Working Draft: NVMe to RF/SF Model Mapping

Richelle Ahlvers, Broadcom
SSM TWG Chair

2/26/2020
The information contained in this publication is subject to change without notice. The SNIA makes no warranty of any kind with regard to this specification, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The SNIA shall not be liable for errors contained herein or for incidental or consequential damages in connection with the furnishing, performance, or use of this specification.
Suggestions for revisions should be directed to http://www.snia.org/feedback/.
This proposal is a snapshot of work in progress by the NVMe Task Force of the SNIA Scalable Storage Management TWG, and does not represent the work of the SSM TWG at this point.

Contents are subject to change at any time.
Mapped Models and Documented Permutations

- **Device Model – NVMe**
  - Simple SSD
    - Default Endurance Group / Default Set
    - Single Endurance Group / Single Set
  - JBOFMode
  - Fabric Attach Array

- **Subsystem (Fabric) Model – NVMe-oF**
  - Fabric-attached subsystem presenting logical subsystem, controller, namespace, port and allowed host
  - Simple SSD with NVMe-oF Attach
    - Default Endurance Group / Default Set

- Add NVMe Domains concept

Yellow – TBD Mockups
Simple SSD Implementation: Default Endurance Group and Set (Not Implemented)
Simple SSD Implementation: Simple Endurance Group and Set

Redfish/Swordfish

- Physical Element Representation
- Storage
- StorageController
- StoragePool
- CapacitySource: StoragePool
- Volume
- Temperature Sensors
- Predictive Storage Failure / SMART Sensors

NVMe

- Connected to: (e.g.) Host PCIe Root Complex
- ComputerSystem: (Host Connection)
- Subsystem (NVMe)
- I/O Controller (NVMe)
- Single Endurance Group
- Single Set
- Namespace (NVMe)
Potential/Proposed New Properties

Storage (Subsystem)
- Identifiers – used for NQN / NGUID / etc. (unique ids)
- NVMe-oFANACharacteristics
- AccessState: Optimized, NonOptimized, Inaccessible, PersistentLoss

StorageController (I/O Controller)
- "RecommendedArbitrationBurstSize": "0",
- ComplexType: "ControllerMultiPathIoAndNamespaceSharingCapabilities":
- "PortCount": (enum values) "SinglePort", "MultiplePort",
- "ControllerCount": (enum values) "SingleController", "MultipleController"
- "ControllerAssociation": (enum values) "SRIOV", "PCIeOrFabric"
- "MaximumDataTransferSizeMPSMIN": "5",
- "NVMeVersion": "1.3",
- "OptionalAsyncEventSupported":enums:
  - EnduranceGroupEventAggregateLog
  - LBAStatusInformationNotices,
  - PredictableLatencyEventAggregateLogChangeNotices
  - AsymmetricNamespaceAccess
  - FirmwareActivationNotices
  - NamespaceAttributeNotices
  - ChangedNamespaceListLogPage
- ComplexType "NVMeControllerAttributes":
  - "ReportsUUIDList"
  - "SupportsSQAssociations"
  - "ReportsNamespaceGranularity"
  - "SupportsKeepAliveTimer"

StoragePool (for Endurance Group)
- "SupportsPredictableLatencyMode"
- "SupportsEnduranceGroups"
- "SupportsReadRecoveryLevels"
- "SupportsNVMSets"
- "SupportsExceedingPowerOfNonOperationalState"
- "Supports128BitHostId"

StoragePool (for Set)

Volume (Namespace)
- "MetadataTransferredAtEndOfDataLBA": "true"
- "NVMeVersion": "1.3",
- "Identifiers" - Add NGUID, IEEE OUI (subset of EUI) and additional formats for Namespace
Potential/Proposed New Properties

Storage (Subsystem)
  - NQNNotAllowedToCreate
    - DurableNameFormat
    - DurableName
New Objects / Schema: Domain

- Domains live off ServiceRoot
- Domain
  - DomainType == NVMe, Fault, Power, Cooling,
  - DomainMembers
  - Links -> AssociatedDomains
- NVMe Specific Properties
  - TotalDomainCapacityBytes
  - UnallocatedDomainCapacityBytes
  - MaximumCapacityPerEnduranceGroupBytes
  - AvailableFirmwareImages: []
    - Version, Vendor, NVMeDeviceType (sample properties, still TBD)
Work in progress – potential properties still to be added
Additional Properties to Add: SMART

- Predictive Storage Failure / SMART Sensors to StorageControllers
  - Drive uses: `<Property Name="FailurePredicted" Type="Edm.Boolean">`
  - StatusIndicator set to PredictiveFailureAnalysis
  - (other values OK/Fail/Rebuild/HotSpare)
  - In NVMe: Multiple SMART triggers in different objects

- Additional SMART triggers
  - PMR Unreliable
  - Power Backup failed
  - Media in Read only
  - Overall subsystem degraded
  - Spare capacity worn out
  - Temperature over critical warning threshold (not a permanent trigger?)

- Q: Do these go at the controller or namespace level? How to bundle vs distribute the info?

- ToDo: Offline investigation / proposal needed
Additional Properties to Add

- **Predictive:**
  - `<Property Name="PredictedMediaLifeLeftPercent" Type="Edm.Decimal">`

- **In NVMe:**
  - Has values at both EnduranceGroup and Subsystem Level
Additional Properties to Add

- Look at NVMe power states
  - RF Power is actual current power state: On/Off/PoweringOn/PoweringOff
- NVMe is selecting a profile for acceptable performance ranges supported – needs characterization as advanced power management functionality
  - Curtis: This is an internal set of properties, don’t see a reason to expose via RF/SF at this point.

- Use Redfish Basic Power State info for devices;
- ToDo: Add a new NVMe Power Profile equivalent model
Deferred to client input:

- Additional properties for hot plug (since not using drive model)
  - Q: What properties already exist, and what objects are they in?
  - Properties exist at the PCIe level.
  - Only NVMe level – no properties.
    - Request to shutdown may be used to anticipate or prep for a hot plug.
    - “Surprise removal” – shutdown notification.
    - Plug-in after surprise – dealt with differently than after a different kind of removal.
      Reporting write-cache discarded (due to re-insertion after surprise removal?). Others?

- Ask clients more details about what use cases, what properties they are looking for. Hot add / hot remove / prepare for hot plug?
Queue depth reporting -- add to controller

- Q: Queue size vs Queue depth?
- Queue depth not defined in the spec
- Queue size can be reported; negotiated value between host and target. Max value.
Additional Properties to Add

- **LED management**
  - Comparison with SES LED mgmt / model
  - Review RF enhanced LED management proposal (Slawek)
Additional Properties to Add

理财产品 / log Data

- Mapped to RF/SF schema / properties
  - SMART mapped properties
  - VPD Data (partial)

- Binary blob pass-through
  - Vendor-unique (pure blob)
    - Mapped into OEM Extensions
  - Binary blob with NVMe standard defined header
  - VPD Data (partial)

ToDo: Look at standard NVMe telemetry definition to see if / how it can be mapped to RF/SF