

# OPEN

Compute Project

## Hyperscale NVMe Boot SSD Specification

Version 1.0

Author: Karthik Shivaram, Ross Stenfort - Meta

Author: Andrés Lagar-Cavilla, Charles Kunzman, Mike Branch, Alex Eisner, Michael Leung, Aaron Lee, Nicholas Maddy - Google

<b>License OWF Option</b>	<b>6</b>
<b>Overview</b>	<b>7</b>
<b>Scope</b>	<b>7</b>
<b>Drive Capacity Requirement</b>	<b>7</b>
<b>NVM Express Requirements</b>	<b>7</b>
Overview	7
NVMe Reset Supported	8
NVMe Controller Configuration and Behavior	8
Shutdown Notification Implementation (Graceful Power Cycle)	8
Time to Ready	8
NVMe Admin Command Set	9
UUID for OCP NVMe SSD Specific Information	10
Format NVM Implementation	10
Namespace Management	11
Identify Controller & Namespace	12
Firmware Update/Downgrade Requirements	12
Log Pages Support	13
Extended SMART Log Requirements	14
SMART / Health Information Extended (Log Identifier C0h)	15
Hardware Revision Log (Log Identifier C6h)	24
Set/Get Features Requirements	30
General Get Feature Requirements	30
Volatile Write Cache Settings	30
Power Management	30
Host Controlled Thermal Management	30
Clear PCIe Correctable Error Counters (Feature Identifier C3h) Set Feature	30
NVMe I/O Command Set	31
De-Allocation Requirements	32
Optional NVMe Features	32
<b>PCIe Requirements</b>	<b>33</b>
Overview	33
Compliance	33
Lane Width & Link Speed	33

Maximum Payload Size	33
Lane reversal	33
Boot Requirements	33
Reset Support	34
PCIe Error Logging	34
Low Power Modes	34
<b>Reliability</b>	<b>35</b>
UBER	35
SSD Operational Life	35
AFR (Annual Failure Rate)	35
End to End Data Protection	35
Background Data Refresh (BDR)	36
Background Data Flush (BDF)	36
Data Integrity	36
Command & Completion Timeout	37
EOL Requirements	37
Degraded Mode Behavior	38
<b>Endurance</b>	<b>38</b>
Endurance data	38
Retention conditions	39
Shelf Life	39
Endurance Targets	40
Wear Leveling	40
<b>Performance</b>	<b>41</b>
Boot SSD Performance Requirements	41
Boot Drive Sustained Bandwidth and IOPS Targets	41
Boot Drive Performance Measurement Process (PMP)	41
Boot-Drive Latency (QoS) Targets:	44
TRIM Performance	45
TRIM Rate Targets	45
BootBench	45
Max Latency	45
Resctl-Bench	46
<b>Security</b>	<b>48</b>

Basic Security Requirements	48
Secure Boot	51
Secure boot rooted in hardware.	51
Core Root of Trust Measurement.	51
<b>Debug &amp; Failure Analysis Support</b>	<b>52</b>
Debug Log Requirements	52
NVMe CLI Management Utility	53
NVMe CLI Plug-in Requirements	53
NVMe CLI Plug-In Nomenclature/Functional Requirements	54
NVMe-CLI Command Output Data Format	56
Human-Readable/ Plain Text Format	56
JSON Format	58
Performance Monitoring	60
<b>Mechanical</b>	<b>60</b>
Form factor	60
<b>Electrical</b>	<b>60</b>
Power consumption	60
Power Consumption Methodology & Requirements	60
Host Based Power & Thermal Management	61
Voltage Detector	61
SMBus Support	62
PCIe Link Equalization	62
GND Pins	62
<b>Thermal</b>	<b>62</b>
Operating Conditions	62
Data Center Altitude	62
Operational Temperature/ Relative Humidity	62
Non-Operational Temperature/Relative Humidity	63
Thermal Throttling	63
<b>Out-of-Band Management Support</b>	<b>64</b>
NVMe Basic Management Command (Appendix A) Requirements	64
VPD	65
NVMe Basic Management Command (Appendix A NVMe-MI Spec.) Data Format	65
NVMe-MI Requirements	67

<b>Labeling Requirements</b>	<b>68</b>
Label Requirements	68
<b>Environmental Considerations</b>	<b>74</b>
RoHS Compliance	74
ESD Compliance	74
<b>Sustainability Requirements</b>	<b>74</b>
<b>Shock and Vibration Requirements</b>	<b>74</b>
<b>Appendix A: CLA Format</b>	<b>76</b>
<b>Appendix B: Meta Unique Requirements</b>	<b>78</b>
Performance:	78
IO.go Targets	78
FileDelete & FileAppend Targets	78
Label:	78
Customer Defined Separator	78
<b>Appendix C: Google Unique Requirements</b>	<b>79</b>
Security:	79
Label:	79
Customer Defined Separator	79
Endurance Targets:	79
Google requires the device to meet the following requirements	79
<b>Appendix D: Guidance on Implementation of GUIDs</b>	<b>80</b>

## 1 License OWF Option

### 1.1. OPTION B: Open Web Foundation (OWF) CLA

Contributions to this Specification are made under the terms and conditions set forth in Open Web Foundation Contributor License Agreement (“OWF CLA 1.0”) (“Contribution License”) by:

Google

Meta

You can review the signed copies of the applicable Contributor License(s) for this Specification on the OCP website at <http://www.opencompute.org/products/specsanddesign> Usage of this Specification is governed by the terms and conditions set forth in **Open Web Foundation Final Specification Agreement (“OWFa 1.0”) (“Specification License”)**.

You can review the applicable Specification License(s) executed by the above referenced contributors to this Specification on the OCP website at <http://www.opencompute.org/participate/legal-documents/>

#### Notes:

1) The following clarifications, which distinguish technology licensed in the Contribution License and/or Specification License from those technologies merely referenced (but not licensed), were accepted by the Incubation Committee of the OCP:

NONE

2) The above license does not apply to the Appendix or Appendices. The information in the Appendix or Appendices is for reference only and non-normative in nature.

NOTWITHSTANDING THE FOREGOING LICENSES, THIS SPECIFICATION IS PROVIDED BY OCP "AS IS" AND OCP EXPRESSLY DISCLAIMS ANY WARRANTIES (EXPRESS, IMPLIED, OR

OTHERWISE), INCLUDING IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, FITNESS FOR A PARTICULAR PURPOSE, OR TITLE, RELATED TO Date: XXXX, 2XXXX Page 5 Open Compute Project • THE SPECIFICATION. NOTICE IS HEREBY GIVEN, THAT OTHER RIGHTS NOT GRANTED AS SET FORTH ABOVE, INCLUDING WITHOUT LIMITATION, RIGHTS OF THIRD PARTIES WHO DID NOT EXECUTE THE ABOVE LICENSES, MAY BE IMPLICATED BY THE IMPLEMENTATION OF OR COMPLIANCE WITH THIS SPECIFICATION. OCP IS NOT RESPONSIBLE FOR IDENTIFYING RIGHTS FOR WHICH A LICENSE MAY BE REQUIRED IN ORDER TO IMPLEMENT THIS SPECIFICATION. THE ENTIRE RISK AS TO IMPLEMENTING OR OTHERWISE USING THE SPECIFICATION IS ASSUMED BY YOU. IN NO EVENT WILL OCP BE LIABLE TO YOU FOR ANY MONETARY DAMAGES WITH RESPECT TO ANY CLAIMS RELATED TO, OR ARISING OUT OF YOUR USE OF THIS SPECIFICATION, INCLUDING BUT NOT LIMITED TO ANY LIABILITY FOR LOST PROFITS OR ANY CONSEQUENTIAL, INCIDENTAL, INDIRECT, SPECIAL OR PUNITIVE DAMAGES OF ANY CHARACTER FROM ANY CAUSES OF ACTION OF ANY KIND WITH RESPECT TO THIS SPECIFICATION, WHETHER BASED ON BREACH OF CONTRACT, TORT (INCLUDING NEGLIGENCE), OR OTHERWISE, AND EVEN IF OCP HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

## 2 Overview

This document is to define the requirements for a Hyperscale NVMe™ Boot SSD for use in data centers.

## 3 Scope

This document covers requirements for a PCIe-attached Boot SSD using NVM Express.

## 4 Drive Capacity Requirement

Requirement ID	Description
BOOT-CAP-1	The device should be offered in the following usable user capacities: <ul style="list-style-type: none"><li>128GB, 256GB, 512GB, 1024GB and 2048GB</li></ul>

## 5 NVM Express Requirements

### 5.1 Overview

Requirement ID	Description
NVMe-1	The device shall comply with all required features of the NVMe 1.4c Specification and those features amended by this specification.
NVMe-2	Any optional features supported by the device not described in this document shall be clearly documented and disclosed to the customer.
NVMe-3	Any vendor unique features supported by the device not described in this document shall be clearly documented and disclosed to the customer.

## 5.2 NVMe Reset Supported

Requirement ID	Description
NVMeR-1	NVMe Controller Reset (CC.EN cleared to 0b) shall be supported.

## 5.3 NVMe Controller Configuration and Behavior

Requirement ID	Description
NVMe-CFG-1	The default arbitration shall be Round-Robin.
NVMe-CFG-2	The device shall support a Maximum Data Transfer Size (MDTS) value of at least 256KB.
NVMe-CFG-3	The device firmware shall support reporting of Controller Fatal Status (CSTS.CFS).

### 5.3.1 Shutdown Notification Implementation (Graceful Power Cycle)

Requirement ID	Description
SHN-1	The Shutdown Notification completion (CSTS.SHST) shall be received within 5s of setting CC.SHN bit to 1b, if RTD3 entry latency is not supported.
SHN-2	The device shall support the CC.SHN Normal and Abrupt Shutdown Notifications.
SHN-3	When safe shutdown is completed successfully, the device should not enter a rebuild/recovery mode on the next power on.
SHN-4	Shutdown Notification shall trigger flushing of all content within the device's internal volatile areas (For example: SRAM/ DRAM cache).
SHN-5	In case of Normal shutdown operation when CSTS.SHST is set to 10b, no data loss is tolerated.
SHN-6	An unexpected shutdown event shall not make the device non-functional under any conditions.

### 5.3.2 Time to Ready

Requirement ID	Description
TTR-1	The device is expected to service I/O and ADMIN commands as soon as CSTS.RDY is set to 1b.
TTR-2	The device shall keep CSTS.RDY = 0 until the device is able to service commands.



TTR-3A	The device shall become ready indicated by setting CSTS.RDY=1 within 20s in case of an unexpected shutdown.																					
TTR-3B	The device shall become ready within 8s in case of an expected shutdown.																					
TTR-4	The CAP.TO register shall reflect the worst-case time required to be ready as defined in TTR-3A.																					
TTR-5	<div>After CSTS.RDY is set to 1b, the device shall complete the following commands in 1 second from when a valid command is fetched from the submission queue.</div> <table><tr><th>Type of Command</th><th>Command Description</th><th>Command Opcode</th></tr><tr><td rowspan="7">ADMIN</td><td>Identify</td><td>0x6</td></tr><tr><td>Get Feature (FID 0x1, 0x6, 0x7, 0x8, 0x9, 0xB)</td><td>0xA</td></tr><tr><td>Set Feature (FID 0x1, 0x6, 0x7, 0x8, 0x9, 0xB)</td><td>0x9</td></tr><tr><td>Delete I/O Submission Queue</td><td>0x0</td></tr><tr><td>Delete I/O Completion Queue</td><td>0x4</td></tr><tr><td>Create I/O Submission Queue</td><td>0x1</td></tr><tr><td>Create I/O Completion Queue</td><td>0x5</td></tr><tr><td>IO</td><td>Read</td><td>0x2</td></tr></table>	Type of Command	Command Description	Command Opcode	ADMIN	Identify	0x6	Get Feature (FID 0x1, 0x6, 0x7, 0x8, 0x9, 0xB)	0xA	Set Feature (FID 0x1, 0x6, 0x7, 0x8, 0x9, 0xB)	0x9	Delete I/O Submission Queue	0x0	Delete I/O Completion Queue	0x4	Create I/O Submission Queue	0x1	Create I/O Completion Queue	0x5	IO	Read	0x2
Type of Command	Command Description	Command Opcode																				
ADMIN	Identify	0x6																				
	Get Feature (FID 0x1, 0x6, 0x7, 0x8, 0x9, 0xB)	0xA																				
	Set Feature (FID 0x1, 0x6, 0x7, 0x8, 0x9, 0xB)	0x9																				
	Delete I/O Submission Queue	0x0																				
	Delete I/O Completion Queue	0x4																				
	Create I/O Submission Queue	0x1																				
	Create I/O Completion Queue	0x5																				
IO	Read	0x2																				
TTR-6	When the CC.SHN register is written to notify the device to shutdown it shall not be assumed that power will be lost even after CC.EN is cleared to 0b and CC.SHN is set to 00b. Under these conditions the device shall continue to meet the reliability requirements.																					

#### 5.4 NVMe Admin Command Set

The device shall support the following mandatory and optional NVMe admin commands:

Requirement ID	Description
NVMe-AD-1	The device shall support all mandatory NVMe admin commands.
NVMe-AD-2	<p>Identify – The following mandatory and optional CNS values shall be supported:</p> <ul style="list-style-type: none"> <li>• CNS 0x0 (Identify Namespace Data Structure)</li> <li>• CNS 0x1 (Identify Controller Data Structure)</li> <li>• CNS 0x2 (Active Namespace List)</li> <li>• CNS 0x3 (Namespace Identification Descriptor list)</li> <li>• CNS 0x12 (Controller List of controllers attached to the specified NSID)</li> <li>• CNS 0x13 (Controller List of controllers that exist in the NVM subsystem)</li> </ul>
NVMe-AD-3	Namespace Management command shall be supported.
NVMe-AD-4	Namespace Attachment command shall be supported.
NVMe-AD-5	<p>Format NVM command shall be supported. The following Secure Erase Settings (SES) settings shall be supported:</p> <ul style="list-style-type: none"> <li>• 001b (<i>User Data Erase</i>)</li> <li>• 010b (<i>Crypto Erase</i>)</li> </ul>

NVMe-AD-6	The device shall support the Sanitize command and meet NIST SP800-88r1 Purge requirements. Block Erase (010b) and Crypto Erase (100b) sanitize operations shall be supported.
NVMe-AD-7	The device shall enable reads to sanitized LBAs to meet validation of sanitized areas per NIST SP800-88r1. Specifically, No Deallocate After Sanitize shall not be supported. Sanitize Capabilities (SANICAP) bit No Deallocate Inhibited (NDI) shall be set to 1 to indicate that Deallocate will always be performed during Sanitize. Consequently, LBAs shall return all 0s or all 1s after a successful Sanitize command in any mode.
NVMe-AD-8	If a Read occurs to a Sanitized LBA prior to that LBA being written, the device shall complete the Read and return successful completion status.
NVMe-AD-9	Firmware Image Download command shall be supported.
NVMe-AD-10	Firmware Commit command shall be supported.
NVMe-AD-11	Device Self-Test command shall be supported.
NVMe-AD-12	The device shall support Identify command UUID List functionality (CNS value 17h).

#### 5.4.1 UUID for OCP NVMe SSD Specific Information

A UUID has been defined for use in commands to ensure that the vendor specific Log Identifiers and Feature Identifiers used in this specification access the functionality defined in this specification (i.e., and do not access other vendor specific functionality that may use the same vendor specific identifiers).

Requirement ID	Description
UUID-1	The UUID List (NVMe-AD-12) shall contain a UUID List Entry that contains the UUID value EDDACD68-474C-4354-8758-D8182AA54B32. The Identifier Association field in that UUID List Entry shall be cleared to 00b.
UUID-2	The Get Features and Set Features commands shall support UUID Index functionality.
UUID-3	A Get Features command or a Set Features command with: <ul style="list-style-type: none"> <li>the UUID Index of the UUID (UUID-1) in the UUID List (NVMe-AD-12) or a zero UUID Index;</li> <li>and a vendor-specific Feature Identifier that is used in this specification (See Section 5.4.7) shall access the vendor specific Feature defined in this specification.</li> </ul>
UUID-4	The Get Log Page command shall support UUID Index functionality.
UUID-5	A Get Log Page command with: <ul style="list-style-type: none"> <li>the UUID Index of the UUID (UUID-1) in the UUID List (NVMe-AD-12) or a zero UUID Index;</li> <li>and a vendor-specific Log Page Identifier that is used in this specification (See Section 5.4.6) shall access the vendor specific Log Page defined in this specification.</li> </ul>

#### 5.4.2 Format NVM Implementation

This section describes the behavior of NVM Format ADMIN command.

Requirement ID	Description
FORMAT-1	Format NVM Behavior: <ul style="list-style-type: none"> <li>- <i>Crypto Erase</i> option which deletes all the encryption keys shall be supported.</li> <li>- <i>User Data Erase</i> option shall be supported and wipes (i.e. erases from the physical media) the full Namespace Capacity (including grown bad-blocks, within deallocated LBAs, over-provisioned blocks etc.)</li> </ul>
FORMAT-2	Format shall not be blocked by any Security Protocol Command (such as <i>Block SID</i> or <i>Security Freeze Lock</i> ) issued by the BIOS and the device is in an UNLOCKED state.
FORMAT-3	Format shall not be blocked due to issuance of TCG BlockSID command.
FORMAT-4	A successful completion status (00h) of Format NVM command with the SES field set to 0x1 (User Data Erase) will mean that all user data has been wiped off (i.e., erased from the physical media) the device.
FORMAT-5	SMART Attributes (including <i>Extended SMART</i> ) shall not be reset on completion of a Format command.

### 5.4.3 Namespace Management

The namespace management command along with the attach/detach commands may be used to increase SSD over-provisioning beyond the default minimum over-provisioning.

Requirement ID	Description																		
NSM-1	From the factory, the device shall have one namespace whose size is the maximum capacity supported.																		
NSM-2	<div>The device shall support being configured to the following OP settings using Namespace Management for a 512GB device capacity:</div> <table><tr><th>Typical OP Usage</th><th>LBA Count (Hex) (512B)</th><th>LBA Count (Dec) (512B)</th></tr><tr><td>OP = 7%</td><td>0x3771BA50</td><td>930200144</td></tr><tr><td>OP = 20%</td><td>0x2FBDA570</td><td>800957808</td></tr><tr><td>OP = 28%</td><td>0x2AECB148</td><td>720154952</td></tr><tr><td>OP = 50%</td><td>0x1DCF0958</td><td>500107608</td></tr><tr><td>OP=Default (0%)</td><td>0x3B9E12B0</td><td>1000215216</td></tr></table>	Typical OP Usage	LBA Count (Hex) (512B)	LBA Count (Dec) (512B)	OP = 7%	0x3771BA50	930200144	OP = 20%	0x2FBDA570	800957808	OP = 28%	0x2AECB148	720154952	OP = 50%	0x1DCF0958	500107608	OP=Default (0%)	0x3B9E12B0	1000215216
Typical OP Usage	LBA Count (Hex) (512B)	LBA Count (Dec) (512B)																	
OP = 7%	0x3771BA50	930200144																	
OP = 20%	0x2FBDA570	800957808																	
OP = 28%	0x2AECB148	720154952																	
OP = 50%	0x1DCF0958	500107608																	
OP=Default (0%)	0x3B9E12B0	1000215216																	
NSM-3	Asynchronous event notification related to Namespace Management shall be supported, specifically <i>Namespace Attribute Changed</i> shall be supported.																		
NSM-4	When creating a namespace, the default “Formatted LBA Size” parameter (FLBAS = 0) in the Identify Namespace Data Structure (Byte 26) shall correspond to the default sector size set at the factory.																		
NSM-5	When formatting the device with the Format command, the default “LBA Format” parameter (LBAF = 0) in Command DWORD 10 bits 3:0 shall correspond to the default sector size set at the factory.																		
NSM-6	Namespace Utilization (NUSE) shall be supported. The NUSE shall be equal to the number of logical blocks currently allocated in the namespace. NUSE shall																		

	<p>not be hardcoded to be equal to NCAP. See below for an example on a 200GB device:</p> <ol style="list-style-type: none"> <li>1. <i>After a Format NVM command User Data Erase (SES = 001b) or NVM Deallocate is completed, NUSE would be zero. And the usage data would reflect that: 0.00GB.</i></li> <li>2. <i>After writing 1GB worth of data, the usage data may show the following: 1.00GB.</i></li> <li>3. <i>After filling the device, the usage data may show the following: 200.00GB.</i></li> <li>4. <i>If the host issues a 10GB de-allocate command and de-allocated the data, the usage data would show the following: 190.00GB.</i></li> </ol>
NSM-7	An over-provisioned device shall provide performance and endurance benefits of over-provisioning.

#### 5.4.4 Identify Controller & Namespace

Requirement ID	Description
IDENTIFY-1	The field tNVMCAP shall be populated to reflect the total NVM capacity in the NVM subsystem, in bytes.
IDENTIFY-2A	<p>The device shall support 512-byte logical block sizes as the default sector size with the following LBA Format structure settings:</p> <ul style="list-style-type: none"> <li>• <i>Relative Performance(RP): 0x0 (Best)/ 0x1 (Better)/ 0x2 (Good)</i></li> <li>• <i>LBA Data Size (LBADS): 9 (512 byte sector size)</i></li> <li>• <i>Metadata Size (MS): 0x0 (No Metadata)</i></li> </ul>
IDENTIFY-2B	<p>The device shall support 4096-byte logical block size with the following LBA Format structure settings:</p> <ul style="list-style-type: none"> <li>• <i>Relative Performance(RP): 0x0 (Best)/ 0x1 (Better)/ 0x2 (Good)</i></li> <li>• <i>LBA Data Size (LBADS): 12 (4096 byte sector size)</i></li> <li>• <i>Metadata Size (MS): 0x0 (No Metadata)</i></li> </ul>
IDENTIFY-3	The device shall support a minimum of 1 Namespace.
IDENTIFY-4	The Serial Number (SN), Model Number (MN), and FRU Globally Unique Identifier (FGUID) shall not change under any conditions.
IDENTIFY-5	The Serial Number and Model Number shall not contain any of the customer-defined separators (see LABL-5).

#### 5.4.5 Firmware Update/Downgrade Requirements

Requirement ID	Description
FWUP-1	Devices shall not have any restrictions on the number of firmware downloads supported.
FWUP-2	<p>The Firmware Commit command with the following Commit Action (CA) codes shall be supported:</p> <ul style="list-style-type: none"> <li>• <i>000b – Download only.</i></li> <li>• <i>001b – Download and activate upon reset.</i></li> <li>• <i>010b – Activate upon reset.</i></li> </ul>

	<ul style="list-style-type: none"> <li>011b – Activate immediately without reset.</li> </ul>
FWUP-3	For firmware commit action 011b (firmware activation without reset), the device shall complete the firmware activation process and be ready to accept host IO and admin commands within 10 seconds from the receipt of the Firmware Commit command. The Maximum Time for Firmware Activation (MTFA) field shall not exceed 64h.
FWUP-4	When attempting to downgrade to incompatible firmware revision, the device shall return Firmware Activation Prohibited (13h) status to a Firmware Commit command. All lower security revision firmware images shall be considered incompatible.
FWUP-5	The device shall not become non-functional upon upgrade/ downgrade under any conditions.
FWUP-6	The firmware image in each valid firmware slot shall have multiple copies of the firmware image for reliability. The corruption of a single copy of the firmware image shall not result in the device no longer functioning.
FWUP-7	Unless others specified by this specification, firmware activation shall not cause data to be lost or destroyed. e.g. User Data, Log Pages (e.g. SMART Data), Internal Logs & Data-Structures etc.
FWUP-8	Firmware activation without reset shall preserve the running state of the device (e.g., Opal locking state, Set Features, Timestamp, etc.).

#### 5.4.6 Log Pages Support

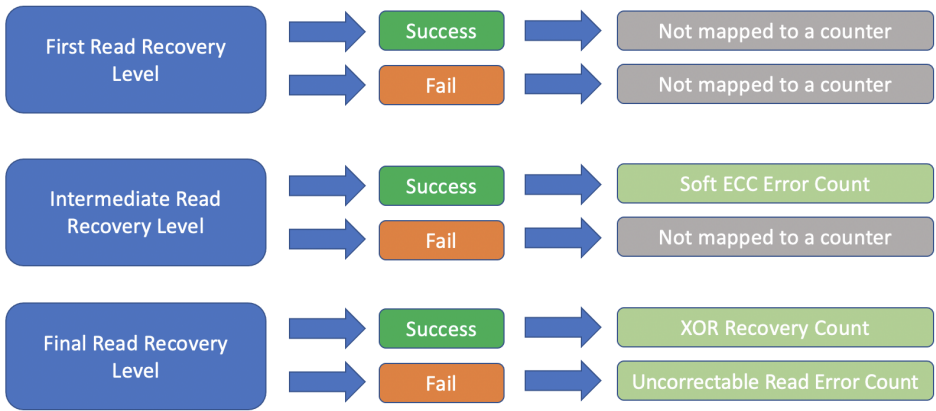
This section describes log page support for this device.

Requirement ID	Description																
LP-1	All mandatory log pages as specified in NVMe specifications 1.4c.																
LP-2	The Percentage Used field in the SMART / Health Information (Log Identifier 02h) shall be based on the average P/E cycle of the device. The Percentage Used field in SMART/ Health Information shall not be affected by the Power on Hours (POH) of the device.																
LP-3	Under no conditions shall the Percentage Used field in the SMART / Health Information (Log Identifier 02h) be reset.																
LP-4	Commands Supported and Effects (Log Identifier 05h) shall be supported.																
LP-5	Telemetry Host-Initiated (Log Identifier 07h) shall be supported.																
LP-6	Telemetry Controller-Initiated (Log Identifier 08h) shall be supported.																
LP-7	Persistent Event Log (Log Identifier 0Dh) shall be supported.																
LP-8	<p>The following Persistent Event Log types shall be supported:</p> <table> <tr> <th>Type</th><th>Event</th></tr> <tr> <td>01h</td><td>SMART / Health Log Snapshot</td></tr> <tr> <td>02h</td><td>Firmware Commit</td></tr> <tr> <td>03h</td><td>Timestamp Change</td></tr> <tr> <td>04h</td><td>Power-on or Reset</td></tr> <tr> <td>05h</td><td>NVM Subsystem Hardware Error</td></tr> <tr> <td>06h</td><td>Change Namespace</td></tr> <tr> <td>07h</td><td>Format NVM Start</td></tr> </table>	Type	Event	01h	SMART / Health Log Snapshot	02h	Firmware Commit	03h	Timestamp Change	04h	Power-on or Reset	05h	NVM Subsystem Hardware Error	06h	Change Namespace	07h	Format NVM Start
Type	Event																
01h	SMART / Health Log Snapshot																
02h	Firmware Commit																
03h	Timestamp Change																
04h	Power-on or Reset																
05h	NVM Subsystem Hardware Error																
06h	Change Namespace																
07h	Format NVM Start																

		08h	Format NVM Completion	
		09h	Sanitize Start	
		0Ah	Sanitize Completion	
		0Ch	Telemetry Log Created	
		0Dh	Thermal Excursion	
LP-9	The device shall support extended SMART log page (LID = C0h). Refer to section 5.4.6.2.			

#### 5.4.6.1 Extended SMART Log Requirements

Requirement ID	Description
SLOG-1	All values in the Vendor Log pages defined by this specification shall be persistent across graceful power cycles & resets unless otherwise specified. In the event of unclean shutdown, data loss shall be limited to recent data per SLOG-11.
SLOG-2	All counters defined by this specification shall be saturating counters unless otherwise specified. A saturating counter is defined as a counter that stops incrementing when it reaches its maximum value (as limited by the number of bits allocated to the counter or logical value limits), and does NOT roll over to 0.
SLOG-3	Unless otherwise specified, the device shall update all SMART and Extended SMART log page values in the background at least once every ten minutes (as indicated by BK-FL-1).
SLOG-4	All values defined by this specification in logs shall be little endian format unless otherwise specified.
SLOG-5	A normalized counter in the Extended SMART Log, unless otherwise specified, shall be reported as the following: 100% shall represent the number at factory exit. 1% shall represent the minimum amount to be reliable. A value of 0% means the device shall no longer be considered reliable. 100% shall be represented as 64h.
SLOG-6	Devices shall support the attributes listed in Section 5.14.1.2 of the NVMe specification and extended log page C0h as described in <a href="#">section</a> 5.4.6.2 of this specification.
SLOG-7	A Read of either the SMART /Health Information (Log Identifier 02h) or SMART / Health Information Extended (Log Identifier C0h) shall not require an update of the SMART values except for SMART attributes listed in SLOG-10.
SLOG-8	Firmware Download & Activate without reset event shall not corrupt user data, SMART data or telemetry logs. Firmware Download & Activate with reset event shall not corrupt user data, SMART data or telemetry controller initiated log.
SLOG-9	NVMeR-1, PCIeRST-1, PCIeRST-2 shall not corrupt any of the SMART attributes.
SLOG-10	The composite and raw temperature sensor values and EXT-SMART-16 shall be updated when the log page is accessed.
SLOG-11	The device may not lose data in either the SMART / Health Information (Log Identifier 02h) or SMART / Health Information Extended (Log Identifier C0h) which is more than 10 minutes old across a normal/ unsafe/ abrupt shutdown.

SLOG-12	<p>The device shall map the Read Error Handling counters as follows based on a NAND page:</p> <ul style="list-style-type: none"> <li>Soft ECC Error Count will be incremented when the first level of error recovery is entered which may involve read retry, soft LDPC error recovery or other techniques before proceeding to XOR for recovery.</li> <li>If XOR Recovery is supported, then XOR Recovery Count will be incremented every time a XOR/ RAID recovery is triggered.</li> <li>Uncorrectable Read Error Counter will be incremented every time the device returns a Unrecovered Read Error (81h) status as defined in the NVMe Specifications.</li> </ul>  <pre> graph LR     subgraph "First Read Recovery Level"         F1[First Read Recovery Level] --&gt; S1[Success]         F1 --&gt; F1_1[Fail]         S1 --&gt; N1[Not mapped to a counter]         F1_1 --&gt; N1_1[Not mapped to a counter]     end     subgraph "Intermediate Read Recovery Level"         F2[Intermediate Read Recovery Level] --&gt; S2[Success]         F2 --&gt; F2_1[Fail]         S2 --&gt; S2_1[Soft ECC Error Count]         F2_1 --&gt; N2[Not mapped to a counter]     end     subgraph "Final Read Recovery Level"         F3[Final Read Recovery Level] --&gt; S3[Success]         F3 --&gt; F3_1[Fail]         S3 --&gt; S3_1[XOR Recovery Count]         F3_1 --&gt; S3_2[Uncorrectable Read Error Count]     end </pre> <p>Figure 1 Read Error-Handler Extended SMART Rules</p>
SLOG-13	Extended SMART log page identifier C0h shall be extractable & decoded using a NVMe-CLI Plugin.
SLOG-14	Reading the extended SMART log shall not block IO or have an impact on device performance.
SLOG-15	All counters in the extended SMART log page shall be tracked and incremented at all times and under all device states (e.g. active IO, background operation etc.)

#### 5.4.6.2 SMART / Health Information Extended (Log Identifier C0h)

This vendor-specific log page, C0h shall be 512-bytes with the following functional requirements and field format:

Req ID	Attribute Name	# of Bytes	Byte Address	Field Description
EXT-SMART-1	Physical Media Units Written – TLC	16	15:0	Contains the number of 512-byte data units written to the TLC media; this value includes metadata written to the non-out-of-band area in the media. When the LBA size is a value other than 512 bytes, the device shall

				convert the amount of data written to 512-byte units. This value is reported in thousands (i.e., a value of 1 corresponds to 1000 units of 512 bytes written) and is rounded up. It must be possible to use this attribute to calculate the Write Amplification Factor (WAF).									
EXT-SMART-2	Physical Media Units Written – SLC	16	31:16	Contains the number of 512-byte data units written to the SLC media; this value includes metadata written to the non-out-of-band area in the media. When the LBA size is a value other than 512 bytes, the device shall convert the amount of data written to 512-byte units. This value is reported in thousands (i.e., a value of 1 corresponds to 1000 units of 512 bytes written) and is rounded up. It must be possible to use this attribute to calculate the Write Amplification Factor (WAF).									
EXT-SMART-3	Bad User NAND Block Count	8	39:32	Raw and normalized count of the number of user NAND blocks that have been retired. On factory exit, the normalized value shall be set to 0x64 and the Raw count shall be set to zero. It should be noted there are 2 bytes for normalized and 6 bytes for raw count. See normalized definition in SLOG-5. <table><tr><th># of Bytes</th><th>Byte Address</th><th>Field Description</th></tr><tr><td>2</td><td>33:32</td><td>Normalized value</td></tr><tr><td>6</td><td>39:34</td><td>Raw count</td></tr></table>	# of Bytes	Byte Address	Field Description	2	33:32	Normalized value	6	39:34	Raw count
# of Bytes	Byte Address	Field Description											
2	33:32	Normalized value											
6	39:34	Raw count											
EXT-SMART-4	XOR Recovery count	8	47:40	Total number of times XOR was invoked to recover data, if XOR Recovery is supported. Data recovery may have succeeded or failed.									
EXT-SMART-5	Uncorrectable read error count	8	55:48	Total count of NAND reads that were not correctable by read retries, all levels of ECC, or XOR. This is a count of the number of times data recovery fails and an uncorrectable read error is returned to the host.									
EXT-SMART-6	SSD End to End correction counts	24	79:56	A count of the detected and corrected errors by the SSD end to end error correction which includes DRAM, SRAM, or other storage element ECC/CRC protection mechanism (not NAND ECC). All correctable errors must result									




				<p>in a counter increase no matter what type of data the memory is protecting. All detected and uncorrectable errors must result in a counter increase. It should be noted there are 8 bytes for count of detected errors and 8 bytes for count of un-correctable errors and 8 bytes for correctable errors.</p> <table><tr><th># of Bytes</th><th>Byte Address</th><th>Field Description</th></tr><tr><td>8</td><td>63:56</td><td>Corrected Errors</td></tr><tr><td>8</td><td>71:64</td><td>Detected Errors</td></tr><tr><td>8</td><td>79:72</td><td>Uncorrected E2E Errors</td></tr></table>	# of Bytes	Byte Address	Field Description	8	63:56	Corrected Errors	8	71:64	Detected Errors	8	79:72	Uncorrected E2E Errors
# of Bytes	Byte Address	Field Description														
8	63:56	Corrected Errors														
8	71:64	Detected Errors														
8	79:72	Uncorrected E2E Errors														
EXT-SMART-7	System data % life-used	1	80	<p>A normalized cumulative count of the number of erase cycles per block since leaving the factory for the system (FW and metadata) area. Starts at 0 and increments. 100 indicates that the estimated endurance has been consumed. Value is allowed to exceed 100 up to 255.</p> <p>The firmware may report instantaneous value by calculating information on-the-fly when a Get Log Page command is issued by the host to Log Page C0h.</p>												
	Reserved	3	83:81													
EXT-SMART-8	User data erase counts (program/ erase counter)	48	131:84	<p>The maximum and minimum erase counts across all NAND blocks in the drive. The host shall not be able to reset this counter. It should be noted there are 8 bytes for the maximum and 8 bytes for the minimum.</p> <table><tr><th># of Bytes</th><th>Byte Address</th><th>Field Description</th></tr><tr><td>8</td><td>91:84</td><td>Minimum User Data Erase Count (TLC)</td></tr><tr><td>8</td><td>99:92</td><td>Maximum User Data Erase Count (TLC)</td></tr><tr><td>8</td><td>107:100</td><td>Average User Data Erase Count (TLC)</td></tr></table>	# of Bytes	Byte Address	Field Description	8	91:84	Minimum User Data Erase Count (TLC)	8	99:92	Maximum User Data Erase Count (TLC)	8	107:100	Average User Data Erase Count (TLC)
# of Bytes	Byte Address	Field Description														
8	91:84	Minimum User Data Erase Count (TLC)														
8	99:92	Maximum User Data Erase Count (TLC)														
8	107:100	Average User Data Erase Count (TLC)														

				<table><tr><td>8</td><td>115:108</td><td>Minimum Erase Count (SLC)</td></tr><tr><td>8</td><td>123:116</td><td>Maximum Erase Count (SLC)</td></tr><tr><td>8</td><td>131:124</td><td>Average Erase Count (SLC)</td></tr></table>	8	115:108	Minimum Erase Count (SLC)	8	123:116	Maximum Erase Count (SLC)	8	131:124	Average Erase Count (SLC)
8	115:108	Minimum Erase Count (SLC)											
8	123:116	Maximum Erase Count (SLC)											
8	131:124	Average Erase Count (SLC)											
EXT-SMART-9	Program fail count	8	139:132	<p>Raw and normalized count of total program failures. Normalized count starts at 100 and shows the percent of remaining allowable failures. It should be noted there 2 bytes for normalized and 6 bytes for raw count.</p> <table><tr><th># of Bytes</th><th>Byte Address</th><th>Field Description</th></tr><tr><td>2</td><td>133:132</td><td>Normalized Program Fail Count</td></tr><tr><td>6</td><td>139:134</td><td>Raw Program Fail Count</td></tr></table>	# of Bytes	Byte Address	Field Description	2	133:132	Normalized Program Fail Count	6	139:134	Raw Program Fail Count
# of Bytes	Byte Address	Field Description											
2	133:132	Normalized Program Fail Count											
6	139:134	Raw Program Fail Count											
EXT-SMART-10	Erase Fail Count	8	147:140	<p>Raw and normalized count of total erase failures in the user &amp; system area. Normalized count starts at 100 and shows the percent of remaining allowable failures. It should be noted there are 2 bytes for normalized and 6 bytes for raw count.</p> <table><tr><th># of Bytes</th><th>Byte Address</th><th>Field Description</th></tr><tr><td>2</td><td>141:140</td><td>Normalized Erase Fail Count</td></tr><tr><td>6</td><td>147:142</td><td>Raw Erase Fail Count</td></tr></table>	# of Bytes	Byte Address	Field Description	2	141:140	Normalized Erase Fail Count	6	147:142	Raw Erase Fail Count
# of Bytes	Byte Address	Field Description											
2	141:140	Normalized Erase Fail Count											
6	147:142	Raw Erase Fail Count											
EXT-SMART-11	PCIe Correctable Error count	8	155:148	Summation counter of all PCIe correctable errors (Bad TLP, Bad DLLP, Receiver error, Replay timeouts, Replay rollovers). These counts shall only increment during run time. They shall not increment during training or power fail.									
EXT-SMART-12	% Free Blocks (User)	1	156	A normalized count of the number of blocks that are currently free (available) out of the total pool of spare (invalid) blocks. Free blocks mean both blocks that have been erased and blocks that have all invalid									

				<p>data. Invalid blocks are blocks that are either marked invalid by drive FW OR by the host (via TRIM or overwrite).</p> <p>For example, if the total number of spare blocks are 100 and garbage collection has been able to reclaim 20 blocks, then this field reports 20%.</p> <p>The firmware may report instantaneous value by calculating information on-the-fly when a Get Log Page command is issued by the host to Log Page C0h.</p>						
	Reserved	3	159:157							
EXT-SMART-13	Security Version Number	8	167:160	This is the security version number of the firmware image. The firmware increments this number any time it includes a fix of a security issue. This is not a saturating counter.						
EXT-SMART-14	% Free Blocks (System)	1	168	<p>A normalized count of the number of blocks that are currently free (available) out of the total pool of spare (invalid) blocks allocated for system area (<i>internal firmware management</i>) of the drive.</p> <p>Free blocks mean blocks that have been erased &amp; ready to use for internal firmware system area operations.</p> <p>The firmware may report instantaneous value by calculating information on-the-fly when a Get Log Page command is issued by the host to Log Page C0h .</p>						
	Reserved	3	171:169							
EXT-SMART-15	NVMe Stats	26	197:172	<p>The counter shall be persistent across graceful power-cycles.</p> <table><tr><th># of Bytes</th><th>Byte Address</th><th>Field Description</th></tr><tr><td>16</td><td>187:172</td><td>Contains the number of Data Set Management (Deallocate) commands completed by the device.</td></tr></table>	# of Bytes	Byte Address	Field Description	16	187:172	Contains the number of Data Set Management (Deallocate) commands completed by the device.
# of Bytes	Byte Address	Field Description								
16	187:172	Contains the number of Data Set Management (Deallocate) commands completed by the device.								

						This shall not depend on completion of the deallocate operation.
				8	195:188	Total Namespace Utilization. Shall be the sum of the Namespace Utilization field defined in the Identify Namespace Data Structure bytes 23:16 across all configured namespaces.
				2	197:196	Contains the number of times the device has successfully completed a Format NVM or Sanitize command. Include all sub-options of the command.
EXT-SMART-16	Background Back-Pressure Gauge	1	198	<p>This field shall indicate an estimate and provide insight into an occurrence of an IO Stall event by monitoring the level of pending background tasks (GC, WL, Refresh, Block-Erase, TRIM etc.). The field shall be updated on-the-fly when a Get-Log-Page is issued:</p> <ul style="list-style-type: none"> <li>• A value of 100% means the firmware is under high-pressure to perform a lot of Background Tasks (&amp; a lot of activity is left pending). This may also mean that on-going IO is stalled to process a Background task in Foreground.</li> </ul>		

				<ul style="list-style-type: none"><li>A value of 0 means the firmware is under no-pressure to perform any maintenance (i.e. <i>all background tasks are complete</i>)</li><li>The counter shall not be persistent across power-cycles.</li></ul> <div><div>Too many Background Tasks are pending (all BG tasks are running in Foreground)</div><div>Moderate amount of B/G activity</div><div>Too few/ no pending Background Task</div></div>									
	Reserved	3	201:199										
EXT-SMART-17	Soft ECC error count	8	209:202	Total count of NAND reads that were not correctable by first level ECC and requires invoking an intermediate recovery. This shall cover all NAND read accesses. Data recovery may have succeeded or failed. If the device has more than one intermediate recovery level, then this counter only increments when intermediate recovery level 1 is invoked.									
EXT-SMART-18	Refresh count	8	217:210	A count of the number of blocks that have been re-allocated to maintain data integrity. This counter does not include creating free space due to garbage collection.									
EXT-SMART-19	Bad System NAND block count	8	225:218	<p>Raw and normalized count of the number of system NAND blocks that have been retired. On factory exit, the normalized value shall be set to 0x64 and the Raw count shall be set to zero. It should be noted there are 2 bytes for normalized and 6 bytes for raw count. See normalized definition in SLOG-5.</p> <table><tr><th># of Bytes</th><th>Byte Address</th><th>Field Description</th></tr><tr><td>2</td><td>219:218</td><td>Normalized value</td></tr><tr><td>6</td><td>225:220</td><td>Raw count</td></tr></table> <p>A value of 0xFFFF_FFFF_FFFF_FFFF indicates that the Bad User NAND block count field above represents all blocks on the device and the Bad System NAND block count field is invalid.</p>	# of Bytes	Byte Address	Field Description	2	219:218	Normalized value	6	225:220	Raw count
# of Bytes	Byte Address	Field Description											
2	219:218	Normalized value											
6	225:220	Raw count											

EXT-SMART-20	Endurance Estimate	16	241:226	This field is an estimate of the total number of data bytes that may be written to the device over its lifetime assuming a write amplification of 1 and shall be calculated based on the <i>Block Endurance</i> field of the NAND Parameter Page (Byte 105, 206). (i.e., no increase in the number of write operations performed by the device beyond the number of write operations requested by a host). This is a static value. This value will be based on total NAND physical capacity.						
EXT-SMART-21	Thermal Throttling Count	2	243:242	The field counts the number of thermal throttling events. This shall be set to zero on factory exit.						
EXT-SMART-22	Unaligned I/O	8	251:244	This is a count of the number of write IOs performed by the device that are not aligned to the indirection unit size (IU) of the device. Alignment indicates only the start of each IO. The length does not affect this count. This counter shall reset on a power cycle. This counter shall not wrap. This shall be set to zero on factory exit.						
EXT-SMART-23	Physical Media Units Read	16	267:252	Contains the number of bytes read from the media from both the user and system areas. It shall be represented in 512-byte units. When the LBA size is a value other than 512 bytes, the device shall convert the amount of data read to 512-byte units. This value is reported in thousands (i.e., a value of 1 corresponds to 1000 units of 512 bytes read) and is rounded up.						
EXT-SMART-24	Command Timeout	16	283:268	<p>Internal counter tracking command timeout, a command timeout is considered as an event every time a host IO command (Read, Write, Trim) exceeds the drive’s command timeout threshold. All commands within the drive’s internal queue should be accounted for, and command execution will be measured from the time the command is fetched by the device till a completion is posted.</p> <p>The supplier shall also provide documentation of this threshold.</p> <table><tr><td># of Bytes</td><td>Byte Address</td><td>Field Description</td></tr><tr><td></td><td></td><td></td></tr></table>	# of Bytes	Byte Address	Field Description			
# of Bytes	Byte Address	Field Description								

				<table><tr><td>4</td><td>271:268</td><td># of READ CMDs exceeding threshold</td></tr><tr><td>4</td><td>275:272</td><td># of WRITE CMDs exceeding threshold</td></tr><tr><td>4</td><td>279:276</td><td># of TRIMs CMDs exceeding threshold</td></tr><tr><td>4</td><td>283:280</td><td>Reserved</td></tr></table>	4	271:268	# of READ CMDs exceeding threshold	4	275:272	# of WRITE CMDs exceeding threshold	4	279:276	# of TRIMs CMDs exceeding threshold	4	283:280	Reserved
4	271:268	# of READ CMDs exceeding threshold														
4	275:272	# of WRITE CMDs exceeding threshold														
4	279:276	# of TRIMs CMDs exceeding threshold														
4	283:280	Reserved														
EXT-SMART-26	PCIe Link Retraining Count	8	291:284	This is a count of the number of PCIe Link Retraining events. This count shall only increment during run time. It shall not increment during link training or a power cycle. This shall be cleared to zero on factory exit.												
EXT-SMART-27	Power State Change Count	8	299:292	Summation counter of the number of power state changes ( <i>power state change can be either host or device initiated</i> ). This count shall only increment during run time. This shall be set to zero on factory exit.												
EXT-SMART-28	BSSD Version	8	307:300	Version of Boot SSD Specifications this device conforms to. <table><tr><th>Byte Address</th><th>Field Description</th></tr><tr><td>301:300</td><td>Major version Field. Shall be set to 0001h.</td></tr><tr><td>303:302</td><td>Minor version Field. Shall be set to 0000h.</td></tr><tr><td>305:304</td><td>Point version Field. Shall be set to 0000h.</td></tr><tr><td>307:306</td><td>Errata version Field. Shall be set to 0000h.</td></tr></table>	Byte Address	Field Description	301:300	Major version Field. Shall be set to 0001h.	303:302	Minor version Field. Shall be set to 0000h.	305:304	Point version Field. Shall be set to 0000h.	307:306	Errata version Field. Shall be set to 0000h.		
Byte Address	Field Description															
301:300	Major version Field. Shall be set to 0001h.															
303:302	Minor version Field. Shall be set to 0000h.															
305:304	Point version Field. Shall be set to 0000h.															
307:306	Errata version Field. Shall be set to 0000h.															
EXT-SMART-29	FTL Unit Size	4	311:308	Display Indirection Unit (IU) size. The value is reported in terms of a power of two (2^n). Units are in Bytes, so "12" means the FTL unit size is 4096 Bytes or 4KiB.												
EXT-SMART-30	TCG Ownership Status	4	315:312	TCG Ownership Status: 00h - ownership status could not be determined (error condition).												

				01h - C_PIN_SID == MSID. 02h - C_PIN_SID != MSID. 03h - BlockSID is enabled.
EXT-SMART-31	Reserved	178	493:316	Shall be set to 0x0.
EXT-SMART-32	Log Page Version	2	495:494	This indicates the version of the mapping this log page uses. Shall be set to 0x0001
EXT-SMART-33	Log Page GUID	16	511:496	Shall be set to 0xC46DD7920F1E4266A178D8AC78884365.

#### 5.4.6.3 Hardware Revision Log (Log Identifier C6h)

This log provides a host with information regarding the hardware revision of the device and components on the device. A value of 0x0 shall be interpreted as the field is not populated.

Requirement ID	Byte Address	Field	# of Bytes	Field Description
HWREV-1	0	Global Device HW revision	1	Global Device Hardware Revision.  The hardware revision shall be incremented every time there is a design change to the hardware.  Non-production devices shall have a value from 0x01 to 0x0F and production devices shall have a value from 0x10 to 0x20. The values 0x00 and 0x21 to 0xFF are Reserved
HWREV-2	1	ASIC Revision	1	ASIC Controller Revision



HWREV-3	2	PCB Manufacturer Code	1	PCB Manufacture Code
HWREV-4	3	DRAM Manufacturer Code	1	DRAM Manufacture Code
HWREV-5	4	NAND Manufacture Code	1	NAND Manufacture Code
HWREV-6	5	PMIC 1 Manufacture Code	1	PMIC 1 Manufacture Code
HWREV-7	6	PMIC 2 Manufacture Code	1	PMIC 2 Manufacture Code
HWREV-8	7	Other Component 1 Manufacture Code	1	Other Component 1 Manufacture Code
HWREV-9	8	Other Component 2 Manufacture Code	1	Other Component 2 Manufacture Code

HWREV-10	9	Other Component 3 Manufacture Code	1	Other Component 3 Manufacture Code
HWREV-11	10	Other Component 4 Manufacture Code	1	Other Component 4 Manufacture Code
HWREV-12	11	Other Component 5 Manufacture Code	1	Other Component 5 Manufacture Code
HWREV-13	12	Other Component 6 Manufacture Code	1	Other Component 6 Manufacture Code
HWREV-14	13	Other Component 7 Manufacture Code	1	Other Component 7 Manufacture Code
HWREV-15	14	Other Component 8 Manufacture Code	1	Other Component 8 Manufacture Code
HWREV-16	15	Other Component 9 Manufacture Code	1	Other Component 9 Manufacture Code

HWREV-17	63:16	Reserved	48	Reserved. Shall be set to 0x0.
HWREV-18	79:64	Device Manufacturing Detailed Information	16	Device Manufacturing Date and Lot Code Detailed Information
HWREV-19	95:80	ASIC Detailed Information	16	ASIC Date and Lot Code Detailed Information
HWREV-20	111:96	PCB Detailed Information	16	PCB Date and Lot Code Detailed Information
HWREV-21	127:112	DRAM Detailed Information	16	DRAM Date and Lot Code Detailed Information
HWREV-22	143:128	NAND Detailed Information	16	NAND Date and Lot Code Detailed Information

HWREV-23	159:144	PMIC 1 Detailed Information	16	PMIC 1 Date and Lot Code Detailed Information
HWREV-24	175:160	PMIC 2 Detailed Information	16	PMIC 2 Date and Lot Code Detailed Information
HWREV-25	191:176	Other Component 1 Detailed Information	16	Other Component 1 Date and Lot Code Detailed Information
HWREV-26	207:192	Other Component 2 Detailed Information	16	Other Component 2 Date and Lot Code Detailed Information
HWREV-27	223:208	Other Component 3 Detailed Information	16	Other Component 3 Date and Lot Code Detailed Information
HWREV-28	239:224	Other Component 4 Detailed Information	16	Other Component 4 Date and Lot Code Detailed Information
HWREV-29	255:240	Other Component 5	16	Other Component 5 Date and Lot Code Detailed Information

		Detailed Information		
HWREV-30	271:256	Other Component 6 Detailed Information	16	Other Component 6 Date and Lot Code Detailed Information
HWREV-31	287:272	Other Component 7 Detailed Information	16	Other Component 7 Date and Lot Code Detailed Information
HWREV-32	303:288	Other Component 8 Detailed Information	16	Other Component 8 Date and Lot Code Detailed Information
HWREV-33	319:304	Other Component 9 Detailed Information	16	Other Component 9 Date and Lot Code Detailed Information
HWREV-34	351:320	Serial Number	32	Device Serial Number
HWREV-35	493:352	Reserved	142	Reserved. Set to 0x0.
HWREV-36	494:495	Log Page Version	2	This indicates the version of the mapping this log page uses. Shall be set to 0001h.

HWREV-37	511:496	Log Page GUID	16	Shall be set to  aab005f5-135e-4815 -ab89-05ba8be2bf3c
----------	---------	---------------	----	---

### 5.4.7 Set/Get Features Requirements

#### 5.4.7.1 General Get Feature Requirements

Requirement ID	Description
GETF-1	For any Get Feature Identifier defined in this section, Selection (SEL) values 00b to 11b in DWORD 10 shall be supported.
GETF-2	If the feature requested by Set Feature is not supported, then a status error code of 02h (Invalid Field in Command) shall be returned.

#### 5.4.7.2 Volatile Write Cache Settings

Requirement ID	Description
VWC-1	Volatile Write Cache (Feature Identifier 06) shall be supported even if the device does not have a Volatile Write Cache.

#### 5.4.7.3 Power Management

Requirement ID	Description
PM-1	Power Management (Feature Identifier 02h) and the Power State descriptor table shall be supported.

#### 5.4.7.4 Host Controlled Thermal Management

Requirement ID	Description
HCTM-1	Host Controlled Thermal Management (Feature Identifier 10h) shall be supported.

#### 5.4.7.5 Clear PCIe Correctable Error Counters (Feature Identifier C3h) Set Feature

Requirement ID	Dword	Field	Bits	Field Description
CPCIE-1	0	Command Identifier (CID)	31:16	Shall be set as defined in the NVMe Specification version specified in NVMe-1.
CPCIE -2	0	PRP or SGL for Data Transfer (PSDT)	15:14	Shall be cleared to 00b.
CPCIE -3	0	Reserved	13:10	Shall be cleared to zero
CPCIE -4	0	Fused Operation (FUSE)	9:8	Shall be cleared to 00b.

CPCIE-5	0	Opcode (OPC)	7:0	Shall be set to 09h.
CPCIE-6	1	Namespace Identifier (NSID)	31:0	Shall be cleared to zero.
CPCIE-7	2:3	Reserved	31:0	Shall be cleared to zero.
CPCIE-8	4:5	Metadata Pointer (MPTR)	31:0	Shall be cleared to zero.
CPCIE-9	6:9	Data Pointer (DPTR)	31:0	Shall be cleared to zero.
CPCIE-10	10	Save (SV)	31	The device shall not support setting this bit to 1b. If the controller receives this Set Features command with the bit set to 1b, then the device shall abort the command with a status of Feature Identifier Not Saveable.
CPCIE-11	10	Reserved	30:8	Shall be cleared to zero.
CPCIE-12	10	Feature Identifier (FID)	7:0	Shall be set to C3h.
CPCIE-13	11	Clear PCIe Error Counters	31	Set to 1b to clear all PCIe correctable error counters in the SMART / Health Information Extended (Log Identifier C0h).  The NVMe CLI plug-in command “clear-pcie-correctable-errors” can also perform this operation.
CPCIE-14	11	Reserved	30:0	Shall be cleared to zero.
CPCIE-15	12:13	Reserved	31:0	Shall be cleared to zero.
CPCIE-16	14	UUID Index	31:0	Shall be set per <a href="#">UUID-3</a>
CPCIE-17	15	Reserved	31:0	Shall be cleared to zero

## 5.5 NVMe I/O Command Set

The device shall support the following mandatory and optional NVMe IO commands:

Requirement ID	Description
NVMe-IO-1	The device shall support all mandatory NVMe I/O commands.
NVMe-IO-2	The device shall support the Dataset Management command. The device shall support the Attribute – Deallocate (AD) bit.



NVME-IO-3	Product Documentation must contain the maximum time taken to complete a full drive deallocate.
NVMe-IO-4	If the device supports Write Uncorrectable command, then uncorrectable errors (e.g., read errors) that are a consequence of a prior Write Uncorrectable command shall not be counted in the Smart / Health Information (Log Identifier 02h or Log Identifier C0h) Media and Data Integrity Errors field.

### 5.5.1 De-Allocation Requirements

Requirement ID	Description
TRIM-1	The device shall support Deallocate/TRIM.
TRIM-2	The Identify Namespace - Deallocate Logical Block Features (DLFEAT) field shall be supported.
TRIM-3	If data has been de-allocated and not written to when an unsafe power down event happens, the data shall be 0, 1 or unchanged when read, according to the value of the DLFEAT field.
TRIM-4	De-allocated addresses shall provide the performance and reliability benefits of over-provisioned space.
TRIM-5	The device shall support Garbage Collection during periods of no IO.
TRIM-6	Read latency shall not change more than 5% from baseline when the host is issuing De-Allocate/TRIM commands.
TRIM-7	Read latency shall not change more than 5% from baseline when the device is performing Idle garbage collection.

## 5.6 Optional NVMe Features

The device shall also support the following optional NVMe features:

Requirement ID	Description
NVMe-OPT-1	Timestamp SET-FEATURE (FID 0xE) shall be supported to align the device's internal logs.
NVMe-OPT-2	The device shall never set the Synch field bit to 1b in the Timestamp (Feature Identifier 0Eh).
NVMe-OPT-3	The device shall only clear the Timestamp Origin field to 000b in the Timestamp (Feature Identifier 0Eh) on a main power cycle (cold boot).
NVMe-OPT-4	NVMe APST shall be disabled by default and shall not be enabled by the host.

## 6 PCIe Requirements

### 6.1 Overview

Requirement ID	Description
PCIe-1	The device shall be compliant to PCIe base specification 3.1a (or later).
PCIe-2	The device shall support common clock with or without spread spectrum.

### 6.2 Compliance

Requirement ID	Description
PCIe-CONF-1	The SSD supplier shall provide a copy of the PCI-SIG <a href="#">compliance</a> test report.

### 6.3 Lane Width & Link Speed

Requirement ID	Description
LWLS-1	The device shall support a x4 PCIe lane width.
LWLS-2	The device shall support a minimum of PCIe Gen3 as factory default.
LWLS-3	The device shall train to x1 when only one upstream lane is available, to x2 when the upstream device provides only 2 lanes per device and to x4 when 4 lanes are present.

### 6.4 Maximum Payload Size

Requirement ID	Description
MPS-1	The device shall support a PCIe Maximum Payload Size (MPS) of 256 bytes or larger.

### 6.5 Lane reversal

Requirement ID	Description
LR-1	The device shall support lane reversal with all lanes connected or partially connected lanes (e.g., a x4 device shall support it for x4, x2, and x1 connections).

### 6.6 Boot Requirements

Requirement ID	Description
BOOT-1	The device shall not support a PCI Expansion/ Option ROM.
BOOT-2	The device shall support booting using UEFI, and Coreboot (see <a href="https://www.linuxboot.org/">https://www.linuxboot.org/</a> ).

## 6.7 Reset Support

Requirement ID	Description
PCIeRST-1	All three PCIe Conventional Resets (Cold, Warm, Hot) shall be supported.
PCIeRST-2	PCIe Function Level Reset shall be supported.

## 6.8 PCIe Error Logging

The following table indicates the implementation details of where the PCIe physical layer error counters shall be logged. This is in addition to the aggregated PCIe error counters defined in SMART / Health Information Extended (Log Identifier C0h).

Requirement ID	Event	Logging Mechanism
PCIERR-1	Unsupported Request Error Status (URES)	Uncorrectable Error Status Register Offset 04h in PCIe Base Specification
PCIERR-2	ECRC Error Status (ECRCES)	
PCIERR-3	Malformed TLP Status (MTS)	
PCIERR-4	Receiver Overflow Status (ROS)	
PCIERR-5	Unexpected Completion Status (UCS)	
PCIERR-6	Completer Abort Status (CAS)	
PCIERR-7	Completion Timeout Status (CTS)	
PCIERR-8	Flow Control Protocol Error Status (FCPES)	
PCIERR-9	Poisoned TLP Status (PTS)	
PCIERR-10	Data Link Protocol Error Status (DLPES)	
PCIERR-11	Advisory Non-Fatal Error Status (ANFES)	Correctable Error Status Register Offset 10h in PCIe Base Specification
PCIERR-12	Replay Timer Timeout Status (RTS)	Correctable PCIe Error Count in the SMART / Health Information Extended (Log Identifier C0h).
PCIERR-13	REPLAY_NUM Rollover Status (RRS)	
PCIERR-14	Bad DLLP Status (BDS)	
PCIERR-15	Bad TLP Status (BTS)	
PCIERR-16	Receiver Error Status (RES)	

## 6.9 Low Power Modes

Requirement ID	Description
LPWR-1	If Active State Power Management (ASPM) is supported, the device default state shall be disabled.

LPWR-2	The device shall support LTR (Latency Tolerance Reporting).
--------	---

## 7 Reliability

### 7.1 UBER

Requirement ID	Description
UBER-1	The device shall support an Uncorrectable Bit Error Rate (UBER) of < 1 sector per 10 <sup>15</sup> bits read.

### 7.2 SSD Operational Life

Requirement ID	Description
SSDOP-1	The warranty and design shall support a 5 year operational life.

### 7.3 AFR (Annual Failure Rate)

Requirement ID	Description													
AFR-1	The device shall meet an MTBF of 2.0 million hours (AFR of <= 0.44% per JEDEC JESD 218) under the following environmental conditions:													
	<table><tr><th>Specification</th><th>Environment</th><th>Requirement</th></tr><tr><td rowspan="2">Temperature</td><td>Operational</td><td><ul style="list-style-type: none"><li>● 0°C to 50°C (32°F to 112°F)</li><li>● &lt; 20°C (68°F) per hour gradient</li></ul></td></tr><tr><td>Non-Operational</td><td><ul style="list-style-type: none"><li>● -40°C to 70°C (-40°F to 158°F)</li><li>● &lt; 30°C (86°F) per hour gradient</li></ul></td></tr><tr><td rowspan="2">Humidity</td><td>Operational</td><td><ul style="list-style-type: none"><li>● 10% to 90% non-condensing</li><li>● Yearly weighted average: &lt; 80% RH<ul style="list-style-type: none"><li>○ 90% of year: &lt; 80%</li><li>○ 10% of year: 80% to 90%</li></ul></li><li>● Maximum dewpoint: 29.4°C (85°F)</li></ul></td></tr><tr><td>Non-Operational</td><td><ul style="list-style-type: none"><li>● 5% to 95% non-condensing</li><li>● 38°C (100.4°F) maximum wet bulb temperature</li></ul></td></tr></table>	Specification	Environment	Requirement	Temperature	Operational	<ul style="list-style-type: none"><li>● 0°C to 50°C (32°F to 112°F)</li><li>● &lt; 20°C (68°F) per hour gradient</li></ul>	Non-Operational	<ul style="list-style-type: none"><li>● -40°C to 70°C (-40°F to 158°F)</li><li>● &lt; 30°C (86°F) per hour gradient</li></ul>	Humidity	Operational	<ul style="list-style-type: none"><li>● 10% to 90% non-condensing</li><li>● Yearly weighted average: &lt; 80% RH<ul style="list-style-type: none"><li>○ 90% of year: &lt; 80%</li><li>○ 10% of year: 80% to 90%</li></ul></li><li>● Maximum dewpoint: 29.4°C (85°F)</li></ul>	Non-Operational	<ul style="list-style-type: none"><li>● 5% to 95% non-condensing</li><li>● 38°C (100.4°F) maximum wet bulb temperature</li></ul>
	Specification	Environment	Requirement											
	Temperature	Operational	<ul style="list-style-type: none"><li>● 0°C to 50°C (32°F to 112°F)</li><li>● &lt; 20°C (68°F) per hour gradient</li></ul>											
		Non-Operational	<ul style="list-style-type: none"><li>● -40°C to 70°C (-40°F to 158°F)</li><li>● &lt; 30°C (86°F) per hour gradient</li></ul>											
	Humidity	Operational	<ul style="list-style-type: none"><li>● 10% to 90% non-condensing</li><li>● Yearly weighted average: &lt; 80% RH<ul style="list-style-type: none"><li>○ 90% of year: &lt; 80%</li><li>○ 10% of year: 80% to 90%</li></ul></li><li>● Maximum dewpoint: 29.4°C (85°F)</li></ul>											
Non-Operational		<ul style="list-style-type: none"><li>● 5% to 95% non-condensing</li><li>● 38°C (100.4°F) maximum wet bulb temperature</li></ul>												
AFR-2	The supplier shall provide AFR de-rating curves for the Temperature range shown in requirement AFR-1 for up to 70°C (158°F).													
AFR-3	The supplier shall provide MTBF derating curve for combined Temperature, shown under <a href="#">Operational Conditions</a> Section of this spec.													
AFR-4	The supplier must provide the temperature and humidity conditions used to determine the MTBF mentioned in AFR-1.													

## 7.4 End to End Data Protection

Requirement ID	Description
E2E-1	All user data and metadata shall be protected using overlapping protection mechanisms throughout the entire read and write path in the device including all storage elements (registers, caches, SRAM, DRAM, NAND, etc.).
E2E-2	At least one bit of correction and 2 bits of detection is required for all memories. This shall be for all memories regardless of function.
E2E-3	The entire DRAM addressable space needs to be protected with at least one-bit correction and 2 bits of detection scheme. This includes but not limited to the following: <ul style="list-style-type: none"><li>• Flash translation layer (FTL)</li><li>• Mapping tables</li><li>• Journal entries</li><li>• Firmware scratch pad</li><li>• System variables</li><li>• Firmware code</li></ul>
E2E-4	The device shall include a mechanism to protect against returning the data from the wrong logical block address (LBA), including previous copies from the same LBA, to the host. Device shall perform LBA integrity checking on all transfers to and from the media.
E2E-5	All device metadata, firmware, firmware variables, and other device system data shall be protected by at least a single bit detection scheme.
E2E-6	Any memory buffers that are utilized to accelerate data transfer (read-ahead buffers for example) shall follow the protection scheme outlined in E2E-4.

## 7.5 Background Data Refresh (BDR)

Requirement ID	Description
BK-DR-1	The device shall support background data refresh while the device is powered on to ensure there is no data-loss due to power-on retention issues.
BK-DR-2	The device shall be designed and tested to support the normal NAND operating temperature. For example, if the SSD is cooled to a composite temperature of 70C, this may imply the actual NAND temperature is 80C. This shall be taken into account when implementing this feature.
BK-DR-3	Background data refresh shall cover the entire device and be designed to continuously run in the background and not just during idle periods.

## 7.6 Background Data Flush (BDF)

Requirement ID	Description
BK-FL-1	The device shall flush data to NVM ( <i>including SMART data</i> ) in internal cache (DRAM, SRAM) at least every 10 mins to minimize data-loss in case of an unexpected power loss scenario.

## 7.7 Data Integrity

Requirement ID	Description
DI-1	The SSD supplier shall provide clear documentation on the worst case data loss conditions to the customers.
DI-2	In case of a successful Normal shutdown operation (CC.SHN = 1 set by the NVMe device driver), all data must be committed to persistent storage.

## 7.8 Command & Completion Timeout

All commands sent to the SSD shall adhere to the following command timeouts.

Requirement ID	Description
CTO-1	NVMe ADMIN Commands shall take no more than 10 seconds from submission to completion.
CTO-2	The only exceptions to CTO-1 shall be Format, Self-Test and Sanitize and the TCG commands Revert, Revert SP and Change Key.
CTO-3	NVMe I/O Commands shall take no more than 8 seconds from submission to completion.
CTO-4	The vendor shall support and disclose the vendor-specific timeout ranges for ranges A, B, C, D supported for the Completion Timeout Ranges Supported as defined in the PCIe Base Specification.
CTO-5	Disabling of PCIe Completion Timeout shall also be supported by the device.
CTO-6	Device supplier shall disclose any I/O scenario that could violate the timeout requirements in CTO-1 through CTO-4.

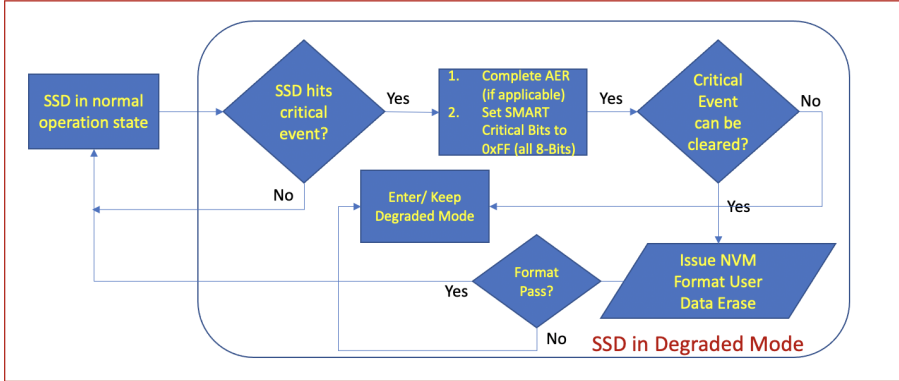
## 7.9 EOL Requirements

This section describes the behavior of the SSD & requirements pertaining to device end of life.

Requirement ID	Description
EOL-1	<p>Various types of samples are required for EOL (End of Life) testing:</p> <ol style="list-style-type: none"><li>1. Beginning of Life (Short-stroked if required by EOL-2).</li><li>2. End of Life (Short-stroked if required by EOL-2).</li><li>3. End of Life (Not short stroked if different than #2).</li></ol> <p>EOL is defined as the Total Bytes Written (TBW) specification has been surpassed using a write workload with a WAF=1. or the non-volatile media endurance limit (e.g., NAND cycling limit) has been reached; whichever is earliest.</p>
EOL-2	The device shall not continue to operate in a read/write mode if data integrity is lost.
EOL-3	Upon reaching 100% of specified device endurance, the device shall notify the host with an AEN.
EOL-4	The device shall switch to Read Only Mode (ROM) when the available spares field in the SMART / Health Information (Log Identifier 02h) reaches 0%. A value of 0%

	<p>represents the device state where there is an insufficient number of spare blocks to support Host writes. After the drive switches to read-only mode, bit 2 and bit 3 of the Critical Warning field of Section 5.14.1.2 SMART Attributes in the NVMe specification version 1.4c shall be set to 1b.</p> <p>The device shall set bit 0 of the Critical Warning field to 1b in the SMART Attributes specified in Section 5.14.1.2 of the NVMe specification version 1.4c and generate a Critical Warning async notification (AEN) when the available spares value falls below the available spare threshold.</p>
--	---

## 7.10 Degraded Mode Behavior

Requirement ID	Description
DGDST-1	All conditions in which the drive enters a Degraded Mode shall be disclosed by the SSD vendor. For example, Read-Only state, Assert, CFS, etc.
DGDST-2	Format NVM command shall be supported when the device is in a Degraded state.
DGDST-3	Device shall allow READ NVMe command when it is in a Degraded state.
DGDST-4	All debug logs, including telemetry & SMART shall be accessible while the device is in a Degraded state.
DGDST-5	All relevant SMART Critical Warning bits as defined by the NVMe Spec. shall be supported.
DGDST-6	The device shall set all bits of the SMART Critical Warning (0xFF) structure to indicate a Firmware Assert degraded state.
DGDST-7	<p>If degraded mode can be cleared, the method below shall be followed to clear and exit degraded mode:</p>  <pre> graph TD     Start[SSD in normal operation state] --&gt; Event{SSD hits critical event?}     Event -- No --&gt; Degraded[Enter/ Keep Degraded Mode]     Event -- Yes --&gt; AER[1. Complete AER (if applicable) 2. Set SMART Critical Bits to 0xFF (all 8-Bits)]     AER --&gt; Cleared{Critical Event can be cleared?}     Cleared -- No --&gt; Degraded     Cleared -- Yes --&gt; Format{Format Pass?}     Format -- No --&gt; Degraded     Format -- Yes --&gt; Erase[/Issue NVM Format User Data Erase/]     Erase --&gt; Degraded     </pre> <p>Note: The responsibility of deciding which Critical Event can be cleared by a FORMAT is made by the Device Firmware whereas the FORMAT issuance is the responsibility of the host.</p>

## 8 Endurance

### 8.1 Endurance data

Requirement ID	Description
ENDU-1	<p>The device documentation shall include the number of physical bytes able to be written to the device assuming a write amplification of 1 (WAF=1). The units should be gigabytes (10<sup>9</sup> bytes). This will be used in the formula for pDWPd:</p> $\text{Physical Drive Writes per Day (pDWPd)} = \frac{\text{Physical Bytes Written @ WAF = 1}}{(5 \text{ years} \times 365 \text{ days} \times \text{usable capacity})}$
ENDU-2	The device vendor shall document the methodology in which ENDU-1 was validated.
ENDU-3	The Percentage Used in the SMART / Health Information (Log Identifier 02h) shall track linearly with bytes written and at 100% it shall match the EOL value specified in ENDU-1 ( <i>number of physical bytes able to be written</i> ).

### 8.2 Retention conditions

Requirement ID	Description						
RETEN-1	<p>The SSD data retention values are as defined in the table below:</p> <table><tr><th>Parameter</th><th>Boot Device Requirement</th></tr><tr><td>Non-operational (Powered-off) data retention (end of life) @ 40°C (104°F).</td><td>At least 1 month</td></tr><tr><td>Powered-on data retention @ conditions specified in <a href="#">Section 14.1</a></td><td>At least 5 years</td></tr></table>	Parameter	Boot Device Requirement	Non-operational (Powered-off) data retention (end of life) @ 40°C (104°F).	At least 1 month	Powered-on data retention @ conditions specified in <a href="#">Section 14.1</a>	At least 5 years
Parameter	Boot Device Requirement						
Non-operational (Powered-off) data retention (end of life) @ 40°C (104°F).	At least 1 month						
Powered-on data retention @ conditions specified in <a href="#">Section 14.1</a>	At least 5 years						
RETEN-2	The device shall not throttle its performance based on the endurance metric (endurance throttling).						
RETEN-3	Operating (Powered-on) data retention, For purposes of this requirement, the assumption is that the Terabytes Written (TBW) capability of the devices is used linearly over the lifetime. This requirement does not imply any specific warranty period.						

### 8.3 Shelf Life

Requirement ID	Description
SHELF-1	A new device may be kept as a datacenter spare at the beginning of life (BOL) and therefore shall be fully functional even if it sits on the shelf for at least 1



	year at 40°C (104°F) before getting installed on the server. The device shall be new in box factory state.
SHELF-2	The SHELF-1 requirement is not limited to new-from-factory drives. Used drives with at least 25 percent lifetime remaining must also be usable after sitting on a shelf for at least 1 year. Such drives must not be completely reverted to factory default state (see LP-3) but are not expected to retain data (see RETEN-1). It is acceptable to require that a FormatNVM of the entire drive be completed successfully before the drive can return to normal operation.
SHELF-3	Supplier shall provide Shelf-life de-rating curve for combined Temperature range, shown under section 14.1.3 Non-Operational environmental conditions.
SHELF-4	Shelf-life shall be documented and provided to the customer.

## 8.4 Endurance Targets

Requirement ID	Usable User Capacity	Minimum TBW @ WAF = 1
ENDU-TGT-1 (No compression, 0% OP)	128GB	Minimum endurance of 768 TBW
ENDU-TGT-2 (No compression, 0% OP)	256GB	Minimum endurance of 1536 TBW
ENDU-TGT-3 (No compression, 0% OP)	512GB	Minimum endurance of 3072 TBW
ENDU-TGT-4 (No compression, 0% OP)	1024GB	Minimum endurance of 6144 TBW
ENDU-TGT-5 (No compression, 0% OP)	2048GB	Minimum endurance of 12288 TBW

## 8.5 Wear Leveling

Requirement ID	Description
WL-1	The device shall utilize the entire physical media capacity range whenever the device needs to wear-level a block (system or user data). The device shall not restrict the wear-leveling range to a subset of the entire physical media capacity unless otherwise specified.

## 9 Performance

### 9.1.1 Boot SSD Performance Requirements

The SSD vendor will supply performance measurement numbers for the following workloads. While for Client SSDs, it's a common practice to test a short-range and report Fresh-Out-Of-Box numbers in the SSD Datasheet, however, hyper-scaler use cases require the testing to be carried out when the drive is in Steady State (*sustained*). The targets that the drive must meet (*while in Steady State*) are as follows:

### 9.2 Boot Drive Sustained Bandwidth and IOPS Targets

**Sustained Performance Targets**

Requirement ID	Workload	Target with 0% OP	Target with 20% OP	Target with 50% OP
SS-PERF-1	100% random 4k read (IOPS)	50K	50k	50K
SS-PERF-2	100% random 4k write (IOPS)	4K	8K	12K
SS-PERF-3	70% random read 4K, 30% random write 4K	15K Rd, 4k Wr (Total 19K)	15K Rd, 8k Wr (Total 23K)	15K Rd, 10k Wr (Total 25K)
SS-PERF-4	100% sequential read 128K (4K align)	250MB/s	250MB/s	250MB/s
SS-PERF-5	100% sequential write 128K (4K align)	200MB/s	200MB/s	200MB/s

### 9.3 Boot Drive Performance Measurement Process (PMP)

Requirement ID	Description
PMP-1	The SSD vendor shall provide sustained mode performance values by using a procedure defined in <a href="#">Figure 4</a> .
PMP-2	The SSD vendor shall disclose the approx. time and procedure used to enter sustained mode. A preferred method using FIO is described below.
PMP-3	The SSD vendor shall measure & report performance numbers with "Internal Cache ON" & "Internal Cache OFF" scenario
PMP-4	The SSD vendor shall measure performance using the full span of LBA range
PMP-5	The SSD vendor shall perform all measurements using the latest kernel for Linux CentOS distribution 8 (5.10 or newer), preferably on a <i>customer recommended OCP Compute Platform</i> .
PMP-6	The SSD vendor shall use the following settings for measurement: <ul style="list-style-type: none"><li>• Queue Depth: Vendor Defined with a minimum of 32 (<i>for IOPS &amp; Throughput Tests</i>), 1 (<i>for Latency Tests</i>)</li><li>• # of Threads: 4 (for Random), 1 (for Sequential)</li><li>• Write/ Internal Cache: ON &amp; OFF</li></ul>
PMP-7	The following metrics are needed for the workloads listed below: <ul style="list-style-type: none"><li>- Throughput</li><li>- IOPS</li></ul>

	<ul style="list-style-type: none"> <li>- Latency (P99, P99.99 and P100 (<i>Max Latency</i>): These are the 99th, 99.99th and 100th (Max) percentile latency respectively.</li> </ul> <p>Workloads:</p> <ul style="list-style-type: none"> <li>- 128K Sequential Writes (4K Aligned)</li> <li>- 128K Sequential Reads (4K Aligned)</li> <li>- 4K, 8K Random Writes (4K Aligned)</li> <li>- 4K, 8K Random Reads (4K Aligned)</li> <li>- 4K, 8K Random Read/Write (70/30) (4K Aligned)</li> <li>- Latency 4K Random Read (QDepth = 1 and 32) <ul style="list-style-type: none"> <li>o P99 Latency</li> </ul> </li> <li>- fb-FioSynthFlash workloads defined in 4.3</li> </ul>
PMP-8	<p>The following kernel &amp; device parameters shall be used:</p> <ul style="list-style-type: none"> <li>- Schedulers: Kyber and None</li> <li>- Device Sector Size: 512 Bytes</li> </ul>
PMP-9	<p>For file-system based testing (e.g. IO.go), please use BTRFS &amp; EXT4 file-system types with the following parameters:</p> <p><b>Example for BTRFS:</b></p> <pre>cd / mkdir data mkdir /data/trimCheck mkfs.btrfs -l 16k -m single /dev/nvmeXn1 /bin/mount -o rw,relatime, discard=async,space_cache,compress-force= zstd:3,ssd,nobarrier,noatime,nodiratime /dev/nvmeXn1 /data/trimCheck</pre> <p><b>Settings for EXT-4:</b></p> <p>Options: (EXT4) type ext4 (rw,relatime,data=ordered)</p>

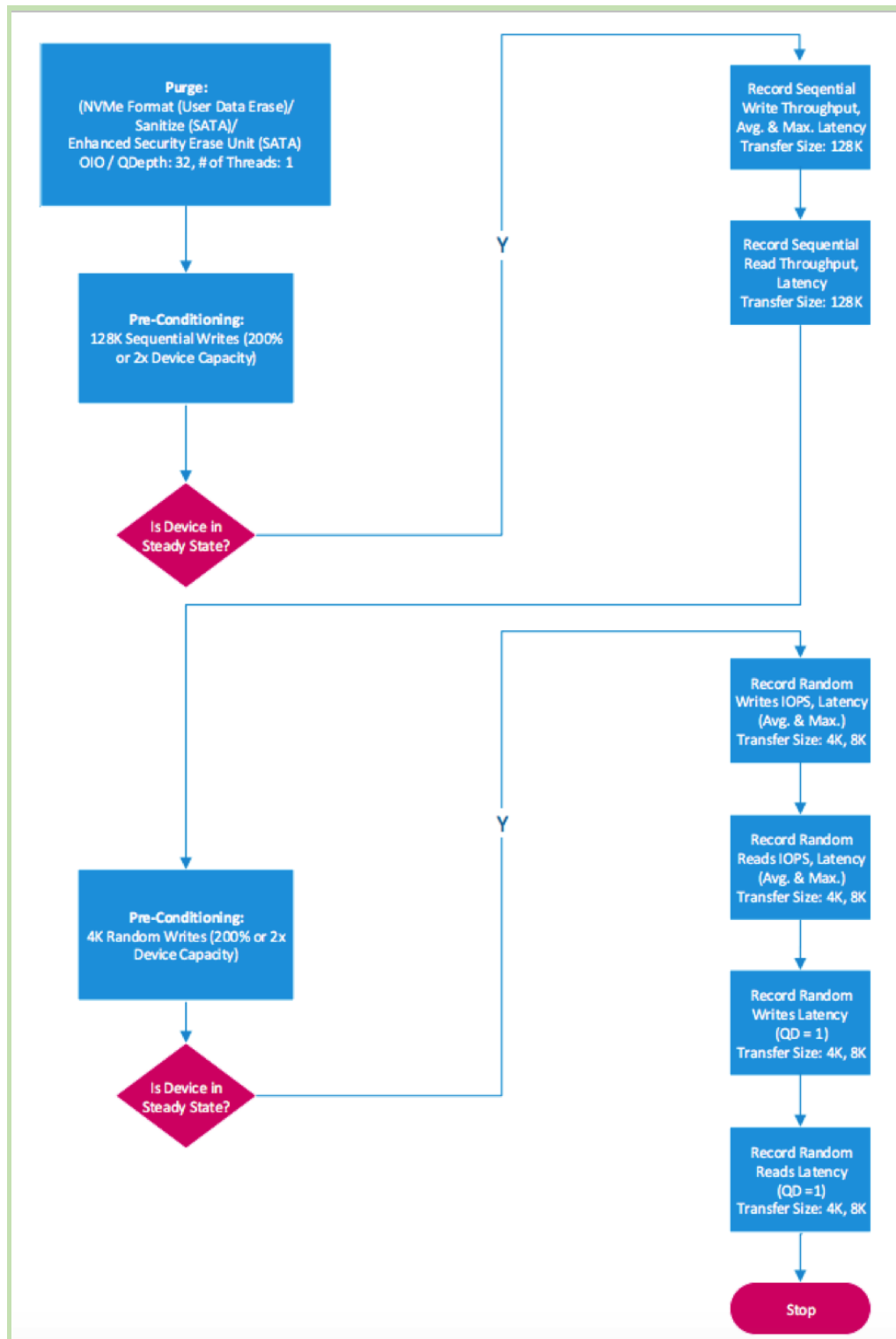


Figure 4 Boot SSD Performance measurement procedure

#### 9.4 Boot-Drive Latency (QoS) Targets:

The following workloads are part of the fb-FioSynthFlash synthetic benchmark package, which can be obtained from [GitHub](#). The supplier shall run the workloads and ensure that the SSD meets/ exceeds the targets set below under drive's normal operating conditions. All targets assume an over-provisioning of 0% (for 256GB) and OP settings per NSM-2 (for 512GB).

Requirement ID	Workloads	Type	IOPS	B/W (MB/s)	P99 (ms)	P99.99 (ms)	Max Latency (ms)
QOS-1	<a href="#">UDB_Boot</a>	Read	N/A	N/A	5	20	500
QOS-2		Write	N/A	N/A	100	200	400
QOS-3	<a href="#">Warmstorage_HXFS_SSD</a>	Read	N/A	N/A	5	20	500
QOS-4		Write	N/A	N/A	40	100	400
QOS-5	<a href="#">Twshared_Pkg_Boot</a>	Read	210	51	60	90	100
QOS-6		Write	600	62	30	60	250
QOS-7	<a href="#">Rsw_Cp_wTRIM</a>	Read	119	27	15	30	80
QOS-8		Write	700	80	30	60	120
QOS-9		Trim	N/A	60	N/A	N/A	N/A
QOS-10	<a href="#">Twi_Iris</a>	Read	420	45	10	50	150
QOS-11		Write	65	6	40	200	400
QOS-12		Trim	N/A	2.4	N/A	N/A	N/A
QOS-13	<a href="#">iDyno_Boot</a>	Read	N/A	75	25	50	90
QOS-14		Write	N/A	15	30	45	60
QOS-15		Trim	N/A	0.1	N/A	N/A	N/A
QOS-16	<a href="#">Stacking</a>	Read	N/A	120	10	50	120
QOS-17		Write	N/A	64	0.6	60	350

QOS-18		Trim	N/A	18	N/A	N/A	N/A
--------	--	------	-----	----	-----	-----	-----

## 9.5 TRIM Performance

TRIM performance is important in several Boot Drive use cases and it's evaluated based on a synthetic benchmark part of fioSynthFlash called *TrimRate*. The targets for these are defined as follows:

### 9.5.1 TRIM Rate Targets

This test measures raw trim performance with no background I/O.

Requirement ID	Description
TRIM-1	64M trim $\geq 50\text{GiB/s}$ & $\leq 10\text{ms}$ P99 trim latency
TRIM-2	3GB trim $\geq 500\text{GiB/s}$ & $\leq 10\text{ms}$ P99 trim latency
TRIM-3	4K trim $\geq 10\text{K trims/sec}$ & $\leq 19\text{ms}$ P99 trim latency

## 9.6 BootBench

This test measures load on a Boot Device in a stacked environment.

1. Download the tool from: [github.com/liu-song-6/bootbench](https://github.com/liu-song-6/bootbench)
2. Execute `./run.py <block-dev>`. FIO 3.20 or *newer* must be installed as a *prerequisite*.
3. Look at the `final_result.txt` file for output at the end of the run!

Requirement ID	Read IOPS (sustained)	Overall Result
BOOT-BENCH-1	At least 60K	Pass, without any errors.

## 9.7 Max Latency

The following benchmarks shall be used to determine the Max Latency targets for a device.

Requirement ID	Workloads	Type	IOPS	B/W (MB/s)	P99 (ms)	P99.99 (ms)	Max Latency (ms)
MAX-LAT-1	<a href="#">Twshared_Pkg_Boot_FullSweep</a>	Read	N/A	N/A	N/A	N/A	100
MAX-LAT-2		Write	N/A	N/A	N/A	N/A	250
MAX-LAT-3	<a href="#">iDyno_Boot_FullSweep</a>	Read	N/A	N/A	N/A	N/A	60
MAX-LAT-4		Write	N/A	N/A	N/A	N/A	90

## 9.8 Resctl-Bench

This test measures how many jobs can be stacked (and the effect of memory offloading) on a Root Disk. To execute the package:

1. Download the tool/ source from [GitHub](#)
2. Execute

Requirement ID	Overall Result
RES-CTL-BENCH-1	<p><a href="#">Execute</a> the benchmark and provide the resulting IO.cost model.</p> <ol style="list-style-type: none"> <li>1. Install resctl-demo and resctl-bench from CentOS repository</li> <li>2. Install any dependency that's required by resctl-bench. Some common libraries include: <ol style="list-style-type: none"> <li>a. <code>cargo coreutils util-linux python3 python3-bcc fio stress oomd adobe-source-code-pro-fonts ImageMagick ghostscript</code></li> <li>b. <code>gnuplot gcc binutils make bison flex pkgconf openssl-devel elfutils-devel</code></li> </ol> </li> <li>3. Verify cGroups2 is setup in the system : <ol style="list-style-type: none"> <li>a. <code>mount -l   grep cgroup</code> and observe: <code>cgroup2</code> on <code>/sys/fs/cgroup</code> type <code>cgroup2</code> (<code>rw,nosuid,nodev,noexec,relatime,seclabel,nsdelegate</code>)</li> </ol> </li> <li>4. If the DUT is connected as a secondary drive (non-root partition), then ensure that the Root Drive is set-up with a BTRFS file-system.</li> <li>5. Setup partition and mount file-system on DUT: <ol style="list-style-type: none"> <li>a. Partition: <code>mkfs.btrfs -l 16k -m single /dev/nvmeXn1</code></li> <li>b. File-System: Assuming <code>/drive1</code> folder exists, run: <code>mount -o rw,relatime,discard=async,space_cache,compress-force=zstd:3,ssd,nobarrier,noatime,nodiratime /dev/nvmeXn1 /drive1/</code></li> </ol> </li> <li>6. Create SWAP file on DUT: <pre>swapoff -a touch swapfile sudo chmod 600 swapfile truncate -s 0 swapfile chattr +C swapfile btrfs property set swapfile compression none sudo fallocate -l &lt;1/3<sup>rd</sup> of Total System Mem&gt; /swapfile sudo mkswap swapfile sudo swapon swapfile sudo swapon --show</pre> </li> <li>7. Command Line to Launch: <ol style="list-style-type: none"> <li>a. <code>resctl-bench -r iocost-tune-log.json run iocost-tune</code></li> </ol> </li> <li>8. Repeat Step#7 3x times</li> <li>9. Command Line to generate PDF Report that outputs the IO.cost model for the DUT. This model can be updated within the kernel under <code>cat</code></li> </ol>

	<p><code>/sys/fs/cgroup/io.cost.model</code> for any system running cgroups v2.</p>
--	---

	<pre>resctl-bench -r iocost-tune-log.json format iocost-tune:pdf=out.pdf</pre>
--	--

	<p>10. Provide the out.pdf report to the customer.</p>
--	--



## 10 Security

### 10.1 Basic Security Requirements

Requirement ID	Description
SEC-1	The device shall support signed firmware binary update which is checked before firmware is activated. The device firmware shall be authenticated using cryptographic keys on every reboot and during firmware update.
SEC-2	The device shall support XTS-AES-256 ( <i>As specified in NIST SP 800-38E, which references IEEE STD 1619</i> ) or AES-256-GCM ( <i>As specified in NIST SP 800-38D</i> ) hardware-based data encryption. AES-256-GCM is the preferred mode. The following AES variants are also acceptable: <ul style="list-style-type: none"> <li>• AES-256-EAX</li> <li>• AES-256-CTR+HMAC-SHA256 (encrypt-then-MAC)</li> <li>• AES-256-HEH</li> <li>• AES-256-CBC+S</li> </ul>
SEC-3	The device shall have anti-rollback protection for firmware. The anti-rollback protection shall be implemented with a security version which is different than the firmware version. If the security version of the firmware being activated is greater or equal to the current security version the firmware may be activated. If the security version of the firmware being activated is not equal or greater than the firmware being activated the firmware update shall fail.
SEC-4	The device shall support Crypto Erase.
SEC-5	The device shall support Secure Boot.
SEC-6	The device shall have a method of identifying a secure boot failure which does not require physical access to the device.
SEC-7	The device's cryptographic module shall be FIPS 140-3 capable per CMVP (not required to get FIPS certificate, except as specified below in a consumer-specific appendix) and shall follow the NIST 800-90 (A, B and C) specification.
SEC-8	The device shall implement only FIPS (not required to get FIPS certificate) and NIST approved implementations and algorithms.
SEC-9	The device shall support Key revocation allowing a new key to be used for firmware validation on boot. Preferred implementation is to allow for up to 8 key revocations.
SEC-10	The device shall support all TCG Storage Security Subsystem Class: Opal Specification Version 2.01 Revision 1.0 or newer mandatory features.
SEC-11	The device shall support TCG Storage Opal SSC Feature Set: Single User Mode Version 1.00, revision 2.00.
SEC-12	The device shall support TCG Storage Opal SSC Feature Set: PSID Version 1.00, revision 1.00.
SEC-12.1	PSID SHALL NOT be deterministic, and SHALL comprise at least 128 bits of entropy, and SHOULD comprise at least 160 bits of entropy. The entropy SHALL be unique for each drive.

	For example, PSID SHALL NOT be derived from the drive serial number or the date of manufacture, because these are deterministic. It SHALL NOT be the serial number of the drive hashed with a secret string shared between all drives, because the entropy SHALL be unique for each drive.
SEC-12.2	The generation method for the PSID SHALL be documented.
SEC-12.3	<p>After the PSID is provisioned in manufacturing, the PSID SHALL NOT be persistently stored in a form that makes direct retrieval of the PSID by firmware possible, either on the drive or outside the drive.</p> <p>One example of this is to store the PSID on the drive using a cryptographic hash such that it can only be used to verify a PSID input from the host.</p>
SEC-12.4	Once manufacturing is complete, the PSID SHALL NOT be stored anywhere other than the physical drive label.
SEC-13	The device shall support TCG Storage Feature Set: Block SID Authentication Version 1.01, revision 1.00 or newer.
SEC-14	Security audits, including firmware source code review, shall be provided to customers. This will include a manifest of source code provenance for the firmware; an audit of firmware build integrity (build infrastructure, signing infrastructure, source code integrity, authorization mechanisms to update source code and generate signed build); a testing gauntlet with detailed fuzzing, penetration and security property testing; the outcome of such testing; verification of artifacts like telemetry and debug logs, etc.
SEC-15	All signing keys shall be stored in a Hardware Security Module (HSM) that is either FIPS 140-2 Level 3 (or greater) certified or FIPS 140-3 Level 3 (or greater) certified.
SEC-16	Use of signing keys should be restricted to a small set of developers, following the principle of least privilege. Number of people with access and their corresponding roles shall be provided. Private signing keys should be generated by automation and not be modifiable by developers. No private signing keys should be transported in laptop or USB devices, or held in workstations.
SEC-17	<p>All debug ports shall be disabled before the device leaves the factory. Alternatively, the ports shall only be accessible in the field after a successful exchange of a challenge-response mechanism using an asymmetric crypto scheme (refer to NIST SP 800-63). The state shall be reset to inaccessible on any reset or power cycle. The debug port state should be reported in a signed attestation at device power on.</p>
SEC-18	<p>All vendor unique commands, log pages or set features that are not explicitly defined in this specification or approved in writing by the customer shall be disabled before the device leaves the factory. Alternatively, the commands/log pages/set features shall only be accessible in the field after a successful exchange of a challenge-response mechanism using an asymmetric crypto scheme (refer to NIST SP 800-63). The state shall be reset to inaccessible on any reset or power cycle. The access state for out-of-spec commands should be reported in a signed attestation at device power on</p>

SEC-19	Adversarial testing using red teams shall be conducted before qualification starts. A report of items attempted, and results shall be provided.
SEC-20	Vendor shall provide timely notification of security issues and delivery of fixes: <ul style="list-style-type: none"> <li>• Vendor shall document all security fixes with each firmware update.</li> <li>• Vendor shall notify the end customer within 7 days of discovering security issues in the device hardware or firmware.</li> <li>• Notification of issues shall include the process and timeline of the vendor's commitment to fix the issue: <ul style="list-style-type: none"> <li>o For privately disclosed vulnerabilities, the duration shall be no longer than 90 days.</li> <li>o For publicly disclosed vulnerabilities, the duration shall be no longer than 7 days.</li> <li>o Vendors shall notify the customers about the known CVEs and security issues and provide security-related updates before public announcement.</li> </ul> </li> </ul>
SEC-21	Supplier shall provide an example of the decoded Telemetry and debugging logs to the customer.
SEC-22	The device shall not include user data in plaintext or ciphertext form, passwords, keys and any secret or sensitive information in any Telemetry or debug logs.
SEC-23	All public keys shall be revocable.
SEC-24	Secure boot flow shall be immutable for exploitation and use public keys to verify the authenticity of the mutable code. For reference see the <a href="#">OCP Hardware Secure Boot White Paper</a> .
SEC-25	Secure firmware update flow shall be immutable for exploitation. There shall be no limits to the number of key rotations for secure update verification.
SEC-26	The device shall delete all keys from volatile memory as soon as they are no longer needed for operation during the current power on state.
SEC-27	The device shall only store host provided passwords, host provided keys, or any host provided secret information in non-volatile memory at any stage in an encrypted form. The encryption key for this protection shall not be stored in non-volatile memory.
SEC-28	The supplier must provide industry certification reports, if available, such as FIPS, NIST for device and device components such as TRNG, RNG, Crypto engine etc.
SEC-29	Log data and user data (data transferred from the Host in Write Commands) shall be stored on separate areas on the device. For example, in the system area and the user data area, respectively.
SEC-30	The vendor shall provide a comprehensive list of what is and what is not in the logs.
SEC-31	All telemetry and debug logs shall only be writable by device firmware after a successful secure boot.
SEC-32	There shall be no limit to the number of updates to the Security Version Number.
SEC-33	For secure boot failures, the device shall support SMBUS-3 using the recovery codes which are defined in the <a href="#">Recovery Document</a> which is referenced by the <a href="#">OCP Hardware Secure Boot White Paper</a> .

SEC-34	When locking is enabled for User Data, the Data Encryption Keys for that User Data must be cryptographically bound to the corresponding unlocking credentials: after a power cycle, the DEK must not be retrievable from the state of the device alone.
SEC-35	The device shall support TCG Storage Opal SSC Feature Set: Additional DataStore Tables Version 1.00, revision 1.00.
SEC-36	The device shall support TCG Storage Opal SSC feature set: Configurable Namespace Locking Version 1.00, revision 1.02.
SEC-37	The device shall support limiting authentication attempts as described in TCG Core §5.3.4.1.1.2 "Authentication Attempt Limits with C_PIN Objects." The TryLimit for each C_PIN should be 50 and shall be between 10 and 100 (inclusive).
SEC-38	Authenticating a C_PIN shall require between 50ms and 5000ms (inclusive), and should require approximately 250ms. The time limit should be enforced cryptographically, e.g. by tuning the number of PBKDF iterations to reach the target time.
SEC-39	Passwords of length 32 bytes (C_PIN PinLength of 32) shall be supported.
SEC-40	The device shall support at least 2 Static ComIDs and should support at least 4.
SEC-41	The device shall support at least 2 Dynamic ComIDs and should support at least 4.
SEC-42	MaxComIDTime shall be between 5 and 240 seconds (inclusive), and should be 60 seconds.
SEC-43	At least 1 Read-Write Session shall be supported per SP, and shall be indicated by MaxSessions. (It is not required to support an active RW session on the AdminSP at the same time as sessions on other SPs, because this would violate the TCG Core spec.)
SEC-44	At least 2 Read-Only Sessions should be supported per SP, and shall be indicated by MaxReadSessions.
SEC-45	<ul style="list-style-type: none"> <li>• TCG SessionTimeout SHALL be supported in the StartSession method call.</li> <li>• DefSessionTimeout SHALL be capable of being Read using the Properties Method.</li> <li>• The default DefSessionTimeout SHALL be 120,000 milliseconds.</li> <li>• MaxSessionTimeout SHALL be 0, This value indicates that there is no limit.</li> <li>• MinSessionTimeout SHOULD have a value of 500 (units in milliseconds).</li> <li>• Support for setting the SPSessionTimeout column of the SPInfo table is RECOMMENDED.</li> </ul>

## 10.2 Secure Boot

The device shall support Secure Boot. There are two fundamental things to address for secure boot:

### 10.2.1 Secure boot rooted in hardware.

#### 10.2.1.1 Core Root of Trust Measurement.

The vendor should follow the recommendations in the [TCG Publication for Hardware Requirements for a Device Identifier Composition Engine](#). DICE coupled with [RIOT Core](#) and Source for [RIOT](#) can help implement Cryptographic Identity with explicit attestation.

Requirement ID	Description
SBT-1	The device shall comply with the <a href="#">FIPS 186-4 Digital Signature Standard (DSS)</a> and the <a href="#">Open Compute Security Project Publication for Secure Boot Requirements</a> .
SBT-2	For Core Root of Trust measurement, each device shall have a Cryptographic Device Identity.
SBT-3	The <a href="#">TCG DICE standard</a> , or hardware based cryptographic identity shall be implemented. Each device shall use DICE to attest to its hardware and firmware identity, and shall use SPDML to attest to other security-relevant configuration.
SBT-4	The device should follow the guidance in the <a href="#">Commercial National Security Algorithm Suite</a> regarding quantum resistant algorithms and key sizes.
SBT-5	Secure boot flow shall be immutable for exploitation and use immutable public keys.

## 11 Debug & Failure Analysis Support

### 11.1 Debug Log Requirements

Requirement ID	Description
DEBUG-1	<p>The device shall support an internal event log that shall be extractable using NVMe-CLI vs-internal-log plug-in command. Below are some examples of data expected to be in the internal event logs.</p> <ul style="list-style-type: none"> <li>● Rolling event-log (with Event ID) &amp; Temperature</li> <li>● PCIe Completion Timeouts</li> <li>● PCIe Interface Errors (Correctable or Uncorrectable errors)</li> <li>● Log of a Critical Error</li> <li>● FATAL Errors/ FW Asserts</li> <li>● Read Only Mode Entry</li> <li>● Ungraceful shutdown events</li> <li>● Brownout Events (<i>if a voltage detector is present</i>)</li> <li>● Read/ Program/ Erase Failure Events</li> <li>● Detection of Bad Blocks/ Grown Bad Blocks</li> <li>● XOR "Error Handling" Triggers</li> <li>● Firmware Upgrade/ Downgrade events</li> <li>● Every time <i>Drive enters/ exits Degraded Mode</i></li> <li>● POH, FW-Version, Timestamp and Power Cycle # of each event occurrence</li> <li>● Exact Physical location of Grown Bad-Blocks (for NAND FA)</li> <li>● End-to-End Data Path Error Event (DRAM, SRAM any other buffers)</li> </ul>

DEBUG-2	The device shall support reporting of Hardware Revision C6h log page.
DEBUG-3	Hardware Revision update shall not trigger a Firmware Upgrade
DEBUG-4	Logging events to the internal log shall not have an impact on device performance.

## 11.2 NVMe CLI Management Utility

The NVMe CLI utility (<https://github.com/linux-nvme/nvme-cli>) shall be used as the management utility for NVMe devices.

Requirement ID	Description
UTIL-1	<p>The SSD supplier must test their SSDs with this utility and ensure compatibility. The following is the minimum list of commands that need to be tested with NVMe CLI:</p> <ul style="list-style-type: none"> <li>• Format (All options specified in IDENTIFY-2 must be verified). <code>nvme format</code></li> <li>• Sanitize. <code>nvme sanitize</code> and <code>nvme sanitize-log</code></li> <li>• FW upgrade &amp; downgrade. <code>nvme fw-download</code> and <code>nvme fw-commit</code></li> <li>• Controller reset to load FW. <code>nvme reset</code> and <code>nvme subsystem-reset</code></li> <li>• Log page reads including vendor log pages. <code>nvme get-log</code></li> <li>• SMART status. <code>nvme smart-log</code></li> <li>• List devices. <code>nvme list</code></li> <li>• Get/set features. <code>nvme get-feature</code> and <code>nvme set-feature</code></li> <li>• Namespace management <code>nvme create-ns</code>, <code>nvme attach-ns</code>, <code>nvme detach-ns</code> and <code>nvme delete-ns</code></li> <li>• Identify controller and namespace. <code>nvme id-ctrl</code>, <code>nvme id-ns</code>, and <code>nvme list-ns</code></li> <li>• Effects log page. <code>nvme effects-log</code></li> </ul>

## 11.3 NVMe CLI Plug-in Requirements

The device supplier shall develop and provide a Linux NVMe CLI plugin that meets the following requirements:

Requirement ID	Description
UTIL-PI-1	A single, common plugin for all the supplier's NVMe-based products. An example of using the NVMe-CLI plug-in is as given below:

	<code>nvme &lt;vendor_name&gt; &lt;command&gt; [&lt;device&gt;] [&lt;args&gt;]</code>
UTIL-PI-2	All plug-in data shall be decoded including into a human readable format and JSON output. Vendor-Unique logs can be in Binary format.
UTIL-PI-3	Access to vendor unique commands
UTIL-PI-4	All of the debug logs, crash dumps, NAND cell threshold voltage distributions shall be retrievable via NVMe-CLI, without using any external probes or tools.
UTIL-PI-5	The plugin's subcommand nomenclature must adhere to <a href="#">Section 16.2.1 - NVMe CLI Plug-In Nomenclature/Functional Requirements</a> and cannot change across versions.
UTIL-PI-6	The plugin shall use the existing NVMe CLI interface to access any vendor unique commands that are supported by the device.
UTIL-PI-7	If the NVMe CLI interface needs to transfer greater than 16 MB of data, the NVMe vendor unique Command shall have the ability to do multiple scatter/gather elements on the data buffer.
UTIL-PL-8	<a href="#">Source code</a> for the plug-in shall be shared with the customer.
UTIL-PL-9	The NVMe-CLI source code shall be backwards compatible when updates are released.

#### 11.4 NVMe CLI Plug-In Nomenclature/Functional Requirements

The NVMe CLI plugin must meet the following naming and functional requirements:

Requirement ID	NVMe CLI Nomenclature	Purpose
UTIL-NM-1	<code>vs-cloud-log</code>	Retrieve the SMART / Health Information Extended (Log Identifier C0h) from <a href="#">Section 5.4.6.2 - SMART / Health Information Extended (Log Identifier C0h)</a> . The output format when this command is issued shall be printed as described in <a href="#">11.5.1</a> and <a href="#">11.5.2</a>
UTIL-NM-2	<code>vs-internal-log</code>	Retrieves internal drive telemetry/debug logging which should include but not limited to events described in DEBUG-1. This is the primary interface to extract all debug data from the drives.
UTIL-NM-3	<code>vs-drive-info</code>	Outputs the following information:  <b>1) Drive_HW_revision</b> – Displays the current HW revision of the drive as described in <a href="#">Section 5.4.6.3 - Hardware Revision Log C6h</a> .  <b>2) FTL_unit_size</b> – Display Indirection Unit (IU) size. Units are in Bytes, so “4096” means the FTL unit size is 4KiB. The source of this information is from EXT-SMART-29 i.e. Bytes 324:327 of the Extended SMART C0h Log Page.

		<p><b>3) Boot Spec. Version</b> – Display the Boot SSD Specification Version compliance. Shall print version “<b>HyperScale Boot Version Spec.” + MAJOR.MINOR</b>” for this version of the HyperScale NVMe Boot SSD Specification. The MAJOR and MINOR version is retrieved from the EXT-SMART-28 field in the <a href="#">Extended SMART log</a>.</p> <p><b>4) Drive Ownership Status</b> - Display the TCG Device Ownership status. Device ownership is defined by inspecting the value of C_PIN_SID, if C_PIN_SID is set to MSID then this field will report <b>UNSET</b>, otherwise report <b>SET</b>. If BlockSID is enabled then report <b>BLOCKED</b>. The source of this information is from EXT-SMART-30.</p>
UTIL-NM-4	clear-pcie-correctable-errors	Calls Clear PCIe Correctable Error Counters (Feature Identifier C3h) Set Feature to clear the correctable PCIe error counters. See <a href="#">Section 5.4.7.5 - Clear PCIe Correctable Error Counters (Feature Identifier C3h) Set Feature</a> for more details.
UTIL-NM-5	log-page-directory	The NVMe command that lists all the log pages and a description of their contents.
UTIL-NM-6	cloud-boot-SSD-version	Prints version “HyperScale Boot Version 1.0”. The data for this resides in Log Page ID 0xC0 (Extended SMART Log Page).
UTIL-NM-7	Help	Display this help menu and usage of all sub-commands
UTIL-NM-8	vs-pcie-stats	<a href="#">PCIe Error Counters</a> data shall be returned when we execute the <b>vs-pcie-stats</b> command in the NVMe-CLI plug-in.
UTIL-NM-9	vs-nand-stats	Statistics related to media health shall be reported via this command. e.g. EXT-SMART-1,2,3,4,5,8,9,10,17,19,20,23,28 from the <a href="#">Extended SMART log</a> data shall be returned when we execute the <b>vs-nand-stats</b> command in the NVMe-CLI plug-in.
UTIL-NM-10	vs-temperature-stats	Temperature related metrics tracked by the SSD firmware shall be reported via the <b>vs-temperature-stats</b> command in the NVMe-CLI plug-in. Drive shall report Temperature Data from all sensors on-board as part of this command.
UTIL-NM-11	vs-device-waf	Calculate the SLC and TLC WAF (Write Amplification Factor) of the device and print/report the ratio. Use data from SMART-Log and



		Extended SMART/ Cloud-Log to calculate this value.
--	--	--

## 11.5 NVMe-CLI Command Output Data Format

Requirement ID	Description
CLI-OP-1	The section outlines the output format for the Cloud Log Page C0h. When command <code>vs-cloud-log</code> is executed by the plug-in the following format shall be followed to output the results.

### 11.5.1 Human-Readable/ Plain Text Format

```
Physical Media Units Written - TLC (Bytes) : 0
Physical Media Units Written - SLC (Bytes) : 576
Bad User NAND Block Count (Normalized)(Int) : 100
Bad User NAND Block Count (Raw)(Int) : 0
XOR Recovery Count (Int) : 0
Uncorrectable Read error count (Int) : 0
SSD End to End correction counts (Corrected Errors)(Int) : 0
SSD End to End correction counts (Detected Errors)(Int) : 0
SSD End to End correction counts (Uncorrected E2E Errors)(Int) : 0
System data %% life-used : 0%
User data erase counts (Minimum TLC)(Int) : 0
User data erase counts (Maximum TLC)(Int) : 0
User data erase counts (Minimum SLC)(Int) : 0
User data erase counts (Maximum SLC)(Int) : 0
User data erase counts (Average SLC)(Int) : 0
User data erase counts (Average TLC)(Int) : 0
Program fail count (Normalized)(Int) : 100
Program fail count (Raw)(Int) : 0
Erase Fail Count (Normalized)(Int) : 100
Erase Fail Count (Raw)(Int) : 0
PCIe Correctable Error count (Int) : 0
%% Free Blocks (User)(Int) : 100%
Security Version Number (Int) : 1
%% Free Blocks (System)(Int) : 88%
NVMe Stats (# Data Set Management/TRIM Commands Completed)(Int) : 0
Total Namespace Utilization (nvme0n1 NUSE)(Bytes) : 0
NVMe Stats (# NVMe Format Commands Completed)(Int) : 100%
Background Back-Pressure Gauge(%) (Int) : 0
Total # of Soft ECC Error Count (Int): 1285
Total # of Read Refresh count (Int) : 0
```

Bad System NAND Block Count (Normalized)(Int) : 100  
Bad System NAND Block Count (Raw)(Int) : 0  
Endurance Estimate (Total Writable Lifetime Bytes)(Bytes) : 300000  
Thermal Throttling Status & Count (Number of thermal throttling events)(Int) : 0  
Total # Unaligned I/O (Int) : 0  
Total Physical Media Units Read (Bytes)(Int) : 425  
Command Timeout (# of READ Command > Threshold → Filled by Vendor)(Int): 0  
Command Timeout (# of WRITE Command > Threshold → Filled by Vendor)(Int): 0  
Command Timeout (# of TRIM Command > Threshold → Filled by Vendor)(Int): 0  
Total PCIe Link Retraining Count (Int): 0  
Active Power State Change Count (Int): 0  
Cloud Boot SSD Spec Version (Int): 1.0.0.0  
Cloud Boot SSD HW Revision (Int): 1.2.1.0 (example)  
TCG Ownership Status: 0x0  
Log Page Version (Int) : 3  
Log Page GUID (Hex) : 0xC46DD7920F1E4266A178D8AC78884365

### 11.5.2 JSON Format

Requirement ID	Description
CLI-OP-2	This section defines a machine-readable JSON format example. The object name specified in the section below shall be followed when printing the output of vs-cloud-log plugin sub-command in a JSON format.

```
{
  "definitions": {},
  "$schema": "http://json-schema.org/draft-07/schema#",
  "$id": "https://example.com/object1635265067.json",
  "title": "This is an example of a json schema for out SMART
attribute defaults",
  "type": "object",
  "required": [
    "physical_media_units_bytes_tlc",
    "physical_media_units_bytes_slc",
    "bad_user_blocks_normalized",
    "bad_user_blocks_raw",
    "xor_recovery_count",
    "uncorrectable_read_errors",
    "corrected_e2e_errors",
    "detected_e2e_errors",
    "uncorrected_e2e_errors",
    "system_data_life_used_pct",
    "min_slc_user_data_erase_count",
    "min_tlc_user_data_erase_count",
    "max_slc_user_data_erase_count",
    "max_tlc_user_data_erase_count",
    "avg_tlc_user_data_erase_count",
    "avg_slc_user_data_erase_count",
    "program_fail_count_normalized",
    "program_fail_count_raw",
    "erase_fail_count_normalized",
    "erase_fail_count_raw",
    "pcie_correctable_errors",
    "pct_free_blocks_user",
    "security_version",
    "pct_free_blocks_system",
    "num_of_trim_commands",
    "total_nuse_bytes",
    "num_of_format_commands",
    "background_pressure_gauge",
    "soft_ecc_error_count",
    "read_refresh_count",
```

```

    "bad_system_block_normalized",
    "bad_system_block_raw",
    "endurance_est_bytes",
    "num_throttling_events",
    "total_unaligned_io",
    "physical_media_units_read_bytes",
    "num_read_timeouts",
    "num_write_timeouts",
    "num_trim_timeouts",
    "pcie_link_retrain_count",
    "active_power_state_change_count",
    "cloud_boot_ssd_spec_ver",
    "cloud_boot_ssd_hw_ver",
    "worst_case_die_location",
    "log_page_ver",
    "log_page_guid",
],
"properties": {
    "avg_slc_user_data_erase_count": {
        "$id": "#root/avg_slc_user_data_erase_count",
        "title": "User data erase counts (Average SLC) (Int)",
        "type": "integer",
        "examples": [
            1000000
        ],
        "default": 0
    },
    "avg_tlc_user_data_erase_count": {
        "$id": "#root/avg_tlc_user_data_erase_count",
        "title": "User data erase counts (Average TLC) (Int)",
        "type": "integer",
        "examples": [
            1000000
        ],
        "default": 0
    },
    "bad_user_blocks_raw": {
        "$id": "#root/bad_user_blocks_raw",
        "title": "Bad User NAND Block Count (Raw) (Int)",
        "type": "integer",
        "examples": [
            100000
        ],
        "default": 0
    }
}

```

```

...
    }
}
}

```

## 11.6 Performance Monitoring

Requirement ID	Description
PERFMON-1	The device shall support the Latency Monitoring Feature Set as described in the Appendix C of <a href="#">OCP NVMe SSD Requirements v2.0</a> Document.

## 12 Mechanical

### 12.1 Form factor

Requirement ID	Description
FF-1	The device shall adhere to the M.2 specification with a size of 22mm x 80mm.
FF-2	PCB height shall be conforming to PCI-SIG M.2 spec R4.0 and shall not exceed the Label-S3 dimensions: Top Max: 1.50mm, Bottom Max: 0mm
FF-3	The device shall use an M key.
FF-4	The bottom-side of the PCB (as defined in M.2 Form-Factor Specification) shall not have any conducting elements (e.g. test-points, debug ports, components etc.).
FF-5	A CAD file for the device shall be provided to the customer.

## 13 Electrical

### 13.1 Power consumption

#### 13.1.1 Power Consumption Methodology & Requirements

Requirement ID	Description
PCM-1	The device max average power consumption for any workload (including Full Drive Erase/ Sanitize) shall not exceed the maximum average power as configured in <a href="#">PWR-1 (power state descriptor table)</a> in a 500ms window with a sampling rate of 2ms or better. The measurement duration shall be at least 15 minutes on a pre-conditioned device.
PCM-2	The device peak power for any workload shall not exceed 11.55W and shall be measured in a 100us window with a sampling rate of 4uS or better. The measurement duration shall be at least 15 minutes on a pre-conditioned device.

### 13.1.2 Host Based Power & Thermal Management

Requirement ID	Description																																								
PWR-1	<p>The following operational power-states table shall be implemented by the device firmware (at a minimum):</p> <table><tr><th>Req. ID</th><th>Power State</th><th>Maximum Power</th><th>Entry Latency (ENTLAT)</th><th>Exit Latency (EXLAT)</th></tr><tr><td>PWR-1a</td><td>0</td><td>11.55W</td><td>IHV*</td><td>IHV</td></tr><tr><td>PWR-1b</td><td>1</td><td>8.25W</td><td>IHV</td><td>IHV</td></tr><tr><td>PWR-1c</td><td>2</td><td>6.5W</td><td>IHV</td><td>IHV</td></tr><tr><td>PWR-1d</td><td>3</td><td>5W</td><td>IHV</td><td>IHV</td></tr><tr><td>PWR-1e</td><td>4</td><td>4W</td><td>IHV</td><td>IHV</td></tr><tr><td>PWR-1f</td><td>5</td><td>3W</td><td>IHV</td><td>IHV</td></tr><tr><td>PWR-1g</td><td>6</td><td>2W</td><td>IHV</td><td>IHV</td></tr></table> <p>*Table entries containing IHV (Independent Hardware Vendor) are to be filled out by the manufacturer. If a given power state is not supported, then the device shall report FFFFFFFFh for the Entry Latency (ENTLAT) and Exit Latency (EXLAT). All other entries in the power state descriptor table are “Don’t Care”.</p>	Req. ID	Power State	Maximum Power	Entry Latency (ENTLAT)	Exit Latency (EXLAT)	PWR-1a	0	11.55W	IHV*	IHV	PWR-1b	1	8.25W	IHV	IHV	PWR-1c	2	6.5W	IHV	IHV	PWR-1d	3	5W	IHV	IHV	PWR-1e	4	4W	IHV	IHV	PWR-1f	5	3W	IHV	IHV	PWR-1g	6	2W	IHV	IHV
Req. ID	Power State	Maximum Power	Entry Latency (ENTLAT)	Exit Latency (EXLAT)																																					
PWR-1a	0	11.55W	IHV*	IHV																																					
PWR-1b	1	8.25W	IHV	IHV																																					
PWR-1c	2	6.5W	IHV	IHV																																					
PWR-1d	3	5W	IHV	IHV																																					
PWR-1e	4	4W	IHV	IHV																																					
PWR-1f	5	3W	IHV	IHV																																					
PWR-1g	6	2W	IHV	IHV																																					
PWR-2	<p>The default power state shall be set as follows:</p> <table><tr><th>Req. ID</th><th>Drive Physical Capacity</th><th>Default Power State</th></tr><tr><td>PWR-2a</td><td>512GB</td><td>PWR-1d (PS3)</td></tr><tr><td>PWR-2b</td><td>256GB</td><td>PWR-1e (PS4)</td></tr><tr><td>PWR-2c</td><td>128GB</td><td>PWR-1f (PS5)</td></tr><tr><td>PWR-2d</td><td>1024GB</td><td>PWR-1b (PS1)</td></tr><tr><td>PWR-2e</td><td>2048GB</td><td>PWR-1b (PS1)</td></tr></table>	Req. ID	Drive Physical Capacity	Default Power State	PWR-2a	512GB	PWR-1d (PS3)	PWR-2b	256GB	PWR-1e (PS4)	PWR-2c	128GB	PWR-1f (PS5)	PWR-2d	1024GB	PWR-1b (PS1)	PWR-2e	2048GB	PWR-1b (PS1)																						
Req. ID	Drive Physical Capacity	Default Power State																																							
PWR-2a	512GB	PWR-1d (PS3)																																							
PWR-2b	256GB	PWR-1e (PS4)																																							
PWR-2c	128GB	PWR-1f (PS5)																																							
PWR-2d	1024GB	PWR-1b (PS1)																																							
PWR-2e	2048GB	PWR-1b (PS1)																																							
PWR-3	The Set Features for Power Management with the SV bit 31 in Command DWORD 10 shall be supported so that the power level can be set and will be saved across power cycles.																																								
PWR-4	If the host selects a non-supported power state, which is higher power than what the device can support, the drive shall accept the command, indicate the drive is operating at the selected power state when a get feature command is executed, and run at the maximum supported power.																																								

### 13.2 Voltage Detector

Requirement ID	Description
VD-1	All brownout scenarios shall be recorded in the device's internal event log if brownout detection is supported.

### 13.3 SMBus Support

Requirement ID	Description
SMBUS-EL-1	The device electricals shall follow the PCI Express M.2 specification.
SMBUS-EL-2	The device's SMBus protocol shall comply with the System Management Bus (SMBUS) Specification <a href="#">Version 3.1</a> .
SMBUS-EL-3	The device should support I3C, at the point in time that I3C is ratified for M.2 by PCI SIG it shall then be a requirement for devices to support I3C. This shall include supporting a I2C fall back method for backward compatibility in legacy systems.

### 13.4 PCIe Link Equalization

Requirement ID	Description
LE-1	All Tx equalization presets (P0 to P10) as specified by the PCIe Spec. shall be supported

### 13.5 GND Pins

Requirement ID	Description
GND-PIN-1	The device shall tie all the GND pins(15 pins) defined in Socket 3 PCIe-based Adapter Pinouts (Key M) of PCI-SIG PCIe M.2 SPEC R4.0 to GND. Pin1 shall be used as a Presence Detect Pin.

## 14 Thermal

### 14.1 Operating Conditions

The SSD shall be operated under the following conditions:

#### 14.1.1 Data Center Altitude

Requirement ID	Description
THRM-1	The data centers may be located at an altitude of up to 2000 meters above sea level. The drive needs to operate in this condition without issues.

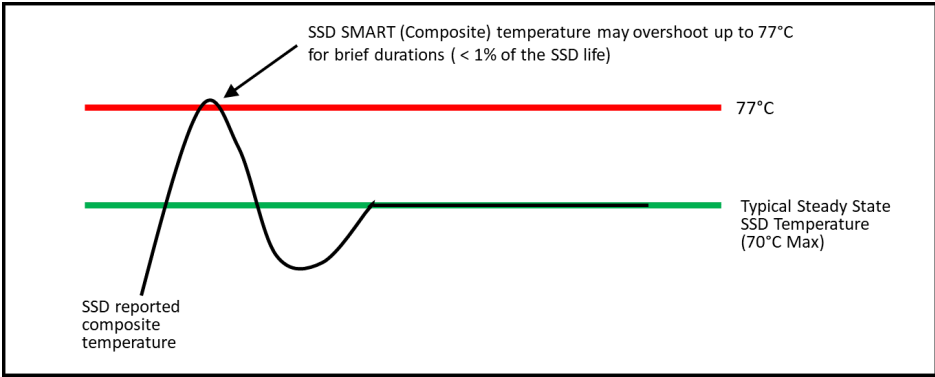
### 14.1.2 Operational Temperature/ Relative Humidity

Requirement ID	Description
OP-TH-1	The SSD shall operate normally with relative humidity to be between 10% and 90%.
OP-TH-2	The SSD shall operate normally ( <i>without losing performance</i> ) with composite temperature ( <i>measured by the thermal sensor</i> ) to be between 0°C and 76°C

### 14.1.3 Non-Operational Temperature/Relative Humidity

Requirement ID	Description
NOP-TH-1	Non-operational relative humidity environment is to be between 10% and 90%.
NOP-TH-2	Non-operational temperature environment is to be between -40°C and 85°C.

## 14.2 Thermal Throttling

Requirement ID	Description
TTHROTTLE-1	The device shall implement a thermal throttling mechanism to protect the device in case of a failure or excursion that causes the device's composite temperature to increase above its maximum specified temperature.
TTHROTTLE-2	When a temperature throttle occurs, an Asynchronous Event Request shall be completed with the Asynchronous Event Type field set to 001b (SMART / Health status) and the Asynchronous Event Information field set to 01h (Temperature Threshold). The device shall set bit 1 of the SMART / Health Information (Log Identifier 02h) Critical Warning field to 1b.
TTHROTTLE-3	<p>Under normal operating conditions the SSD shall not engage in thermal throttling when installed in OCP systems as they will maintain the device temperature at or below 70C (as reported by the thermal sensor). The required behavior is illustrated below:</p>  <p style="text-align: center;"><i>Figure 2 Boot SSD thermal management scheme</i></p>



TTHROTTLE-4	The firmware algorithm shall deploy safeguards to prevent a false activation of either thermal throttling or thermal shutdown. Example of a false activation would be a glitch in any of the sensor readings which would cause the composite temperature to reach the thermal throttling or thermal shutdown limit.
TTHROTTLE-5	A composite temperature of 77°C (170.6°F) shall be used for throttling.
TTHROTTLE-6	Thermal throttling shall not start based on the rate of temperature increase or slew rate.
TTHROTTLE-7	When the device is in the thermal throttling state and the temperature drops back down below 75°C (167°F), the device shall exit the thermal throttling state and an Asynchronous Event Request shall be completed with the Asynchronous Event Type field set to 001b (SMART / Health status) and the Asynchronous Event Information field set to 01h (Temperature Threshold). The device shall clear bit 1 of the SMART / Health Information (Log Identifier 02h) Critical Warning field to 0b.
TTHROTTLE-8	When the device reaches a critical temperature on any component it shall report a composite temperature of 85°C (185°F) and an Asynchronous Event Request shall be completed with the Asynchronous Event Type field set to 001b (SMART / Health status) and the Asynchronous Event Information field set to 01h (Temperature Threshold) before the device shuts off to protect itself.
TTHROTTLE-9	The device shall report a value of 015Eh (77°C) in the Warning Composite Temperature Threshold (WCTEMP) field of the Identify Controller Data structure.
TTHROTTLE-10	The device shall report a value of 0166h (85°C) in the Critical Composite Temperature Threshold (CCTEMP) field of the Identify Controller Data structure.

## 15 Out-of-Band Management Support

### 15.1 NVMe Basic Management Command (Appendix A) Requirements

Requirement ID	Description
SMBUS-1A	The device shall support the NVMe Basic Management Command as defined in Appendix A of the NVMe Management Interface 1.1b specification: ( <a href="https://nvmexpress.org/wp-content/uploads/NVM-Express-Management-Interface-1.1b-2020.10.05-Ratified.pdf">https://nvmexpress.org/wp-content/uploads/NVM-Express-Management-Interface-1.1b-2020.10.05-Ratified.pdf</a> ).
SMBUS-1B	SMBus Block Read protocol and Byte Read protocol shall be supported.
SMBUS-2	The Secure Boot Failure Feature Reporting Supported bit at offset 243 shall be supported and set to 1b to indicate support for SEC-6.
SMBUS-3	When there is a secure boot failure the device shall report the failure with the following behavior:

		Bit	Description
		7	Secure Boot Failure Feature Reporting Supported bit at offset 243 shall be set to 1b. See <a href="#">Command Code 242 (Secure Boot Failure Reporting)</a> for definition.
		6	Secure Boot Failure Status bit at offset 243 shall be set to 1b. See <a href="#">Command Code 242 (Secure Boot Failure Reporting)</a> for definition.
		5	Shall be set to 1b if codes are supported. If this bit is set to 1b then a valid Recovery code shall be entered in <a href="#">byte 244</a> .
		4:0	Reserved. Shall be cleared to zero.
SMBUS-4	The device shall implement the SMBus format for Basic Management commands ( <i>Appendix A of the NVMe Management Interface 1.1b specification</i> ) as shown in <a href="#">section 15.3</a> .		
SMBUS-5	The device shall take no longer than CAP.TO to produce stable SMBUS output values through the NVMe Basic Management Interface protocol after a device power-on		
SMBUS-6	The device shall generate the PEC values specified for each command code in the SMBus data format described in <a href="#">Section 15.3 - NVMe Basic Management Command (Appendix A) Data Format</a> .		
SMBUS-7	All data shall be returned in a big-endian format unless otherwise noted.		
SMBUS-8	The device shall support SMBus ARP.		
SMBUS-9	The default SMBUS/I2C address shall follow the NVMe Management Interface Specification version 1.1b or newer.		
SMBUS-10	SMBUS interface shall not be dependent on the health of PCIe Interface.		

## 15.2 VPD

Requirement ID	Description
VPD-1	VPD support is optional. If implemented, it shall support IPMI Platform Management FRU Information.

## 15.3 NVMe Basic Management Command (*Appendix A NVMe-MI Spec.*) Data Format

For devices supporting Appendix A NVMe-MI the following is required.

Command Code (Decimal)	Offset (Decimal)	Description
0	0	Defined in NVM Express Management Interface 1.1b.
8	8	Defined in NVM Express Management Interface 1.1b.
32	32	Length of GUID. This is the number of bytes until the PEC code. This shall be 16 decimal (0x10).

	48:33	GUID: This is a 16-byte Global Unique Identifier. The GUID shall be 0xC035E2BC3B4A40E982E61BDEC28EFA72
	49	PEC: An 8-bit CRC calculated over the slice address, command code, second slave address and returned data. Algorithm is defined in SMBus Specifications.
50	95:50	Reserved: Shall be set to 0x0.
96	96	Length of Device Info: Indicates number of additional bytes to read before encountering PEC. This value should always be 56 (38h) in implementations of this version of the spec.
	104:97	Firmware Version Number: This field shall indicate the activated firmware version that is running on the device after the firmware activation took place. The format of this field shall be as defined in field Firmware Revision (FR) section 5.15.2.2 Identify Controller Data Structure of the NVMe specification version 1.4c.
	108:105	This is the device raw capacity in GB in Hex (e.g. 2048 GB in raw capacity = 0x800). Does not include any extra spare blocks within the NAND.
	112:109	Reserved: Shall be set to 0x0.
	152:113	Product Part/Model Number. The reason for 40 bytes is to keep this consistent with NVMe that already has this field in the identify command.
	153	PEC: An 8-bit CRC calculated over the slice address, command code, second slave address and returned data. Algorithm in SMBus Specifications.
154	241:154	Reserved: Shall be set to 0x0.
242	242	Length of Secure Boot Status: Indicates number of additional bytes to read before encountering PEC. This value should always be 4 (0x4).
	243	Bit 7: Secure Boot Failure Feature Reporting Supported When set to 0x1 the secure boot feature reporting is supported. When set to 0x0 the secure boot failure feature reporting is not supported.

		<p>Bit 6: Secure Boot Failure Status: When set to 0b there is no secure boot failure. When set to 1b there is a secure boot failure. This bit shall only be set if the Secure Boot Feature Supported bit is set to 1b and there is a secure boot failure.</p> <p>Bit 5: OCP Recovery/ Platform Root-of-Trust for Recovery: When set to 1b, OCP Recovery/ Platform Root-of-Trust for Recovery codes are supported in byte 244. When cleared to 0b OCP Recovery/ Platform Root-of-Trust for Recovery codes are not supported and byte 244 shall be cleared to zero.</p> <p>Bit 4:0 Reserved. Shall be cleared to 0x0.</p>
	244	Recovery Code: OCP Recovery/Platform Root-of-Trust for Recovery code.
	246:245	Reserved: Shall be set to 0x0.
	247	PEC: An 8-bit CRC calculated over the slice address, command code, second slave address and returned data. Algorithm in SMBus Specifications.
248	248	Length of Log Page Version Number: Indicates number of additional bytes to read before encountering PEC. This value should always be 6 (0x06) in implementations of this version of the spec.
	250:249	Log Page Version Number: Indicates the version of this mapping used in the device. Shall be set to '5' (0x05) after an SMBus block read is completed.
	254:251	Reserved: Shall be set to 0000h.
	255	PEC: An 8-bit CRC calculated over the slice address, command code, second slave address and returned data. Algorithm in SMBus Specifications.

#### 15.4 NVMe-MI Requirements

For devices supporting NVMe-MI, all the following requirements apply.

Requirement ID	Description
NVMe-MI-1	The device shall support NVMe Management Interface Specification version 1.1b or later.
NVMe-MI-2	The device shall support MCTP over SMBus.
NVMe-MI-3	The device shall support SMBUS Fixed and Discoverable.

NVMe-MI-4	The device shall NACK any SMBus/I2C addresses not listed in the NVM Express Management Interface 1.1b, or assigned through SMBUS ARP.
NVMe-MI-5	All mandatory NVMe Management Interface commands defined by the NVMe Management Interface Specification 1.1b shall be supported through MCTP over SMBus.
NVMe-MI-6	All mandatory NVMe Admin commands defined by the NVMe Management Interface Specification 1.1b shall be supported through MCTP over SMBus.
NVMe-MI-7	<p>The device shall support the following optional NVMe Management Interface Commands by MCTP over SMBus, even when the PCIe Links are not active:</p> <ul style="list-style-type: none"> <li>● Firmware Activate/Commit</li> <li>● Firmware Image Download</li> <li>● Security Send</li> <li>● Security Receive</li> </ul> <p>The device may support the following optional NVMe Management Interface Commands:</p> <ul style="list-style-type: none"> <li>● Device Self-test</li> <li>● Format NVM</li> <li>● Namespace Management</li> <li>● Namespace Attachment</li> <li>● Sanitize</li> <li>● Set Features</li> </ul> <p>Note that our intention is that in future versions of this specification, the above optional commands will become mandatory.</p>
NVMe-MI-8	<p>The following Log Identifiers shall be supported by MCTP over SMBus, even when the PCIe Links are not active:</p> <ul style="list-style-type: none"> <li>● SMART / Health Information Log (02h)</li> <li>● Error Information Log (01h)</li> <li>● Firmware Slot Information (03h)</li> <li>● Device Self-test (06h)</li> <li>● Sanitize Status (81h)</li> <li>● Persistent Event Log (0Dh)</li> <li>● SMART Extended Log Page (C0h)</li> </ul>

## 16 Labeling Requirements

Below are the requirements for the label placed on devices conforming to this specification.

### 16.1 Label Requirements

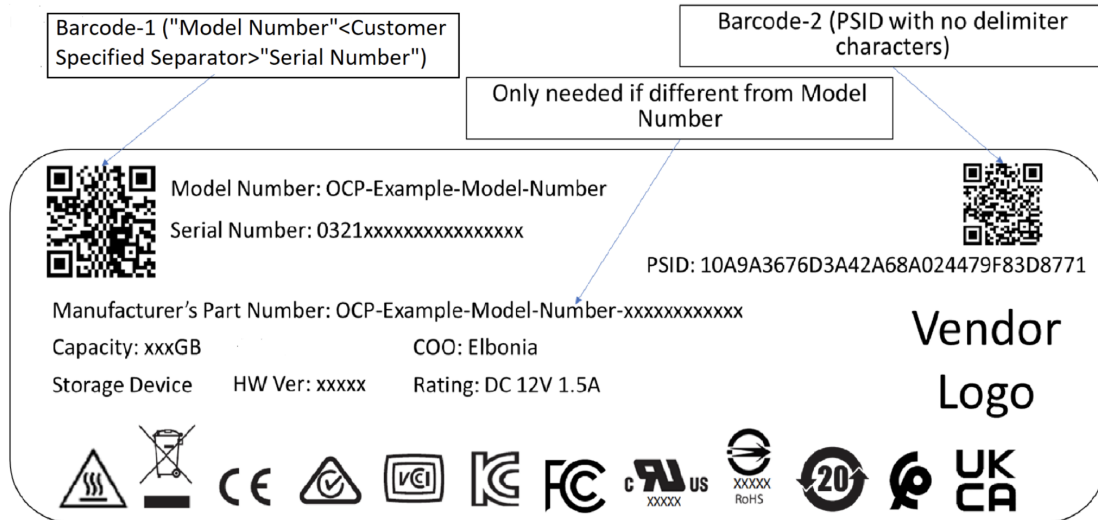


Figure 3 HyperScale Boot Device Label Format







Requirement ID	Description				
LABL-1	The following fields are required information that shall be placed on the label:				
	Item	Format	Text Required	Barcode Required	Barcode Type
	Barcode-1	'Model Number' 'Customer Defined Separator' 'Serial Number' \n.	No	Yes	2d
	Model Number	See LABL-11 (Model Number must match).	Yes	No	N/A
	Serial Number	See LABL-12 (Serial Number must match), LABL-15 (certification logos), and LABL-17 (Serial Number format).	Yes	No	N/A











	Manufacturer's Part Number	Number used for ordering.	Yes, if different from Model Number	No	N/A
	Capacity	Number of GB or TB.	Yes	No	N/A
	STORAGE DEVICE	Text shall be "STORAGE DEVICE".	Yes	No	N/A
	PSID	TCG-OPAL Spec.	Yes	No	N/A
	Barcode-2	'PSID' \n	No	Yes	2d
	HW Revision		Yes	No	N/A
	Regulatory Mark	See LABL-15 (certification logos).		No	N/A
	Country Certification Numbers	See LABL-15 (certification logos).		No	N/A
	Certification Logos	See LABL-15 (certification logos).		No	N/A
	RoHS/ Green	See LABL-15 (certification logos).		No	N/A
LABL-2	The Model Number on the shipping label shall match the Model Number used during qualification.				
LABL-3	The minimum font size shall be 3 points and the typical size should be 6 points.				
LABL-4	For the Capacity field, if there are space constraints, the manufacturer may remove "Capacity:" and just show "XXXGB" or "XXXTB".				
LABL-5	<p>To distinguish Model Number and Serial Number, Barcode-1 shall have a customer defined separator between the Model Number portion and the Serial Number portion. For Example:</p> <p>Model Number: OCP-Example-Model-Number</p> <p>Serial Number: 0321xxxxxxxxxxxxxxxxxx</p> <p>Barcode-1 Readout: OCP-Example-Model-Number&lt;Customer Defined Separator&gt;0321xxxxxxxxxxxxxxxxxx</p> <p>Customer Defined Separator can be as follows:</p> <ol style="list-style-type: none"><li>OCP-Example-Model-Number_0321xxxxxxxxxxxxxxxxxx</li><li>OCP-Example-Model-Number 0321xxxxxxxxxxxxxxxxxx</li></ol>				
LABL-6	There shall be a line with the text "STORAGE DEVICE".				
LABL-7	The following fields are optional information that can be placed on the label at the discretion of the device maker. Placement is also at the device makers discretion if				

such information does not interfere with the mandatory information above. No additional barcode shall be present.

Item	Format	Text Required	Barcode Required	Barcode Type
Processor Code (BA)		Optional	No	N/A
Maker Logo		Optional	No	N/A
Rated Voltage & Current		Optional	No	N/A
Production Date	DDMMYYYY: DD (Date), MM (Month), YYYY (Year)	Optional	No	N/A
Weekly Code	YYWW: YY (Year), WW (Week)	Optional	No	N/A
Warranty VOID IF REMOVED		Optional	No	N/A
Makers Own Label Material Number		Optional	No	N/A
Website, Company Address		Optional	No	N/A
SSD		Optional	No	N/A
Product Series Name		Optional	No	N/A
SA: Value used within manufacturin g		Optional	No	N/A
PBA: Physical Board Address (identifies the physical configuration of the device)		Optional	No	N/A
WWN: World Wide Number (unique for each device)		Optional	No	N/A



LABL-8	To ensure that data-center operations personnel can quickly and easily identify devices that have been ticketed for field replacement, it is mandatory to have the proper identifying fields on the label(s), in the format specified below.																		
LABL-9	The label shall not degrade over the standard device lifetime under standard operating conditions.																		
LABL-10	For M.2 the label shall be placed on the customer specific side of the device as defined in the Appendix section of this spec.																		
LABL-11	The Model Number in Barcode-1, the Model Number printed on the label, the Model Number in the Product Datasheet and the Model Number returned in Identify Controller Data Structure (CNS 01h, byte offset 63:24) shall all match at all times after manufacturing completes. This includes device states, such as firmware authentication failures, that may prevent complete loading of the device firmware. The requirement may be exempted in writing by the customer.																		
LABL-12	The Serial Number in Barcode-1, the Serial Number printed on the label and the Serial Number returned in Identify Controller Data Structure (CNS 01h, byte offset 23:04) shall all match at all times after manufacturing completes. This includes device states, such as firmware authentication failures, that may prevent complete loading of the device firmware																		
LABL-13	This Hardware Revision printed on the label and returned by the NVMe-CLI utility shall match, and shall match Global HW Revision (HWREV-1).																		
LABL-14	All other electronically readable information shall also match their counterparts printed on the label.																		
LABL-15	<p>The following certification logos and their corresponding certifications are required:</p> <table border="1"> <thead> <tr> <th>Regulatory Mark/Text</th><th>Description</th></tr> </thead> <tbody> <tr> <td>Regulatory Model Number</td><td>Unique regulatory Identifier.</td></tr> <tr> <td>Made in XXXX</td><td>Country of Origin.</td></tr> <tr> <td>Manufacturer or Brand name</td><td>Identification of the responsible party for placing the device into the market.</td></tr> <tr> <td>Address of the Manufacturer</td><td>Required for devices with the CE mark or UKCA mark.</td></tr> <tr> <td>Date of Manufacture</td><td>Not needed if embedded in the Serial Number.</td></tr> <tr> <td>Serial Number</td><td>Alphanumeric, 12-20 digits with first 4 digits indicating: Date of Manufacturing in Work Week and Year WWYY1234567890123456.</td></tr> <tr> <td></td><td>[Europe] Compliance with EU WEEE directive 2010/19/EU.</td></tr> <tr> <td></td><td>[Europe] Compliance with EU EMC directive 2014/30/EU and RoHS directive 2011/65/EU.</td></tr> </tbody> </table>	Regulatory Mark/Text	Description	Regulatory Model Number	Unique regulatory Identifier.	Made in XXXX	Country of Origin.	Manufacturer or Brand name	Identification of the responsible party for placing the device into the market.	Address of the Manufacturer	Required for devices with the CE mark or UKCA mark.	Date of Manufacture	Not needed if embedded in the Serial Number.	Serial Number	Alphanumeric, 12-20 digits with first 4 digits indicating: Date of Manufacturing in Work Week and Year WWYY1234567890123456.		[Europe] Compliance with EU WEEE directive 2010/19/EU.		[Europe] Compliance with EU EMC directive 2014/30/EU and RoHS directive 2011/65/EU.
Regulatory Mark/Text	Description																		
Regulatory Model Number	Unique regulatory Identifier.																		
Made in XXXX	Country of Origin.																		
Manufacturer or Brand name	Identification of the responsible party for placing the device into the market.																		
Address of the Manufacturer	Required for devices with the CE mark or UKCA mark.																		
Date of Manufacture	Not needed if embedded in the Serial Number.																		
Serial Number	Alphanumeric, 12-20 digits with first 4 digits indicating: Date of Manufacturing in Work Week and Year WWYY1234567890123456.																		
	[Europe] Compliance with EU WEEE directive 2010/19/EU.																		
	[Europe] Compliance with EU EMC directive 2014/30/EU and RoHS directive 2011/65/EU.																		

		[Australia, New Zealand] Compliance with requirements of the relevant Australian ACMA Standards, under the Radiocommunications Act 1992 and the Telecommunications Act 1997.
		[Japan] Compliance with Japan VCCI requirements.
	 XXXX-XX-XX	[Korea] Compliance with requirements of the Radio Research Laboratory Ministry of Information and Communication Republic of Korea.
	CAN ICES-3(*)/NMB-3(*)	[Canada] Compliance with Canada standard ICES 003. Where * is either A or B.
		[USA] Mandatory. Compliance with United States Federal Communications Commission requirements.
	 XXXXX	[USA] Compliance with UL standards and Canadian Safety Standards.
	 XXXXX RoHS	[Taiwan] Compliance with Taiwan EMC and RoHS.
		[China] Compliance with Chinese environmental requirements. Number inside the circle is usually 10 or 20.
		[Morocco] Compliance with Moroccan EMC standards.
		[United Kingdom] Mandatory after 1/1/2022. Guidance to UKCA marking.
LABL-16	<p>If the surface of any component or casing will reach a temperature of 70°C (158°F) or greater the following warning logo shall be either printed on the label or placed separately on the device:</p> 	
LABL-17	The format of the serial number shall be WWYYSerialNumber. For Example: WWYY1234567890123456.	
LABL-18	Barcodes shall be printed using Datamatrix ECC200 or Model 2 QR code only.	
LABL-19	QR codes shall use a minimum of ECC Level M (15%).	
LABL-20	The density of the barcode shall be 10 mils or larger.	

LABL-21	The label shall only be printed on Polyester or Plastic labels using a Wax/Resin ribbon.
LABL-22	The PSID shall be printed on the label in its direct 32-character alphanumeric representation without any ancillary delimiting characters (e.g., underscore, dash, backslash, forward slash, etc.) and shall exactly match the readout of Barcode-1.

## 17 Environmental Considerations

### 17.1 RoHS Compliance

Requirement ID	Description
RoHS-1	The Supplier shall provide component-level reporting on the use of listed materials by concentration (ppm) for all homogenous materials.

### 17.2 ESD Compliance

Requirement ID	Description
ESD-1	The Device manufacturer needs to provide ESD immunity level (HBM- Human Body Model) measured in accordance with ANSI/ESDA/JEDEC JS-001-2010 spec. (IEC-61000-4-2)
ESD-2	ESD testing shall include testing gold-fingers.

## 18 Sustainability Requirements

Requirement ID	Description
SUS-1	A Life Cycle Assessment (LCA) performed in accordance with ISO 14040 for the device shall be provided to the customer.
SUS-2	Recycled content information per material type shall be provided to the customer.

## 19 Shock and Vibration Requirements

The device vendors shall conform to the following requirements with respect to a M.2 SSD module:

Requirement ID	Description
SV-1a	The non-operational shock requirement is 700G, half-sine, 0.5ms, total 6 shocks, along all three axes (+/-)
SV-1b	The vibration requirement during operation is: 1.8G <sub>rms</sub> , 5-500-5 Hz, Random Vibe, 20 min along all three axes
SV-1c	The vibration requirement during non-operation is: 3.13G <sub>rms</sub> , 5-800-5 Hz, total 6 sweeps along all three axes, 20 minutes per sweep

SV-1d	<p>Validation flow for Shock and Vibration:</p> <ol style="list-style-type: none"> <li>1. UUT (Unit Under Test), test fixture should be visually inspected and ensured that everything is torqued or secured as needed. Pictures of test fixtures with and w/o UUT should be provided.</li> <li>2. Baseline performance of the drive should be gathered and used as a reference against post S&amp;V data to ensure no performance impact incurred.</li> <li>3. Once S&amp;V testing is completed, repeat visual inspection to the UUT and test fixture to ensure no physical damage or performance impact has occurred to the UUT or test fixture.</li> <li>4. Re-run stress test on the UUT in case of non-op test and provide data indicating no performance impact incurred to the unit.</li> </ol>
-------	---

## Appendix A: CLA Format

Requirements	Details	Link to which Section in Spec
Contribution License Agreement	OPTION B: Open Web Foundation (OWF) CLA	
Are All Contributors listed in Sec 1: License?	Yes	
Did All the Contributors sign the appropriate license for this spec? Final Spec Agreement/HW License?	Yes	

Which 3 of the 4 OCP Tenets are supported by this Spec?	Efficiency Scalability Openness Impact	<ul style="list-style-type: none"> <li>● Efficiency: Enables a common design that can be used across multiple end users which reduces the number of SKUs.</li> <li>● Scalability: Simpler design allows for scale, manageability and reliability</li> <li>● Openness: Provides an open specification for hyperscale boot drive SSD requirements.</li> <li>● Impact: This contribution provides the community a public SSD specification for hyperscale boot drive SSD users to focus on. Enables a more robust supply chain and increased quality. Enabling the eco-system to have an open component specification thus enabling innovation.</li> </ul>
Is there a Supplier(s) that is building a product based on this Spec?		
Will Supplier(s) have the product available for GENERAL AVAILABILITY within 120 days?		

## Appendix B: Meta Unique Requirements

### Performance:

#### IO.go Targets

This test measures how long the file system is blocked from writing/overwriting a file while a different file is deleted. The prep for this test is defined in PMP-9.

Requirement ID	Description
Iogo-1	All targets below shall be applicable to file-sizes <= 2048MB.
Iogo-2	Less than 4 file sizes total with latency outliers > 10ms
Iogo-3	No more than 2 latency outliers per file size
Iogo-4	No single latency outlier above 15ms
Iogo-5	The IO.go targets will be met with XFS, EXT-4 & BTRFS file-systems.

#### FileDelete & FileAppend Targets

This test measures how long the file system is blocked from appending to a file while a file is deleted. Ideally there should be no measurable io stalls reported by this tool. Run the benchmark script "filedelete\_whileappending.sh". Results will output to stdout.

Requirement ID	Description
FILE-DELETE-1	No measurable io stalls reported by this tool.
FILE-DELETE-2	Test should be executed for <i>xfs</i> , <i>ext4</i> and <i>btrfs</i> file-systems.

### Label:

#### Customer Defined Separator

Requirement ID	Description
LABEL-FB-1	The customer defined separator for labels printed on Meta SSD's shall be " <u>" (underscore).</u>
LABEL-FB-2	For M.2 the label shall be placed on the bottom side of the device as defined in the M.2 form-factor specification.

## Appendix C: Google Unique Requirements

### Security:

Requirement ID	Description
GOOG-SEC-1	Google to require FIPS 140-3 L1 or higher certification of cryptographic modules within the drive. This requirement is additive to SEC-7 and SEC-15.

### Label:

#### Customer Defined Separator

Requirement ID	Description
LABEL-GOOG-1	The customer defined separator for labels printed on Google SSD's shall be " " ( <i>space</i> ).
LABEL-GOOG-2	For M.2 the label shall be placed on the top side of the device, so they can be accessed and scanned without any disassembly.

### Endurance Targets:

#### Google requires the device to meet the following requirements

Requirement ID	Usable User Capacity <sup>2</sup> /Physical Capacity <sup>3</sup>	Minimum TBW @ 4k random writes
ENDU-GOOG-TGT-1 (No compression, 28% OP <sup>1</sup> )	100GB/128GB	Minimum endurance of 192 TBW
ENDU-GOOG-TGT-2 (No compression, 28% OP)	200GB/256GB	Minimum endurance of 384 TBW
ENDU-GOOG-TGT-3 (No compression, 28% OP)	400GB/512GB	Minimum endurance of 768 TBW
ENDU-GOOG-TGT-4 (No compression, 28% OP)	800GB/1024GB	Minimum endurance of 1536 TBW
ENDU-GOOG-TGT-5 (No compression, 28% OP)	1600GB/2048GB	Minimum endurance of 3072 TBW

<sup>1</sup> OP is calculated from usable user capacity and physical capacity in decimal GB only, which does not include the capacity from the difference between binary GiB and decimal GB.

<sup>2</sup> Usable user capacity is the host/user addressable capacity (TNVMCAP - UNVMCAP) in GB.

<sup>3</sup> Physical capacity is the total formattable NVM capacity (TNVMCAP) with 0% OP in GB.



## Appendix D: Guidance on Implementation of GUIDs

### **GUID#1 0xC46DD7920F1E4266A178D8AC78884365**

EXT-SMART-33 -little endian (since this is a NVMe log)

First Byte: offset 496: 65

Last Byte: offset 511:C4

### **GUID#2 0xAAB005F5135E4815AB8905BA8BE2BF3C**

HWREV-37 -little endian (since this is a NVMe log)

First Byte: offset 496: 3C

Last Byte: offset 511:AA

### **GUID#3 0xC035E2BC3B4A40E982E61BDEC28EFA72**

SMBus Byte 33:48 -big endian (since this is a SMBus ID)

First Byte: offset 33: C0

Last Byte: offset 48:72