purge Device:** Secure Erase to return drive to FOB.
- 2) Run workload independent **preconditioning:** Full span sequential write.
- 3) **Run test:** 4k random write, Queue Depth 32, span 100%.
- 4) **Record results:** Measure during steady state only! Be sure to include information about test system hardware, software, and device under test.

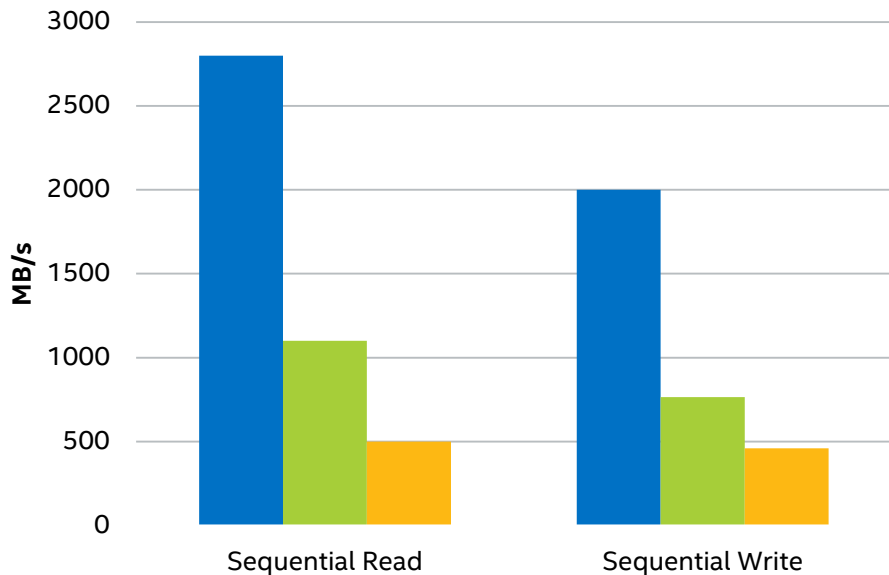
Storage Performance Metrics

- **Bandwidth** (Often MBps)
 - This is commonly measured in megabytes per second. Calculated by the IOPS multiplied by the block size. Bound by the storage interface (example: 6Gbps SATA* max theoretical throughput is 600MBps).
- **IOPS**, Input Output operations Per Second
 - The number of I/O packets that can be transferred per sec; packet sizes can range from 512B to 4MB. IOPS data over time you can really start to paint a picture of what is happening on the drive. Traditionally you would focus on a single average number, but enterprise workloads demand strict stability requirements as well. A new focus is comparing the mean to min on a per second basis (performance consistency).
- **Latency/QoS** (Quality of Service)
 - Latency is in many way related to IOPS. Each command sent takes a period of time to service and return (latency, sometimes called the service time). QoS focuses on the worst outliers in latency over a period of time. In general this is most important for enterprise applications but is applicable to client in extreme cases (a drive timeout is in many ways just a really bad QoS event).

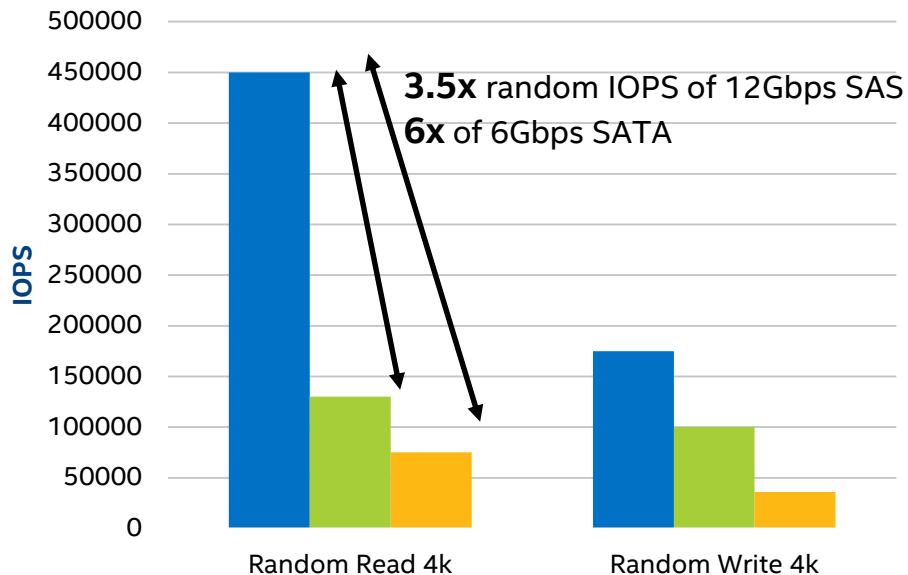
There is no such thing as too much detail in a system report

Up to 6x the Performance of Best in Class SATA* SSDs at Half the Latency and CPU Utilization

Sequential Bandwidth



Random Workload



P3700 Series

SAS 12Gbps

SATA* 6Gbps

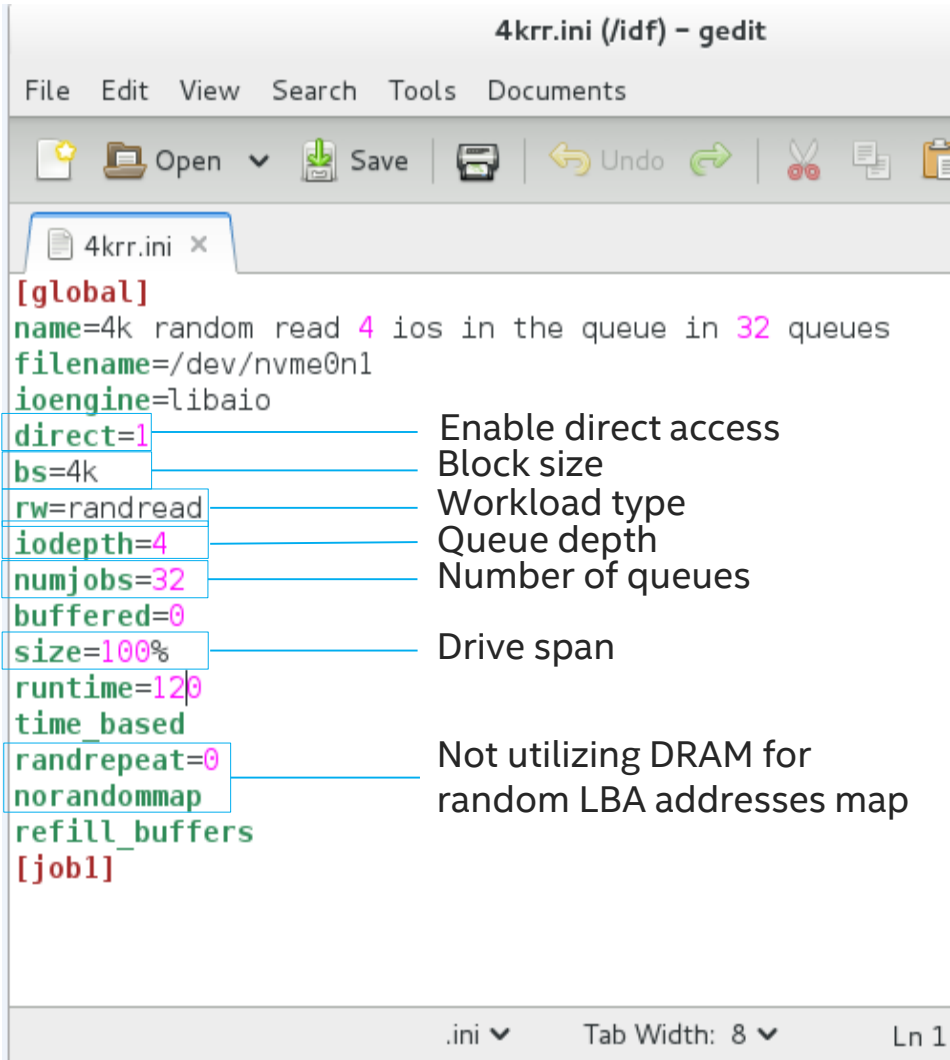
Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance.

Configurations: Performance claims obtained from data sheet, sequential read/write at 128k block size for NVMe Express™ and SATA, 64k for SAS. Intel SSD C P3700 Series 2TB, SAS Ultrastar® SSD1600MM, Intel SSD DC S3700 Series SATA 6Gbps. Intel® Core™ i7-3770K CPU @ 3.50GHz, 8GB of system memory, Windows® Server 2012, IOMeter*. Random performance is collected with 4 workers each with 32 QD

FIO

- FIO is an open-source, flexible IO measuring tool which supports multiple threads, various queue depths and many other parameters
- FIO is an industry standard tool for measuring storage performance in Linux*
- There are 3 ways to run FIO:
 - script file (.\4krr.sh)
 - job/ini file (fio 4krr.ini)
 - old school command line

(fio --name=test1 --filename=/x --size=10G --ioengine=libaio --direct=1 --numjobs=1.....)



The screenshot shows a gedit editor window titled "4krr.ini (/idf) - gedit". The menu bar includes File, Edit, View, Search, Tools, and Documents. The toolbar contains icons for Open, Save, Undo, and other standard editing functions. The editor displays the contents of the 4krr.ini file, which is a configuration for FIO. The configuration is as follows:

```
[global]
name=4k random read 4 ios in the queue in 32 queues
filename=/dev/nvme0n1
ioengine=libaio
direct=1
bs=4k
rw=randread
iodepth=4
numjobs=32
buffered=0
size=100%
runtime=120
time_based
randrepeat=0
norandommap
refill_buffers
[job1]
```

Annotations on the right side of the editor explain the parameters:

- direct=1**: Enable direct access
- bs=4k**: Block size
- rw=randread**: Workload type
- iodepth=4**: Queue depth
- numjobs=32**: Number of queues
- size=100%**: Drive span
- randrepeat=0**: Not utilizing DRAM for random LBA addresses map

The status bar at the bottom indicates ".ini", "Tab Width: 8", and "Ln 1".

Examples of FIO ini Files for Sequential and Random Workloads

Optimal `Queue_Depth * Number_of_queues` for Intel® NVM Express™ drives = 128.

For best random performance create `Number_of_queues = Number_of_logical_cores` with total QD=128 (8 queues each QD=16, 32 queues each QD=4, etc)

For best sequential performance create 1 Queue (thread) with QD=128

```
[global]
name=Random_Read_test1
filename=/dev/nvme0n1
ioengine=libaio
direct=1
bs=4k
rw=randread
iodepth=32
numjobs=8
buffered=0
size=100%
runtime=60
time_based
randrepeat=0
Norandommap
refill_buffers
```

Random

```
[global]
name=Seq_Write_test2
filename=/dev/nvme0n1
ioengine=libaio
direct=1
bs=128k
rw=write
iodepth=128
numjobs=1
buffered=0
size=100%
runtime=60
time_based
randrepeat=0
Norandommap
refill_buffers
```

Sequential

Linux* Performance Testing of a Single NVMe Express™ Drive

Step 5: Measure performance using FIO, iostat and performance counters

Exercise: Open 2 terminal windows

- # **fio /idf/4krr.ini** in 1st terminal window
- # **iostat -c -d -x -t -m /dev/nvme0n1 1** in 2nd terminal window (this will show CPU usage)

```
root@localhost:~# fio
File Edit View Search Terminal Help
job1: (groupid=0, jobs=1): err=0: pid=28856: Thu Aug 6 11:57:11 2015
read : io=15613MB, bw=266455KB/s, iops=66613, runt= 60000msec
slat (usec): min=0, max=21, avg= 1.68, stdev= 0.82
clat (usec): min=8, max=3491, avg=478.20, stdev=208.45
lat (usec): min=10, max=3493, avg=479.94, stdev=208.45
clat percentiles (usec):
| 1.00th=[ 84], 5.00th=[ 167], 10.00th=[ 223], 20.00th=[ 298],
| 30.00th=[ 358], 40.00th=[ 410], 50.00th=[ 462], 60.00th=[ 516],
| 70.00th=[ 572], 80.00th=[ 644], 90.00th=[ 748], 95.00th=[ 852],
| 99.00th=[ 1032], 99.50th=[ 1112], 99.90th=[ 1272], 99.95th=[ 1368],
| 99.99th=[ 2256]
bw (KB /s): min=186416, max=320984, per=12.35%, avg=266412.71, stdev=30728.19
lat (usec): 10=0.01%, 20=0.01%, 50=0.26%, 100=1.19%, 250=11.67%
lat (usec): 500=43.97%, 750=32.82%, 1000=8.69%
lat (msec): 2=1.37%, 4=0.01%
cpu      : usr=9.51%, sys=16.44%, ctx=2801719, majf=0, minf=40
IO depths : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=100.0%, >=64=0.0%
submit    : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
complete  : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.1%, 64=0.0%, >=64=0.0%
issued    : total=r=3996828/w=0/d=0, short=r=0/w=0/d=0, drop=r=0/w=0/d=0
latency   : target=0, window=0, percentile=100.00%, depth=32

Run status group 0 (all jobs):
  READ: io=126402MB, aggrb=2106.8MB/s, minb=248775KB/s, maxb=291740KB/s, mint=60000msec, maxt=60000msec

Disk stats (read/write):
nvme0n1: ios=32312831/0, merge=0/0, ticks=15209895/0, in_queue=15260167, util=100.00%
[root@localhost idf]#
```

IO statistics
Latency statistics

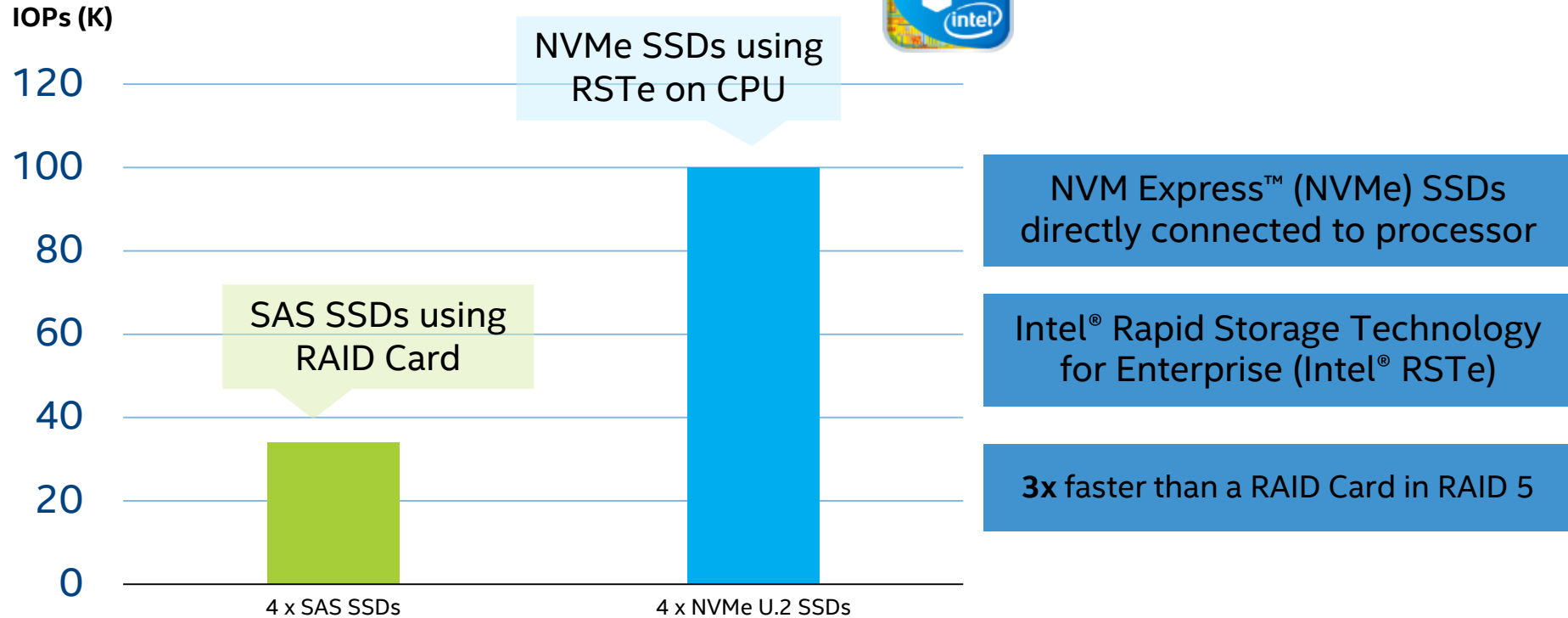
Latency percentiles

Queue Depth percentiles

Drive IO statistics

Wow! I have 2 NVMe Express™ Drives in My System! Now How do I RAID Those?

NVM Express™ RAID with Intel® Rapid Storage Technology for Enterprise



Test configuration- Intel GZ2600 Server, dual socket Intel® Xeon® E5-2699 v3, 32GB DDR4 RAM, Hardware RAID controller. Operating System: Microsoft Windows Server 2012 R2. Configuration 1- 12G SAS . 4ea (Seagate® 1200 SSD ST800FM0043 SSDs-800GB in RAID5 (LSI MegaRAID SAS-3 3108 Controller). Configuration 2 – NVMe-RSTe 4ea (Intel® DC P3700 800GB SSD) in RAID5 array. Intel RSTe 4.3 BETA software. Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance.

Step 6: Create an Intel® RSTe RAID Volume

The lab system already has latest version of mdadm included in the Linux kernel and Intel® Matrix Storage Manager software option can be used*

1. Create a container:

```
# mdadm -C /dev/md/imsmd /dev/nvme0n1 /dev/nvme1n1 -n 2 -e imsm -f
```

2. Create a RAID 0 volume within the container:

```
# mdadm -C /dev/md0 /dev/md/imsmd -n 2 -l 0 -c 128 -f
```

3. Create a partition on RAID 0 volume:

```
# parted /dev/md0 mklabel gpt
```

```
# parted /dev/md0 mkpart primary 2097152B 100%
```

4. Create an xfs filesystem on the RAID volume:

```
# mkfs.xfs -K /dev/md0p1 -f
```

is the root shell prompt and not part of the command. Also all commands must be run as root (administrator).

Step 7: Create and Mount Partition

5. Verify partition creation

parted /dev/md0 print

6. Mount the file system on the RAID volume:

mkdir /mnt/myraid/

mount -o noatime,nodiratime,nobarrier /dev/md0p1 /mnt/myraid

noatime - not recording last file
access time

nodiratime - not recording last directory
access time

nobarrier - no flush cache commands

7. Show the container and RAID volume

cat /proc/mdstat

```
root@localhost:~/Desktop
File Edit View Search Terminal Help
[root@localhost Desktop]# parted /dev/md0 print
Model: Linux Software RAID Array (md)
Disk /dev/md0: 800GB
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number  Start   End     Size    File system  Name      Flags
  1      2097kB  800GB   800GB   xfs          primary

[root@localhost Desktop]# cat /proc/mdstat
Personalities : [raid0]
md0 : active raid0 nvme0n1[1] nvme1n1[0]
      781416448 blocks super external:/md127/0 128k chunks

md127 : inactive nvme1n1[1](S) nvme0n1[0](S)
      2210 blocks super external:imsm
```

Linux* Performance Testing of RSTe RAID'd drive

Step 8: Measure performance of NVMe RAID volume

- Measuring performance with 'fio' over file system

```
# fio /idf/md4krr.ini
```

Please note that FIO will create test files under the directory specified here:

“Directory=/mnt/myraid”

Number of files will match number of jobs for this ini file

- Alternate way for creating test files – using 'dd' on a test partition with file system

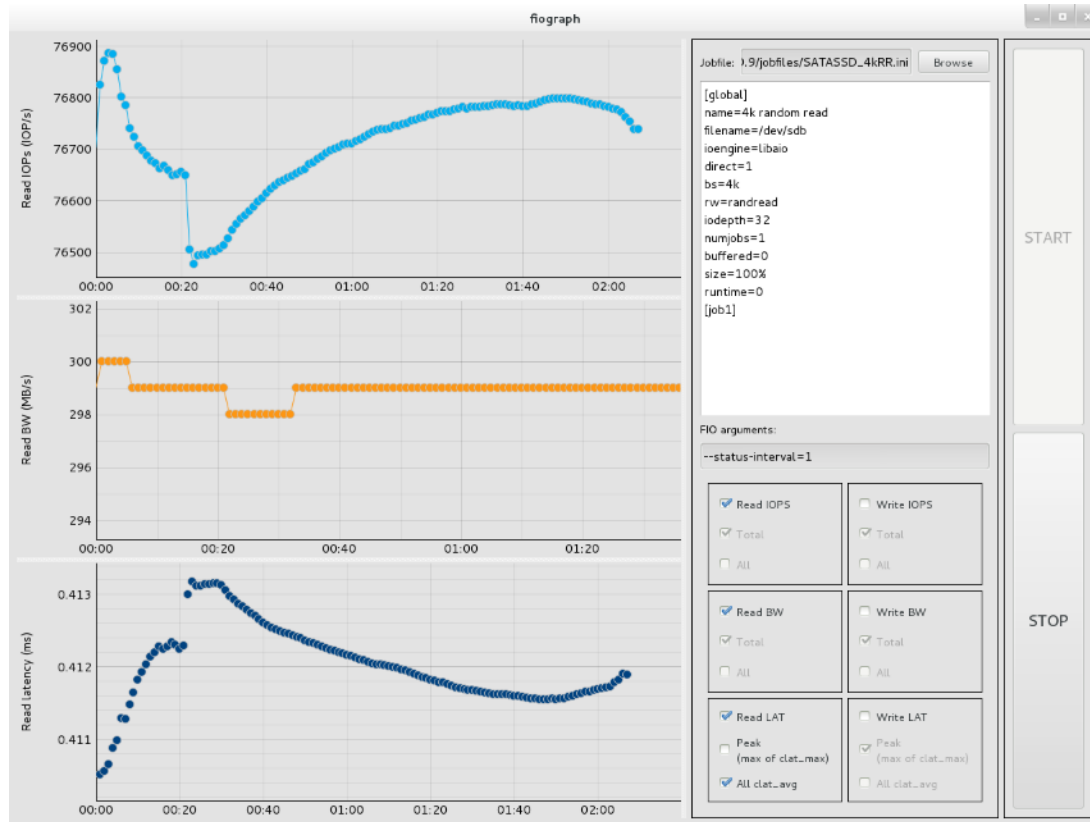
```
# dd if=/dev/zero of=/mnt/myraid/testfile bs=1M count=16384
```

How Can I Monitor NVM Express™ Performance and Debug Application Performance Issues ?

FIO Visualizer Tool

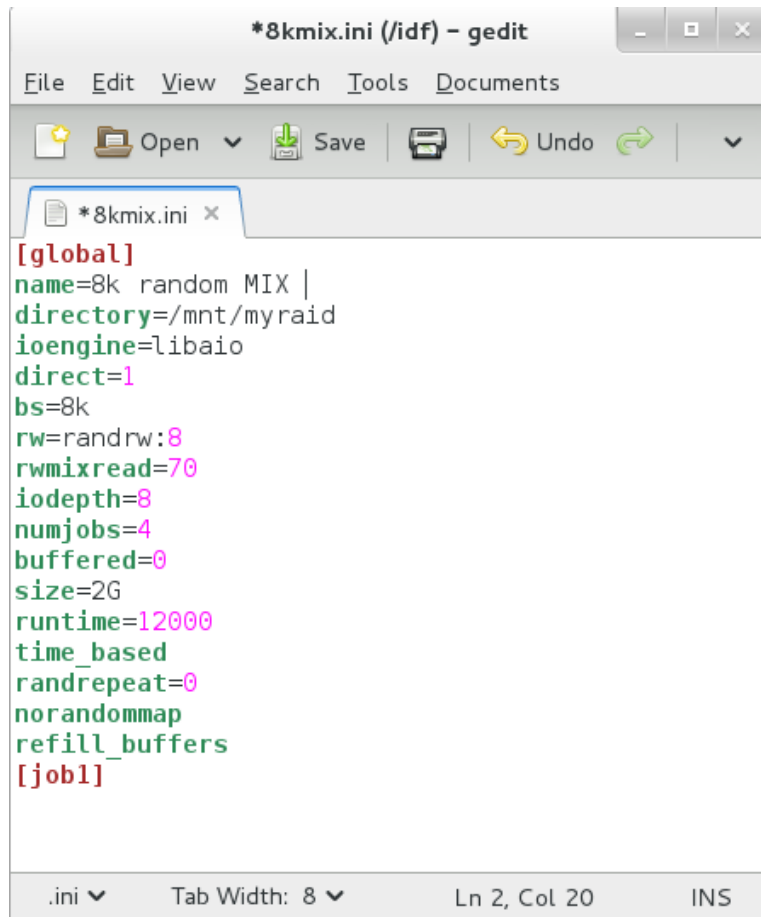
Unique features include:

- Single GUI for FIO tool, de facto standard benchmark tool for Linux*
- Real time. The data gathers by assigned time intervals and immediately displays on graphs.
- Monitors IOPS, bandwidth, latency for reads and writes
- Customizable layout. User defines which parameter needs to be monitored.
- Multithread / multi jobs support makes a value for NVMe Express™ SSD benchmarking
- Workload manager for FIO settings. Comes with base workload settings used in all Intel® SSD datasheets.



Step 9: Measure Performance Using FIO Visualizer and ioprof

- For this step we will be using pre-defined ini file /idf/8kmix.ini which represents semi-random customer workload with following parameters:
 - mix of reads and writes
 - 8KB access
 - addition of random block every 8 consecutive blocks
 - QD=8
 - 4 threads on 2 GB disk span each:



```
*8kmix.ini (/idf) - gedit
File Edit View Search Tools Documents
Open Save Print Undo Redo
*8kmix.ini x
[global]
name=8k random MIX |
directory=/mnt/myraid
ioengine=libaio
direct=1
bs=8k
rw=randrw:8
rwmixread=70
iodepth=8
numjobs=4
buffered=0
size=2G
runtime=12000
time_based
randrepeat=0
norandommap
refill_buffers
[job1]
.ini Tab Width: 8 Ln 2, Col 20 INS
```

Step 9 (cont'd): Measure Performance With FIO Visualizer and ioprof

Open FIO Visualizer and use /idf/8kmix.ini as a job file :

```
# cd /idf/fiovisualizer/
```

```
# ./fio_visualizer.py
```

Click 'Browse', go to '/idf/' folder, select 8kmix.ini

Click 'START'

Keep FIO visualizer running and open 2nd terminal window (adjust the size to half of the screen)

```
# cd /idf/ioprof-master/
```

```
# ./ioprof.pl -m trace -d /dev/md0 -r 30
```

```
# ./ioprof.pl -m post -t md0.tar
```

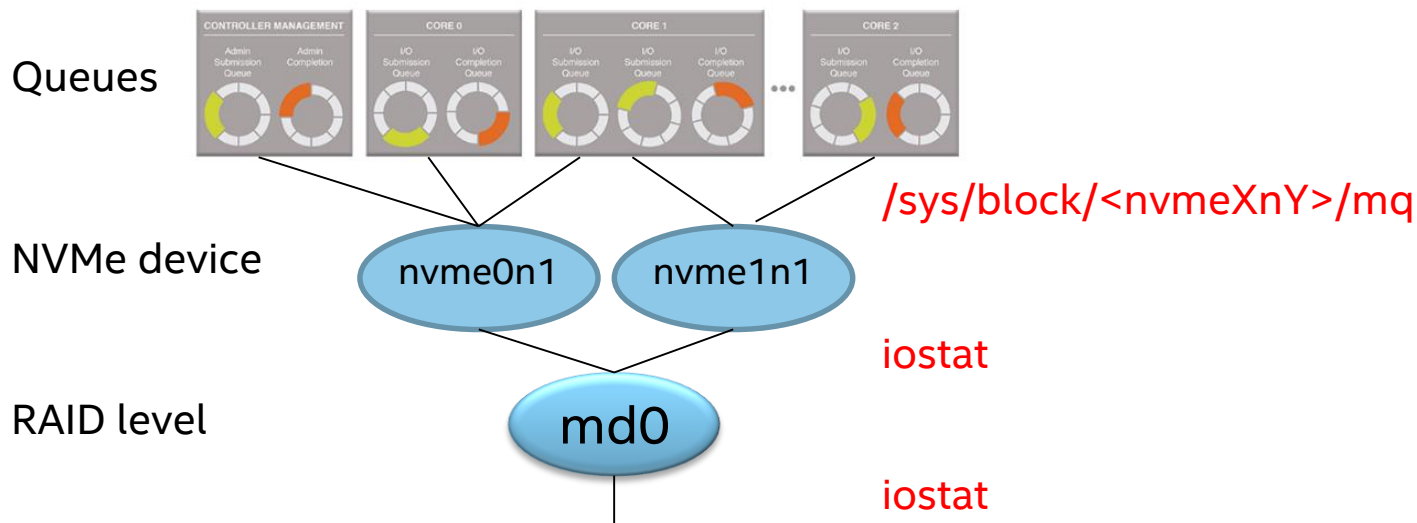
Please note that 'ioprof' should be installed separately as it is not part of standard Linux build*

'ioprof' is an open-source utility developed by Intel to analyze IO intensive workloads in background and report it's statistics

<http://www.01.org/ioprof>

Real-Time Monitoring of NVM Express™ Queues

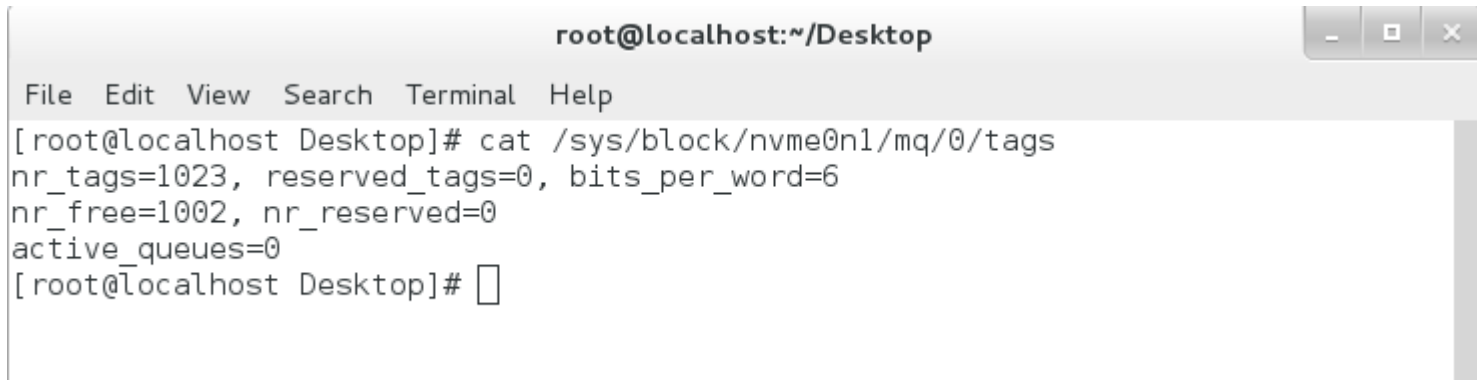
- It is possible to monitor IO utilization on the RAID level and device level via iostat, which is a standard Linux* tool for any block device
- Starting with kernel 3.19 NVM Express™ (NVMe) driver reporting has been expanded. This includes performance counters under `/sys/block/<nvmeXnY>/mq`
- Performance counters provide detailed information per NVMe queue



Step 10: Real-time Monitoring of NVM Express™ Queues

- Open new terminal window and run the command:

cat /sys/block/nvme0n1/mq/X/tags (where X = queue number)



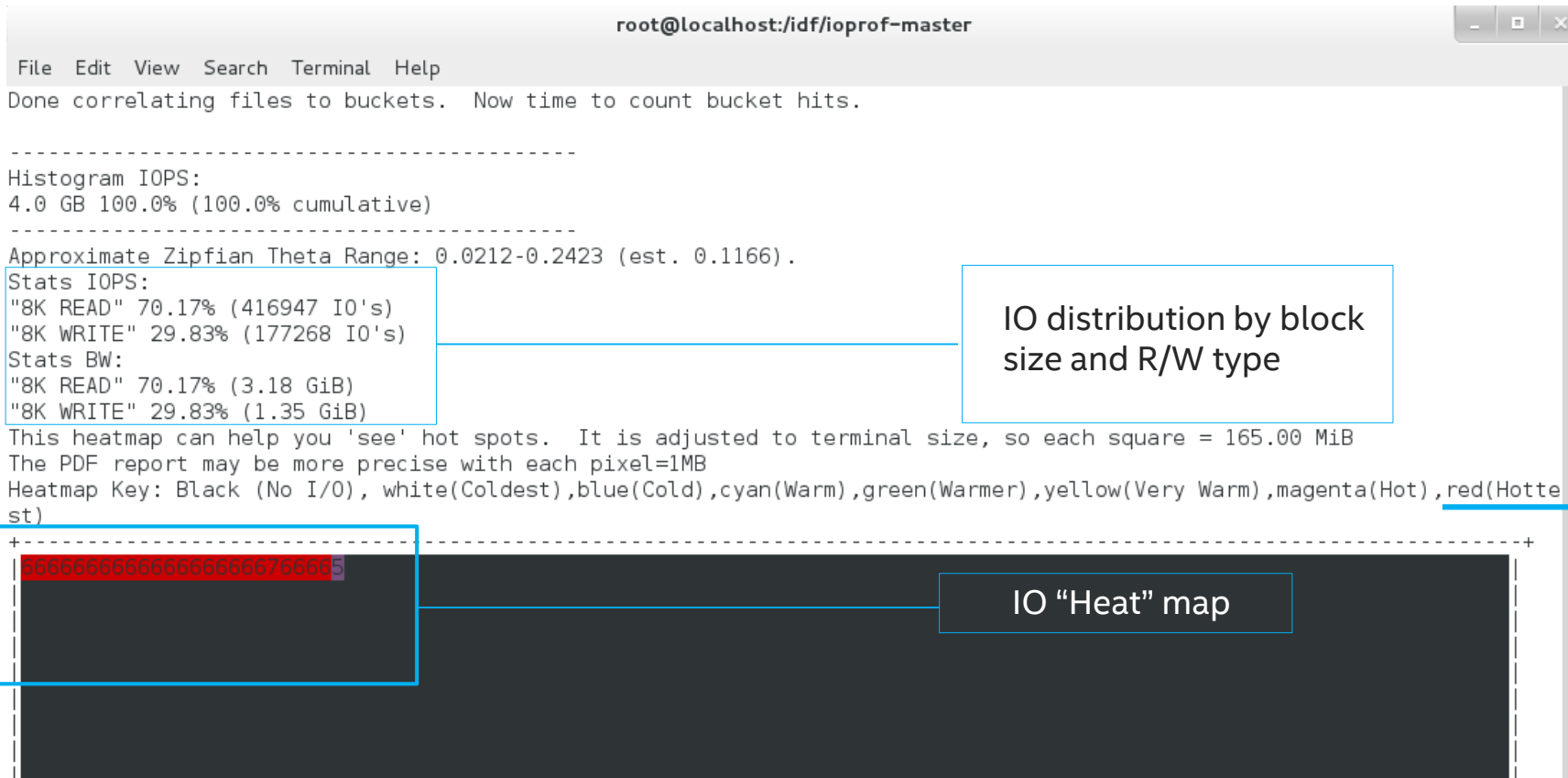
```
root@localhost:~/Desktop
File Edit View Search Terminal Help
[root@localhost Desktop]# cat /sys/block/nvme0n1/mq/0/tags
nr_tags=1023, reserved_tags=0, bits_per_word=6
nr_free=1002, nr_reserved=0
active_queues=0
[root@localhost Desktop]#
```

In this case:

nr_tags - nr_free = 21 → Represents number of submitted IO's waiting in the submission queue 0

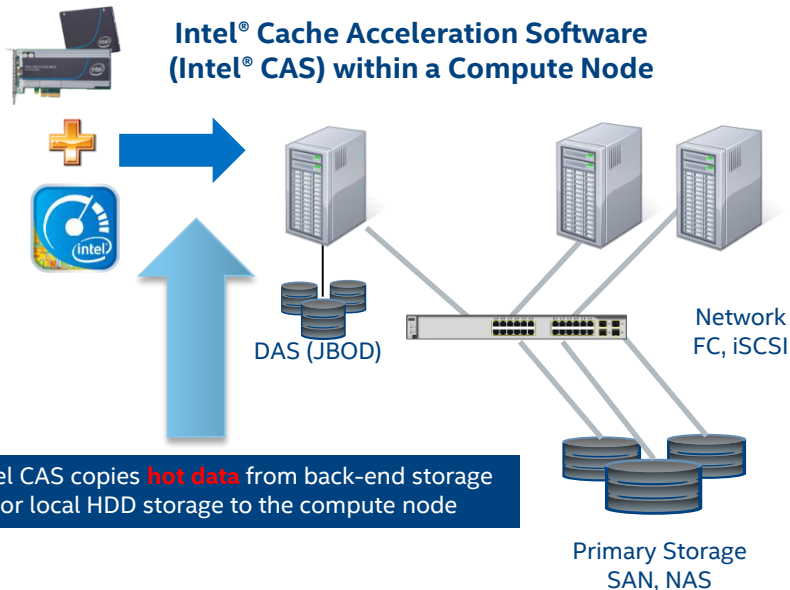
Please note that the number of queue pairs in NVM Express™ driver is equal to number of logical CPU cores in the system but cannot exceed 31.

Step 11: Analyzing ioprof Results



Intel® Cache Acceleration Software: Accelerate Your Data Center

Usage Model



Intel CAS Features & Benefits

Features	Benefits
Ease of Use	<ul style="list-style-type: none">Installed & Configured in <10minApplication acceleration with no modification to the application or storage back-end
Performance	<p>Intel CAS & Intel® SSD Delivers Gains vs HDD Solution up to:</p> <ul style="list-style-type: none">1400X IOPS¹, 57X OLAP², 3X OLTP²5X the Database Transaction throughput30X the Virtualization I/O throughput12X the Microsoft Exchange® Throughput
Flexible & Tuning	<ul style="list-style-type: none">Several Caching policies and options<ul style="list-style-type: none">Warm cache (pinning) for proactive cachingFile/block based caching w/ selective policiesMapping to SSD or System MemoryBroad OS Support (Linux® & Windows®)Virtualization Support (VM, XEN®, KVM, Microsoft®)Operates on any SSD, optimized for Intel SSD DC
Robust Solution	<ul style="list-style-type: none">Intel validated and optimized for Intel SSDs

Intel® CAS Accelerates Existing Applications by Addressing Storage I/O Bottlenecks Without Application or Storage Modifications

1. Configuration used: Intel® Server model 2600GZ (codename Grizzly Pass); Dual Intel® E5-2680 processor (2.7GHz), 32GB memory; Seagate® ST1000NC000 SATA HDD Microsoft Windows 2012R2 SP1, Intel SSD DC P3700 -800GB, Intel® CAS 2.6 release, L2 cache on ; IOMeter 10.22.2009 ; 4K Random Read test; 8-queue depth x 8 workers 2. Configuration used: Intel® Server model 2600GZ (codename Grizzly Pass); Dual Intel Xeon E5-2680 processor (2.7GHz), 96GB DDR3, VMware® 5.5, Intel® SSD DC P3700 -800GB, Intel® CAS 2.6., L2 cache off, 8xSeagate 146GB SAS In RAID5, VMs: Microsoft Server 2008R2, 8GB, 2 Cores, IOMeter workloads: Media Player ,Exchange Server, Web Server, 4K OLTP using QD4.1 Worker

Intel® Cache Acceleration Software Publicly Available Information

Download & Start an Evaluation Today!

Publicly Available Resources Intel.com: www.intel.com/cas

Software Downloads:

- Intel® I/O Assessment Tool – How Intel® Cache Acceleration Software (Intel® CAS) can benefit you
- 30-Day Trial Software - Try Before You Buy¹
- Click through to purchase
- Overview Videos & Documents:
- Intel Caching Explained Animation
- Introduction to Intel CAS
- Intel CAS Product Brief
- Intel CAS Solution Brief

Case Studies and Briefs:

- Accelerating Business Analytics with Intel CAS and Intel® Solid-State Drives
- Accelerating Analysis and Simulation with Intel CAS and Intel Solid-State Drives
- Accelerating MySQL* Database Performance with Intel CAS and Intel Solid-State Drives
- 3rd Party Performance Analysis Case Studies
- **& more coming!**

Contact Intel Authorized Distributors and Resellers for current promotional programs

¹Please see additional information at: intel.com/cas

Step 12: Deleting RAID Configuration

- Un-mounting the drive

```
# umount /mnt/myraid
```

- Stopping RAID

```
# mdadm --stop /dev/md0
```

```
# mdadm --stop /dev/md/ims
```

- Deleting meta-data

```
# mdadm --zero-superblock /dev/nvme0n1
```

```
# mdadm --zero-superblock /dev/nvme1n1
```

**Alright, I Will get an NVM Express™ Drive.
How can I Make Sure it Will Have Enough
Endurance?**

Step 13: Calculating Endurance of an Intel® SSD

In the beginning of the lab we reset Timed Workload SMART attributes by using following command:

```
# isdct set -intelssd 1 EnduranceAnalyzer=reset
```

To check Endurance Analyzer result, issue following command:

```
# isdct show -a -intelssd 0
```

Look for “EnduranceAnalyzer”

Intel Data-Center SSD Endurance is calculated for workload specified in JESD-219 standard: 100% write, 100% drive span, 100% random, mixed blocks size with 67% being 4KB

```
[root@localhost Desktop]# isdct show -a -intelssd 0
ls: cannot access /dev/sg*: No such file or directory
- IntelSSD CVCQ5042001N400AGN -
AggregationThreshold: 0
Aggregation Time: 0
ArbitrationBurst: 0
AsynchronousEventConfiguration: 0
Bootloader: 8B1B012E
DevicePath: /dev/nvme0n1
DeviceStatus: Healthy
EnduranceAnalyzer: 6.06 Years
ErrorString:
Firmware: 8EV10135
FirmwareUpdateAvailable: Firmware is up to date as of this tool release
```

Please note that in order for calculation to be valid, after reset workload should be applied minimum for 60 min and so media would wear out for at least 0.01%

Bonus: How to Over-provision the Drive?

Over-Provisioning

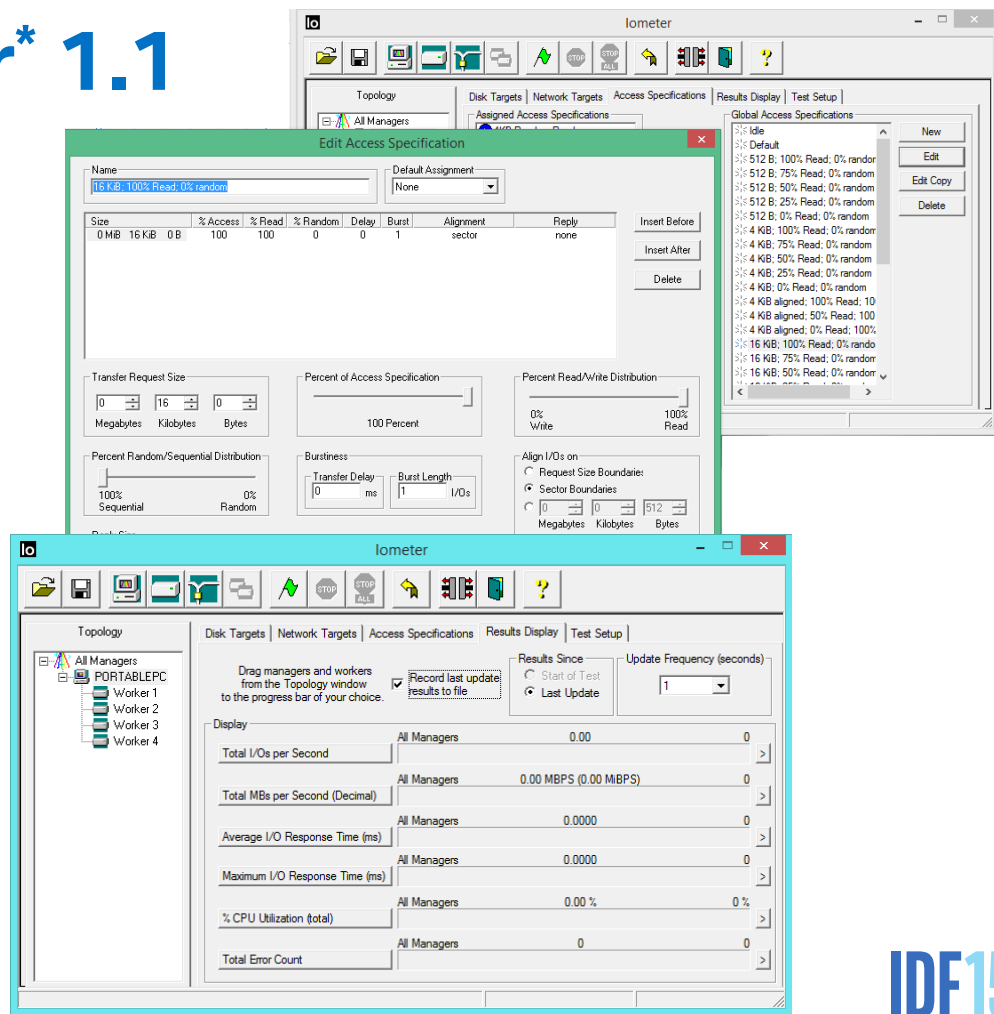
- Over-provisioning (OP) allows user to trade off capacity for lower Write Amplification thus increasing Performance and Endurance.
- OP is mostly applicable to Random Write with 100% drive span workloads. In such cases OP can lead to up to 400% endurance increase.
- Over-provisioning could be set at
 - SW level (limit capacity that SW can access, i.e. data base size)
 - OS level (partition the drive to use not full capacity of the drive)
 - Drive level (set Maximum LBA accessible and visible by user)
- isdct commands:
 - # isdct set -intelssd MaximumLBA=100000 (Max LBA = 100000)
 - # isdct set -intelssd MaximumLBA=100GB (Drive resized to 100GB)
 - # isdct set -intelssd MaximumLBA=80% (Drive resized to 80% of it's capacity)
 - # isdct set -intelssd MaximumLBA=native (return to full capacity)

Bonus #2: I only use Windows*! Which tool to use in this OS?

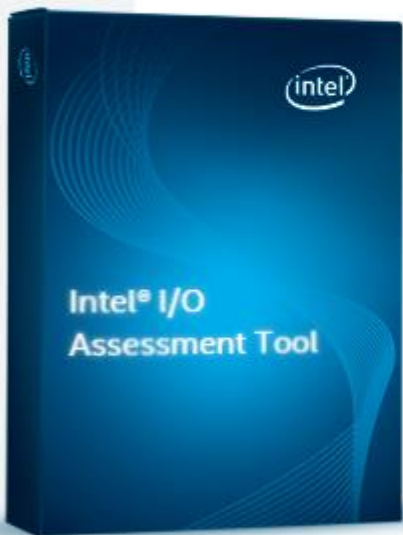
Testing with IOMeter* 1.1

NEW Features (in ver. 1.1)

- Record last update results to file
 - Latency buckets collected
 - Frequency dropdown menu
- Transfer size aligned patterns added
- Added option to align transfers to the request size boundary
- Allows changing data pattern
 - Compressible (Repeating Bytes)
 - Partially-compressible (Pseudo-Random)
 - Incompressible (Full Random)
- Transfer size aligned patterns added
- Added option to align transfers to the request size boundary



Intel® I/O Assessment Tool 1.0

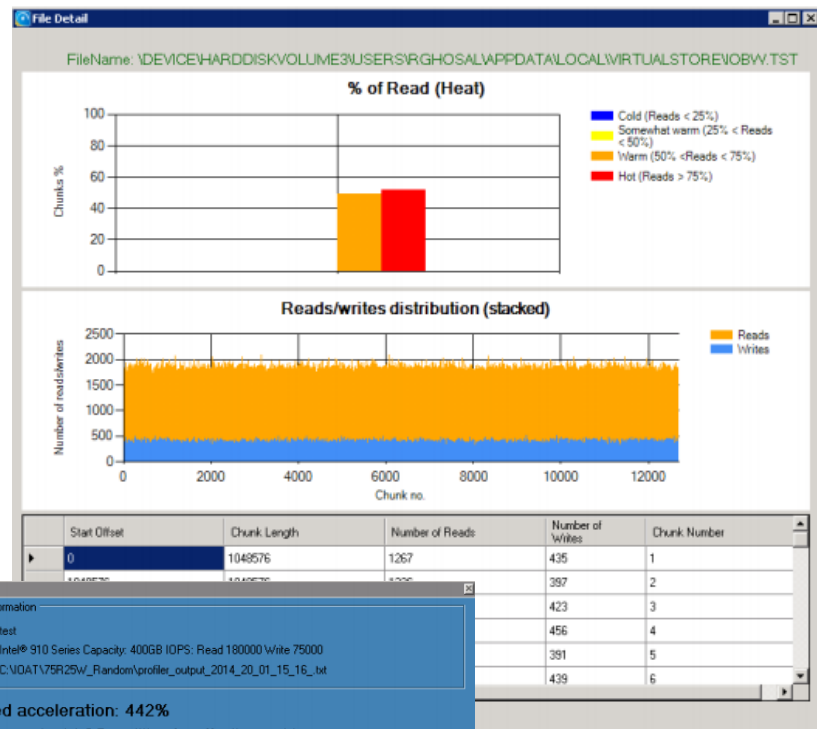


You have registered Intel® I/O Assessment Tool .

File to download:

ioat-01.00.00.1083.msi (3 MB)

Download Now



Summary

Workload Information:

Workload: test

SSD: Intel® 910 Series Capacity: 400GB IOPS: Read 180000 Write 75000

First Logfile: C:\IOAT\75R25W_Random\profiler_output_2014_20_01_15_16_bt

Estimated acceleration: 442%

We recommend a 14 GB partition for effective caching.

Percentage of Read = 75%

Percentage of Pre-read = 67.6%

Percentage of Write = 25%

For detail, please open the output.csv file.

Additional Sources of Information

- A PDF of this presentation is available from our Technical Session Catalog: www.intel.com/idfsessionsSF. This URL is also printed on the top of Session Agenda Pages in the Pocket Guide.
- Demos in the showcase – Intel® Solid State Drive Pavilion
- Additional info in the Intel SSD Pavilion
- More web based info: www.intel.com/ssd

Other Technical Sessions

Session ID	Title	Day	Time	Room
SSDL001/R	Hands-on Lab: How to Unleash Your Storage Performance by Using NVM Express* based PCI Express* Solid-State Drives	Wed	1:15; 4:00	2010
SSDC001	Tech Chat: Benchmarking Data Center Solid-State Drives – Insights Into Industry-Leading NVM Express* SSD Performance Metrics	Thurs	9:30 Thurs	Tech Chat Station 1
SSDC002	Tech Chat: Insights into Intel® Solid-State Drives Data Retention and Endurance	Thurs	9:30 Thurs	Tech Chat Station 2
SSDC003	Tech Chat: NVM Express* Features for High Availability and Storage Eco-System	Wed and Thurs	10:30 Wed 9:30 Thurs	Tech Chat Station 3
SSDS004	The Future of Storage Security	Thurs	1:00	2006
SSDS005	New Software Capabilities and Experiences Through Innovation in Storage Architecture	Thurs	2:15	2006

Legal Notices and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

No computer system can be absolutely secure.

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit <http://www.intel.com/performance>.

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

Statements in this document that refer to Intel's plans and expectations for the quarter, the year, and the future, are forward-looking statements that involve a number of risks and uncertainties. A detailed discussion of the factors that could affect Intel's results and plans is included in Intel's SEC filings, including the annual report on Form 10-K.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Intel, Core, Xeon, and the Intel logo are trademarks of Intel Corporation in the United States and other countries.

*Other names and brands may be claimed as the property of others.

© 2015 Intel Corporation.

Risk Factors

The above statements and any others in this document that refer to plans and expectations for the second quarter, the year and the future are forward-looking statements that involve a number of risks and uncertainties. Words such as "anticipates," "expects," "intends," "plans," "believes," "seeks," "estimates," "may," "will," "should" and their variations identify forward-looking statements. Statements that refer to or are based on projections, uncertain events or assumptions also identify forward-looking statements. Many factors could affect Intel's actual results, and variances from Intel's current expectations regarding such factors could cause actual results to differ materially from those expressed in these forward-looking statements. Intel presently considers the following to be important factors that could cause actual results to differ materially from the company's expectations. Demand for Intel's products is highly variable and could differ from expectations due to factors including changes in business and economic conditions; consumer confidence or income levels; the introduction, availability and market acceptance of Intel's products, products used together with Intel products and competitors' products; competitive and pricing pressures, including actions taken by competitors; supply constraints and other disruptions affecting customers; changes in customer order patterns including order cancellations; and changes in the level of inventory at customers. Intel's gross margin percentage could vary significantly from expectations based on capacity utilization; variations in inventory valuation, including variations related to the timing of qualifying products for sale; changes in revenue levels; segment product mix; the timing and execution of the manufacturing ramp and associated costs; excess or obsolete inventory; changes in unit costs; defects or disruptions in the supply of materials or resources; and product manufacturing quality/yields. Variations in gross margin may also be caused by the timing of Intel product introductions and related expenses, including marketing expenses, and Intel's ability to respond quickly to technological developments and to introduce new products or incorporate new features into existing products, which may result in restructuring and asset impairment charges. Intel's results could be affected by adverse economic, social, political and physical/infrastructure conditions in countries where Intel, its customers or its suppliers operate, including military conflict and other security risks, natural disasters, infrastructure disruptions, health concerns and fluctuations in currency exchange rates. Results may also be affected by the formal or informal imposition by countries of new or revised export and/or import and doing-business regulations, which could be changed without prior notice. Intel operates in highly competitive industries and its operations have high costs that are either fixed or difficult to reduce in the short term. The amount, timing and execution of Intel's stock repurchase program could be affected by changes in Intel's priorities for the use of cash, such as operational spending, capital spending, acquisitions, and as a result of changes to Intel's cash flows or changes in tax laws. Product defects or errata (deviations from published specifications) may adversely impact our expenses, revenues and reputation. Intel's results could be affected by litigation or regulatory matters involving intellectual property, stockholder, consumer, antitrust, disclosure and other issues. An unfavorable ruling could include monetary damages or an injunction prohibiting Intel from manufacturing or selling one or more products, precluding particular business practices, impacting Intel's ability to design its products, or requiring other remedies such as compulsory licensing of intellectual property. Intel's results may be affected by the timing of closing of acquisitions, divestitures and other significant transactions. A detailed discussion of these and other factors that could affect Intel's results is included in Intel's SEC filings, including the company's most recent reports on Form 10-Q, Form 10-K and earnings release.