NVMe™ Annual Update

Sponsored by NVM Express™ organization, the owner of NVMe™, NVMe-oF™ and NVMe-MI™ standards
Speakers

Peter Onufryk

Nick Adams
NVMe™ State of the Union

Peter Onufryk
NVM Express™, Inc.
120+ Companies Defining NVMe™ Together

Board of Directors
13 elected companies, stewards of the technology & driving processes
Chair: Amber Huffman

Marketing Workgroup
NVMexpress.org, webcasts, tradeshows, social media, and press
Co-Chairs: Jonmichael Hands and Cameron Brett

Technical Workgroup
NVMe™ Base and NVMe Over Fabrics
Chair: Peter Onufryk

Management Intf. Workgroup
NVMe Management
Chair: Peter Onufryk
Vice Chair: Austin Bolen

Interop (ICC) Workgroup
Interop & Conformance Testing in collaboration with UNH-IOL
Chair: Ryan Holmqvist

Co-Chairs: Jonmichael Hands and Cameron Brett

Chair: Amber Huffman

Chair: Peter Onufryk
Vice Chair: Austin Bolen

Chair: Ryan Holmqvist

Co-Chairs: Jonmichael Hands and Cameron Brett

Chair: Amber Huffman

Chair: Peter Onufryk
Vice Chair: Austin Bolen

Chair: Ryan Holmqvist
Ten Years of NVMe™
Strong Growth Across Segments

* Projections provided by Forward Insights Q2’19
NVMe™ is The New Language of Storage

<table>
<thead>
<tr>
<th>NVMe™ SSDs</th>
<th>23 Companies Shipping 96 Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVMe Severs</td>
<td>13 Companies Shipping 93 Models</td>
</tr>
<tr>
<td>NVMe AFAs</td>
<td>11 Companies Shipping 21 Models</td>
</tr>
<tr>
<td>NVMe Appliances</td>
<td>8 Companies Shipping 21 Models</td>
</tr>
<tr>
<td>NVMe-oF HBAs/NICs/RNICs</td>
<td>5 Companies Shipping 53 Models</td>
</tr>
<tr>
<td>NVMe-oF Accelerated Adapters</td>
<td>6 Companies Shipping</td>
</tr>
</tbody>
</table>

* Data provided by G2M Research
With Millions of Units Shipping

<table>
<thead>
<tr>
<th>K Units</th>
<th>2016</th>
<th>2017</th>
<th>2018</th>
<th>2019*</th>
<th>2020*</th>
<th>2021*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise</td>
<td>364</td>
<td>749</td>
<td>1,048</td>
<td>2,774</td>
<td>5,740</td>
<td>11,192</td>
</tr>
<tr>
<td>Cloud</td>
<td>2,051</td>
<td>3,861</td>
<td>10,231</td>
<td>17,338</td>
<td>25,891</td>
<td>31,050</td>
</tr>
<tr>
<td>Client</td>
<td>33,128</td>
<td>50,385</td>
<td>82,613</td>
<td>111,888</td>
<td>187,689</td>
<td>243,889</td>
</tr>
</tbody>
</table>

* Projections provided by Forward Insights Q2’19
### 2019 NVMe™ Deliverables

<table>
<thead>
<tr>
<th>NVMe™ Base Specification</th>
<th>NVMe Over Fabrics Specification</th>
<th>NVMe Management Interface Specification</th>
<th>NVMe Plugfest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Version</strong></td>
<td><strong>Date</strong></td>
<td><strong>Version</strong></td>
<td><strong>Date</strong></td>
</tr>
<tr>
<td>NVMe 1.4</td>
<td>6/10/2019</td>
<td>NVMe-oF™ 1.1</td>
<td>Plugfest #11</td>
</tr>
<tr>
<td>NVMe 1.3</td>
<td>5/1/2017</td>
<td>NVMe-oF 1.0</td>
<td>6/24/2019</td>
</tr>
<tr>
<td>NVMe 1.2</td>
<td>11/3/2014</td>
<td>NVMe-oF 1.0</td>
<td>Plugfest #10</td>
</tr>
<tr>
<td>NVMe 1.1</td>
<td>10/11/2012</td>
<td>NVMe-oF 1.0</td>
<td>11/12/2018</td>
</tr>
<tr>
<td>NVMe 1.0</td>
<td>5/14/2008</td>
<td>NVMe-oF 1.0</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NVMe-oF™ 1.1</td>
<td>Plugfest #1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45-day Review</td>
<td>5/13/2013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NVMe-oF 1.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NVMe-oF 1.0</td>
<td></td>
</tr>
</tbody>
</table>
NVMe™ Specification Roadmap

<table>
<thead>
<tr>
<th>Year</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NVMe**
- **NVMe™ 1.2.1 May’16**
  - Transport and protocol
  - RDMA binding

**NVMe-oF**
- **NVMe-oF™ 1.0 May’16**
  - Transport and protocol
  - RDMA binding

**NVMe-MI**
- **NVMe-MI™ 1.0 Nov’15**
  - Out-of-band management
  - Device discovery
  - Health & temp monitoring
  - Firmware Update

**NVMe 1.3 May’17**
- Sanitize
- Streams
- Virtualization

**NVMe 1.4 June’19**
- NVM Sets and IOD
- Persistent Event Log
- Multipathing Enhancements

**NVMe 1.5**
- SANitize
- Streams
- Virtualization

**NVMe-oF-1.1**
- Enhanced Discovery
- TCP Transport Binding

**NVMeoF-1.1**
- Enhanced Discovery
- TCP Transport Binding

**NVMe 2.0**
- Merged w/Fabrics
- Namespace Types
- Alternate Cmd Sets

**NVMe Transport Spec(s)**

**NVMe Cmd Set Spec(s)**

**In 45-day Review**
- NVMe Base Spec
- In-band Mechanism
- Storage Device Extension
Three New Specifications for 2019

NVMe™ 1.4

NVM Sets and IO Determinism enable better performance, isolation, and QoS for hyperscale data centers. Persistent event log provides robust drive history for issue triage and debug. Multipathing provides optimal path for a namespace in multi-controller topologies.

NVMe-oF™ 1.1

Enhanced Discovery for hosts to discover new NVMe devices. TCP Transport Binding NVMe/TCP enables efficient end-to-end NVMe operations with standard IP network with excellent performance and latency characteristics.

NVMe-MI™ 1.1

Enclosure Management enhances NVMe-MI for storage arrays for slot control, LED, and fans. In-band Mechanism opens up the NVMe-MI command set to standard NVMe driver (VPD, FRU). Storage Device Extension extends NVMe-MI to carrier cards and multiple controller devices.
The Evolution of NVMe™

Phase 1
- Unify PCIe® SSDs around a common interface
- Get an in-box driver in all major operating systems

Phase 2
- Scale NVMe™ over arbitrary fabrics

Phase 3
- Standardize NVM enabled storage innovations
- Expand NVMe into new use cases (e.g., Automotive and Storage Arrays)
NVMe™ Continues to Drive Simplicity in A World of Complexity

NVMe™ Base Specification (PCIe + Fabrics)

NVMe Architecture  Admin Command Set  NVMe Features

IO Determinism · Multipath · Sets & Endurance Groups · Namespace Types · Domains & Partitions
Security · Sanitize · Persistent Event Log · Telemetry · Power Management · and many others ....
Increasing the Rate of Innovation Together with Greater Quality

- Formalized sub-teams with publicly published calendars and minutes
- Technical proposal phases with clear entries and exits
- Document repository with revision history
- Integrated draft specification always up to date
- Weekly electronic ballots
NVMe™ has unified client, cloud, and enterprise storage around a common command set and interface

The growth in NVMe adoption continues to accelerate

The NVMe organization has put in place processes and initiatives to support the increased rate of innovation enabled by NVM and new use cases

NVMe remains true to its core principles of simplicity and efficiency as it enters its second decade
NVMe™ Base Spec 1.4 Features Overview

Nick Adams
NVMe™ 1.4 Technical Highlights
IO Determinism

- Service isolation region
- Increase Read IOPs and reduce tail latency
- Provides strict QoS profile
- Significantly improves P99 and P9999 for a well-behaved host

**IO Determinism & NVM Sets work together to provide improved QoS!**

No IO Determinism

- Workload A
- Workload B
- Workload C
- Workload D

4TB

With IO Determinism

- Workload A
- Workload B
- Workload C
- Workload D

1TB 1TB 1TB 1TB
Interface Optimizations

Created new **mechanisms** for Hosts to **optimize** their use of NVMe™ devices

- IO Performance & Endurance Hints
  - Exposes preferred Size, Granularity and Alignment for both Write and Deallocate to the Host
- Endurance Groups
  - Enables drives to be configured for endurance management across one or more NVM Sets
- Namespace Granularity
  - Enables optimal Namespace Size and Capacity by the Host at Namespace creation time
- Submission Queue Associations
  - Enables the Host to associate an IO Submission Queue it created with a specific NVM Set
- Verify Command
  - Verifies ability to read data without sending that data across the bus to the host

**Focused effort on optimizing the Host’s ability to improve the performance and endurance of NVMe devices**
Persistent Memory Regions

• Persistent Memory Region (PMR)
  • PCI memory space on the SSD exposed to the Host
  • May be used to store command data
  • Contents persist across power cycles, resets and disabling of the PMR

• Usage Models for PMR
  • Logs for SW RAID, EC & Databases
  • Journals for File Systems
  • Metadata
  • Staging area for data pre-processing
  • Network transactions
Enhanced Telemetry Capabilities

• The **Persistent** Event Log defines the features necessary to build a scaffolding that enables extensible debug infrastructure that is usable at scale

• Comprehensive set of events defined
  • Health Snapshot
  • Firmware Commits
  • Timestamp Changes
  • Power-on or Resets
  • Thermal Excursions
  • Vendor Specific
  • TCG-defined Events

• Hardware Errors

**Allows SSD customers to get consistent debug capabilities across vendors!**

**Allows SSD vendors an extensible framework for custom debug content!**
Administrative Controller

- New controller type for Enclosure Management
- **Requires minimal functionality** while enabling innovative management solutions
  - NVMe-MI™ in-band support (NVMe-MI Send & Receive)
  - Namespace Management
  - NVM Subsystem Reset
  - Unique PCI interface allows loading a custom driver
- Key Functional Aspects
  - Not required to support IO Queues or Command Set(s)
  - Reduced set of required Admin Commands, Features & Log Pages to support
Maturing the NVMe™ Infrastructure

• Working together to make the NVMe™ infrastructure **robust** and **mature** for the industry.
  • Addressed industry needs across a variety of areas
    • **Rebuild Assist** – Improved Recovery Scenarios
    • **UUIDs for Vendor-specific Info** – Mechanism to ensure vendor-specific events don’t collide
    • **Multi-host Shared Stream Write** – Improved multi-host functionality for Cloud & Enterprise
    • **Enhanced Command Retry** – Adds robustness in heavy load & other abnormal conditions
    • **Namespace Write Protect** – Enables finer granularity control over areas to Write Protect
  • Added clarifications and clean up to over 25 functional areas of the specification

**Contributions to the infrastructure come from broad industry collaboration. This highlights NVMe’s ability to come together as a community & deliver value for the entire industry!**
Refactoring the NVMe™ Family of Specifications
Driving Simplicity in a World of Complexity

Key Aspects Driving the Refactor

- Back to the core values… Fast, Simple, Scalable
- Foster areas of innovation while minimizing impact to broadly deployed solutions
- Creating an extensible spec infrastructure that will take the industry through the next phase of growth for NVMe!
NVMe™ spec family wasn’t structured for extensibility

Similar functions split between Base & Fabrics
PCle® transport integrated into Base
Command Sets not layered to enable extensibility

Need a new structure to enable innovation!
Optimizing the Specifications for Evolution

Transport Separation, Command Set Extensibility, Fabrics Base Integration

NVMe™ Base Specification

- Current Base with removal of PCIe® specifics
- Integrate NVMe-oF™ content
  - Identify, SGLs, Capsules, Discovery, ...

- Adds Fabrics concepts as core to NVMe
- Eliminates duplication in data structures
- Integration of NVMe and NVMe-oF base functions
- Separate command set specs
- Modular transport mapping layer, including PCIe
Questions?