STORAGE DEVELOPER CONFERENCE
SDC 21
BY Developers FOR Developers
Virtual Conference
September 28-29, 2021

Testing NVMe SSDs against the DatacenterSSD Specification

Tools and Methodologies

Presented by David Woolf –
University of New Hampshire InterOperability Lab (UNH-IOL)

Abstract

- The DatacenterSSD specification has been created by a group of hyperscale datacenter companies in collaboration with SSD suppliers and enterprise integrators. What is in this specification? How does it expand on the NVMe specification? How can devices demonstrate compliance? In this talk we'll review important items from the DatacenterSSD specification to understand how it expands on the NVMe family of specifications for specific use cases in a datacenter environment. Since the DatacenterSSD specification goes beyond just an interface specification, we will also show what test setups can be used to demonstrate compliance.

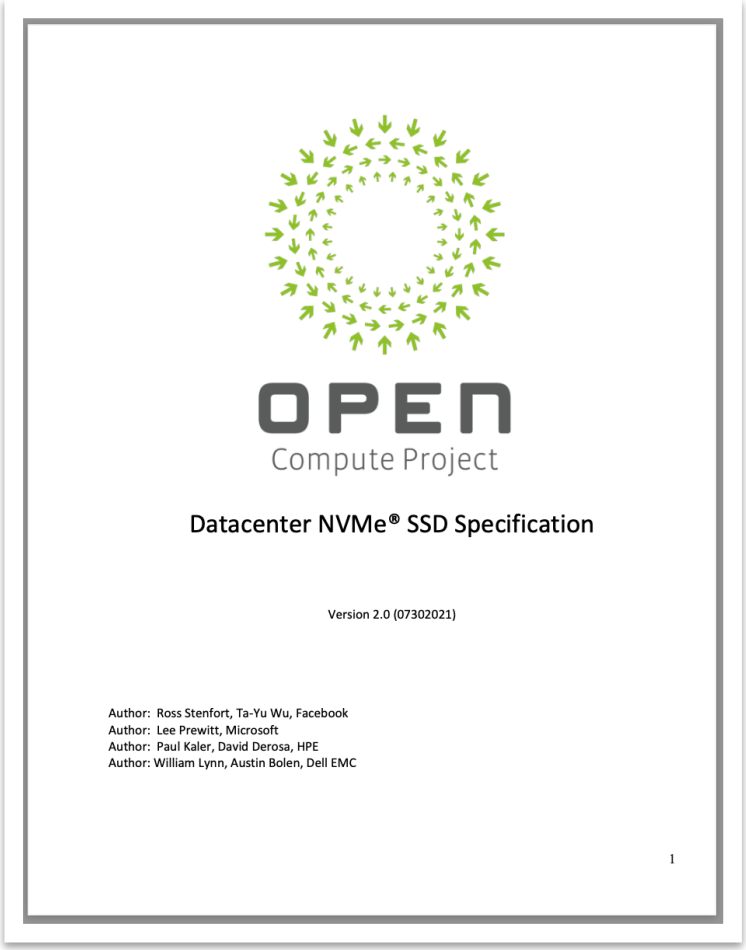
Learning Objectives

1. What is the DatacenterSSD specification?
2. How is the DatacenterSSD specification complementary to the NVMe specification?
3. How can an SSD demonstrate compliance to the DatacenterSSD specification?
4. What tools and equipment are necessary to demonstrate compliance to the DatacenterSSD specification?

What is the DatacenterSSD specification?

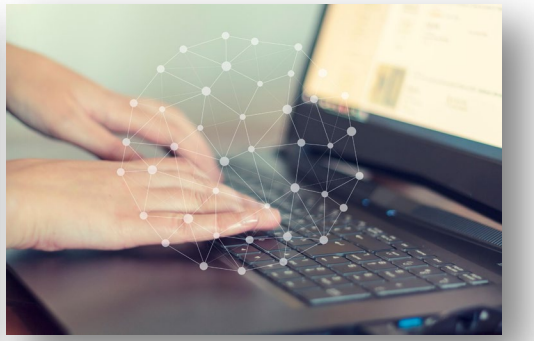
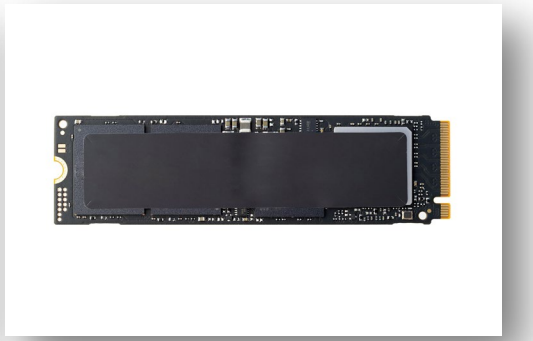
What is the DatacenterSSD specification?

- Originally launched as ‘CloudSSD’ in 2020
- Latest version, v2.0, published July 2021 as the ‘Datacenter NVMe® SSD Specification’ in the OCP Contributions Database:
- Link: https://www.opencompute.org/documents/datacenter-nvme-ssd-specification-v2-0r21-pdf
- Key contributors: Facebook, Microsoft, HPE, Dell



What is the DatacenterSSD specification?

- NVMe is intended as a storage protocol for a variety of transports and use cases.


- Datacenter NVMe SSD Spec is like a profile for implementing NVMe protocol in a way that's conducive to datacenter deployment.

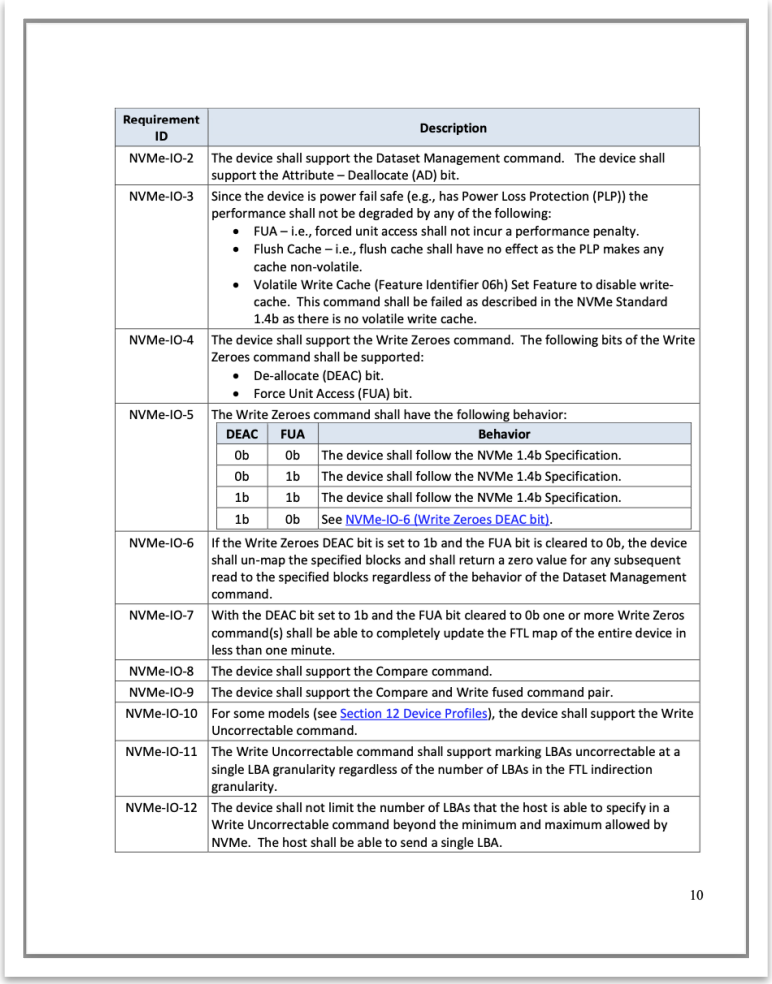
What is the DatacenterSSD specification?

- Datacenter NVMe SSD Spec goes far beyond NVMe protocol implementation:
  - NVMe Requirements
  - PCIe Requirements
  - Reliability
  - Endurance
  - Thermal
  - Out of Band Management
  - Security
  - Device Profiles
  - Labeling
  - RoHS / ESD Compliance
  - Shock and Vibration
  - NVMe CLI
  - Customer Specific Items

This presentation will primarily focus on capabilities that can be tested and observed via the NVMe interface

What is the DatacenterSSD specification?

- Formatted as a requirements list, each item has a unique Requirement ID
- Makes clear what NVMe optional features should be implemented or not.



Requirement ID	Description
NVMe-IO-2	The device shall support the Dataset Management command. The device shall support the Attribute – Deallocate (AD) bit.
NVMe-IO-3	Since the device is power fail safe (e.g., has Power Loss Protection (PLP)) the performance shall not be degraded by any of the following: • FUA – i.e., forced unit access shall not incur a performance penalty. • Flush Cache – i.e., flush cache shall have no effect as the PLP makes any cache non-volatile. • Volatile Write Cache (Feature Identifier 06h) Set Feature to disable write-cache. This command shall be failed as described in the NVMe Standard 1.4b as there is no volatile write cache.
NVMe-IO-4	The device shall support the Write Zeroes command. The following bits of the Write Zeroes command shall be supported: • De-allocate (DEAC) bit. • Force Unit Access (FUA) bit.
NVMe-IO-5	The Write Zeroes command shall have the following behavior:
NVMe-IO-6	If the Write Zeroes DEAC bit is set to 1b and the FUA bit is cleared to 0b, the device shall un-map the specified blocks and shall return a zero value for any subsequent read to the specified blocks regardless of the behavior of the Dataset Management command.
NVMe-IO-7	With the DEAC bit set to 1b and the FUA bit cleared to 0b one or more Write Zeros command(s) shall be able to completely update the FTL map of the entire device in less than one minute.
NVMe-IO-8	The device shall support the Compare command.
NVMe-IO-9	The device shall support the Compare and Write fused command pair.
NVMe-IO-10	For some models (see Section 12 Device Profiles), the device shall support the Write Uncorrectable command.
NVMe-IO-11	The Write Uncorrectable command shall support marking LBAs uncorrectable at a single LBA granularity regardless of the number of LBAs in the FTL indirection granularity.
NVMe-IO-12	The device shall not limit the number of LBAs that the host is able to specify in a Write Uncorrectable command beyond the minimum and maximum allowed by NVMe. The host shall be able to send a single LBA.

DEAC	FUA	Behavior
0b	0b	The device shall follow the NVMe 1.4b Specification.
0b	1b	The device shall follow the NVMe 1.4b Specification.
1b	1b	The device shall follow the NVMe 1.4b Specification.
1b	0b	See NVMe-IO-6 (Write Zeroes DEAC bit).

10

How is the DatacenterSSD specification complementary to the NVMe specification?

How is the DatacenterSSD specification complementary to the NVMe specification?

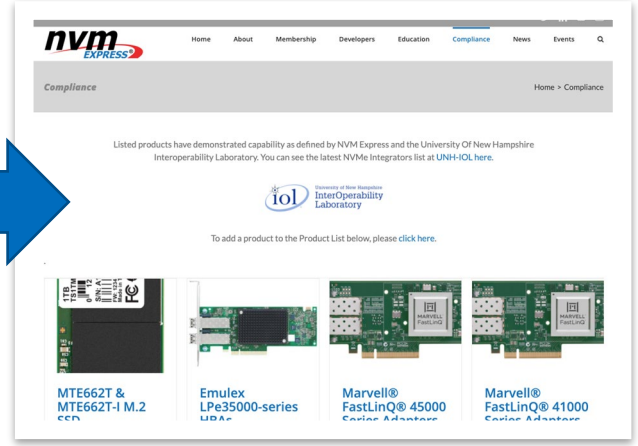
- Datacenter NVMe SSD Spec is like a profile for implementing NVMe protocol in a way that's conducive to datacenter deployment. How?
  1. Leverages NVMe v1.4b compliance
  2. Requiring or prohibiting features that are optional in the NVMe Specification
  3. Setting acceptable ranges/values for parameters defined in the NVMe Specification
  4. Defining new Features and Log Pages from 'Vendor Specific' ranges

How is the DatacenterSSD specification complementary to the NVMe specification?

- Example 1: Leverages NVMe v1.4b compliance

Requirement ID	Description
NVMe-1	The device shall comply with all required features of the NVMe 1.4b Specification. Optional features shall be implemented per the requirements of this specifications.

- NVMe Organization has a compliance program, use that for demonstrating v1.4b compliance
  - The ‘b’ indicates the integration of certain ECNs



How is the DatacenterSSD specification complementary to the NVMe specification?

- Example 2: Requiring or prohibiting features that are optional in the NVMe Specification

Requirement ID	Description
NVMe-IO-2	The device shall support the Dataset Management command. The device shall support the Attribute – Deallocate (AD) bit.
NVMe-IO-3	Since the device is power fail safe (e.g., has Power Loss Protection (PLP)) the performance shall not be degraded by any of the following: • FUA – i.e., forced unit access shall not incur a performance penalty. • Flush Cache – i.e., flush cache shall have no effect as the PLP makes any cache non-volatile. • Volatile Write Cache (Feature Identifier 06h) Set Feature to disable write-cache. This command shall be failed as described in the NVMe Standard 1.4b as there is no volatile write cache.
NVMe-IO-4	The device shall support the Write Zeroes command. The following bits of the Write Zeroes command shall be supported: • De-allocate (DEAC) bit. • Force Unit Access (FUA) bit.

Datacenter NVMe SSD Spec

Bytes	O/M [1]	Description
521:520	M	**Optional NVM Command Support (ONCS):** This field indicates the optional NVM commands and features supported by the controller. Refer to section 6. Bits 15:8 are reserved. Bit 7 if set to '1', then the controller supports the Verify command. If cleared to '0', then the controller does not support the Verify command. Bit 6 if set to '1', then the controller supports the Timestamp feature. If cleared to '0', then the controller does not support the Timestamp feature. Refer to section 5.21.1.14. Bit 5 if set to '1', then the controller supports reservations. If cleared to '0', then the controller does not support reservations. If the controller supports reservations then the following commands associated with reservations shall be supported: Reservation Report, Reservation Register, Reservation Acquire, and Reservation Release. Refer to section 8.8 for additional requirements. Bit 4 if set to '1', then the controller supports the Save field set to a non-zero value in the Set Features command and the Select field set to a non-zero value in the Get Features command. If cleared to '0', then the controller does not support the Save field set to a non-zero value in the Set Features command and the Select field set to a non-zero value in the Get Features command. Bit 3 if set to '1', then the controller supports the Write Zeroes command. If cleared to '0', then the controller does not support the Write Zeroes command. Bit 2 if set to '1', then the controller supports the Dataset Management command. If cleared to '0', then the controller does not support the Dataset Management command. Bit 1 if set to '1', then the controller supports the Write Uncorrectable command. If cleared to '0', then the controller does not support the Write Uncorrectable command. Bit 0 if set to '1', then the controller supports the Compare command. If cleared to '0', then the controller does not support the Compare command.

NVMe v1.4b Spec

How is the DatacenterSSD specification complementary to the NVMe specification?

- Example: Setting acceptable ranges/values for parameters defined in the NVMe Specification

NVMe-CFG-2	The device shall support a Maximum Data Transfer Size (MDTS) value of at least 256KB.

Datacenter NVMe SSD Spec

NVMe v1.4b Spec

How is the DatacenterSSD specification complementary to the NVMe specification?

- Example: Defining new Log Pages from 'Vendor Specific' ranges

Log Identifier	Scope	Log Page Name	Reference Section
C0h	NVM subsystem	SMART / Health Information Extended	4.8.5
C1h	NVM subsystem	Error Recovery	4.8.6
C2h	NVM subsystem	Firmware Activation History	4.8.7
C3h	Controller	Latency Monitor	4.8.9
C4h	NVM subsystem	Device Capabilities	4.8.10
C5h	NVM subsystem	Unsupported Requirements	4.8.11

KEY:
Namespace = The log page contains information about a specific namespace.
Controller = The log page contains information about the controller that is processing the command.
NVM subsystem = The log page contains information about the NVM subsystem.

Datacenter NVMe SSD Spec

Figure 195: Get Log Page – Log Page Identifiers

Log Identifier	Scope	Log Page Name	Reference Section
02h	Controller [1] / Namespace [2]	SMART / Health Information	5.14.1.2
03h	NVM subsystem	Firmware Slot Information	5.14.1.3
04h	Controller	Changed Namespace List	5.14.1.4
05h	Controller	Commands Supported and Effects	5.14.1.5
06h	Controller [3] / NVM subsystem [4]	Device Self-test [5]	5.14.1.6
07h	Controller	Telemetry Host-Initiated [5]	5.14.1.7
08h	Controller	Telemetry Controller-Initiated [5]	5.14.1.8
09h	NVM subsystem	Endurance Group Information	5.14.1.9
0Ah	NVM subsystem	Predictable Latency Per NVM Set	5.14.1.10
0Bh	NVM subsystem	Predictable Latency Event Aggregate	5.14.1.11
0Ch	Controller	Asymmetric Namespace Access	5.14.1.12
0Dh	NVM subsystem	Persistent Event Log [5]	5.14.1.13
0Eh	Controller	LBA Status Information	5.14.1.14
0Fh	NVM subsystem	Endurance Group Event Aggregate	5.14.1.15
10h to 6Fh	Reserved		
70h	Discovery (refer to the NVMe over Fabrics specification)		
71h to 7Fh	Reserved for NVMe over Fabrics implementations		
80h to BFh	I/O Command Set Specific		
C0h to FFh	Vendor specific [5]		

KEY:
Namespace = The log page contains information about a specific namespace.
Controller = The log page contains information about the controller that is processing the command.
NVM subsystem = The log page contains information about the NVM subsystem.

NOTES:
1. For namespace identifiers of 0h or FFFFFFFFh.
2. For namespace identifiers other than 0h or FFFFFFFFh.
3. Bit 0 is cleared to '0' in the DSTO field in the Identify Controller data structure (refer to Figure 251).
4. Bit 0 is set to '1' in the DSTO field in the Identify Controller data structure.
5. Selection of a UUID may be supported. Refer to section 8.24.

NVMe v1.4b Spec

How is the DatacenterSSD specification complementary to the NVMe specification?

- Example: Defining new Features from ‘Vendor Specific’ ranges
  - C0h: Error Injection
  - C1h: Clear FW Update History
  - C2h: EOL/PLP Failure Mode
  - C3h: Clear PCIe Correctable Error Counters
  - C4h: Enable IEEE1667 Silo
  - C5h: Latency Monitor
  - C6h: PLP Health Check Interval
  - C7h: DSSD Power State

Figure 275: Set Features – Feature Identifiers

Feature Identifier	Current Setting Persists Across Power Cycle and Reset [2]	Uses Memory Buffer for Attributes	Feature Name
78h to 7Fh	Refer to the NVMe Management Interface Specification for definition.		
80h to BFh			Command Set Specific (Reserved)
C0h to FFh			Vendor Specific [1, 5]
NOTES:			

NVMe v1.4b Spec

How is the DatacenterSSD specification complementary to the NVMe specification?

- Will these new features and log pages or any other requirements from the DatacenterSSD spec be adopted into the NVMe Specification? 😉

How can an SSD demonstrate compliance to the DatacenterSSD specification?

How can an SSD demonstrate compliance to the DatacenterSSD specification?

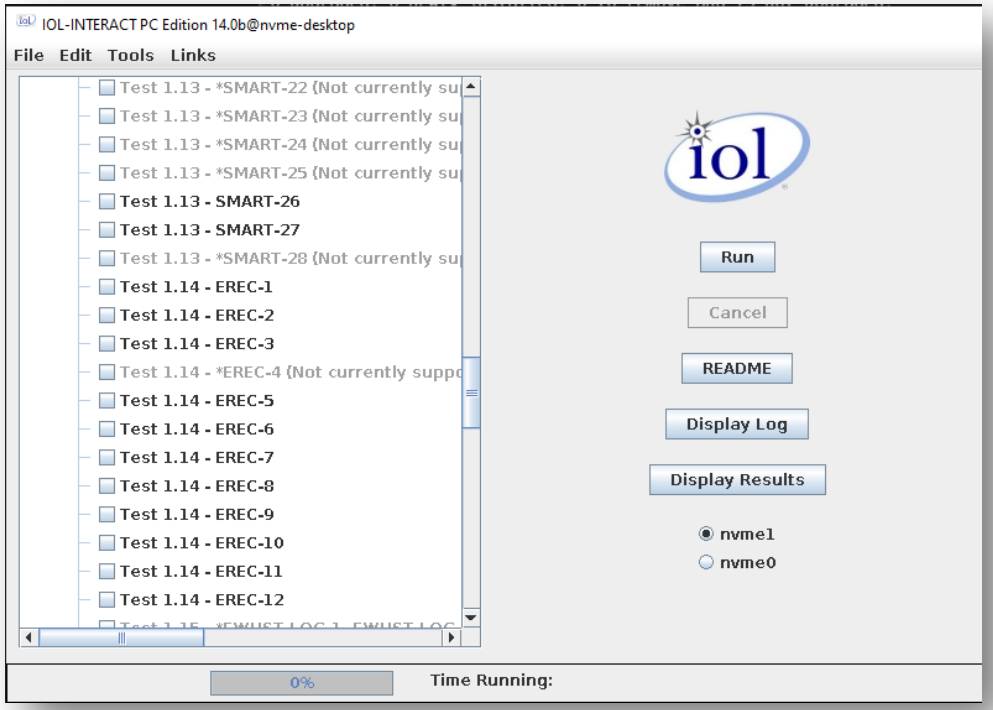
- DatacenterSSD Compliance is multi-faceted
  - Many requirements can be validated by testing at the NVMe interface, and can therefore be validated by third parties, similar to NVMe Compliance Testing
  - Other requirements need special access to prove, or have exceedingly long test times (such as requirements around Reliability and Endurance).
- DatacenterSSD Compliance is multi-disciplinary
  - Toolsets and Skills vary widely:
    - Labeling requirements
    - NVMe Interface requirements
    - NAND Endurance
    - Shock and Vibration
    - Altitude

How can an SSD demonstrate compliance to the DatacenterSSD specification?

- Key contributors are working on a dedicated Test Plan for the DatacenterSSD specification.
  - Expect publication later this year
  - UNH-IOL is partnering on that

How can an SSD demonstrate compliance to the DatacenterSSD specification?

- UNH-IOL will have a DatacenterSSD test service
- Some implementers may plan to test internally
- UNH-IOL is actively developing tools for testing DatacenterSSD requirements at the NVMe interface
  - Some test capability being demonstrated at customer sites



What tools and equipment are necessary to demonstrate compliance to the DatacenterSSD specification?

What tools and equipment are necessary to demonstrate compliance to the DatacenterSSD specification?

- DatacenterSSD spec is multi-faceted, not just an interface spec.
- Tool Needs:
  - ***NVMe protocol tools***
  - ***PCIe protocol tools***
  - ***Power Consumption***
  - ***PCIe Resets***
  - Thermal reporting and shutdown
  - SMBus Access
  - TCG Opal

Some protocol tests require shutdown capability to check persistence etc..

https://quarch.com/products/gen4-edsff-x4-breaker-module/

Conclusions

Conclusions: Learning Objectives

1. What is the DatacenterSSD specification?
   - A profile for implementing NVMe protocol and SSDs in a fashion more easily consumed by Datacenter implementers.
2. How is the DatacenterSSD specification complementary to the NVMe specification?
   - DCSSD Spec leverages existing NVMe capability and extends into new capabilities.
3. How can an SSD demonstrate compliance to the DatacenterSSD specification?
   - Whether testing with a third party or internally, tools are being developed to support DatacenterSSD testing.
4. What tools and equipment are necessary to demonstrate compliance to the DatacenterSSD specification?
   - In addition to SW tools to test at the NVMe interface some HW tools will be necessary for hot swap and power consumption testing

Please take a moment to rate this session.

Your feedback is important to us.

