Flash Memory Summit

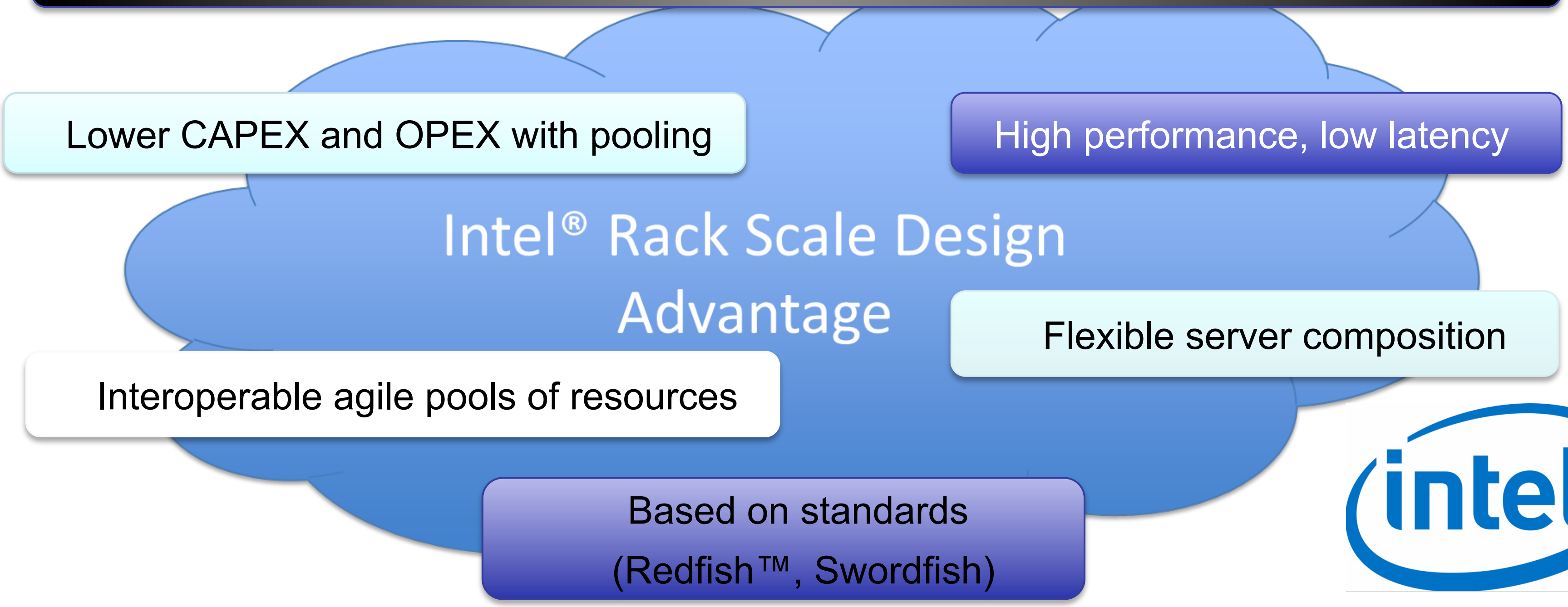
Intel® RSD and NVMe-over-Fabric

Sujoy Sen, Principal Engineer, Intel

Mohan Kumar, Fellow, Intel

Intel® RSD Pooled Technologies

The first industry-standard framework for managing Pooled Technologies



Lower CAPEX and OPEX with pooling

High performance, low latency

Intel® Rack Scale Design Advantage

Flexible server composition

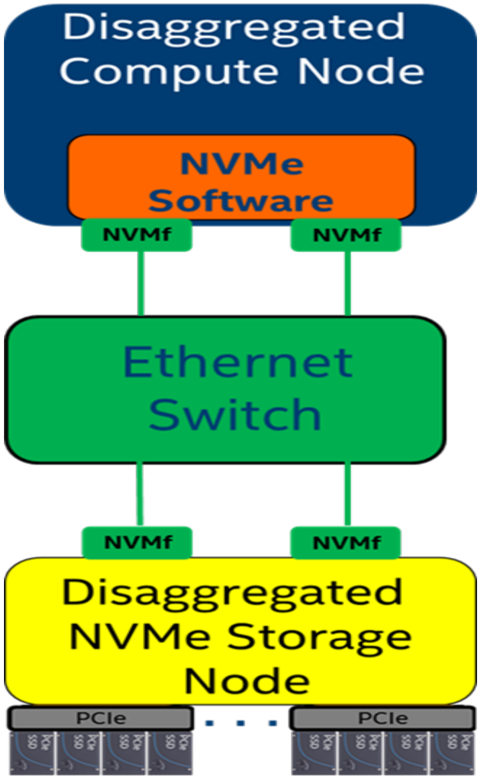
Interoperable agile pools of resources

Based on standards (Redfish™, Swordfish)

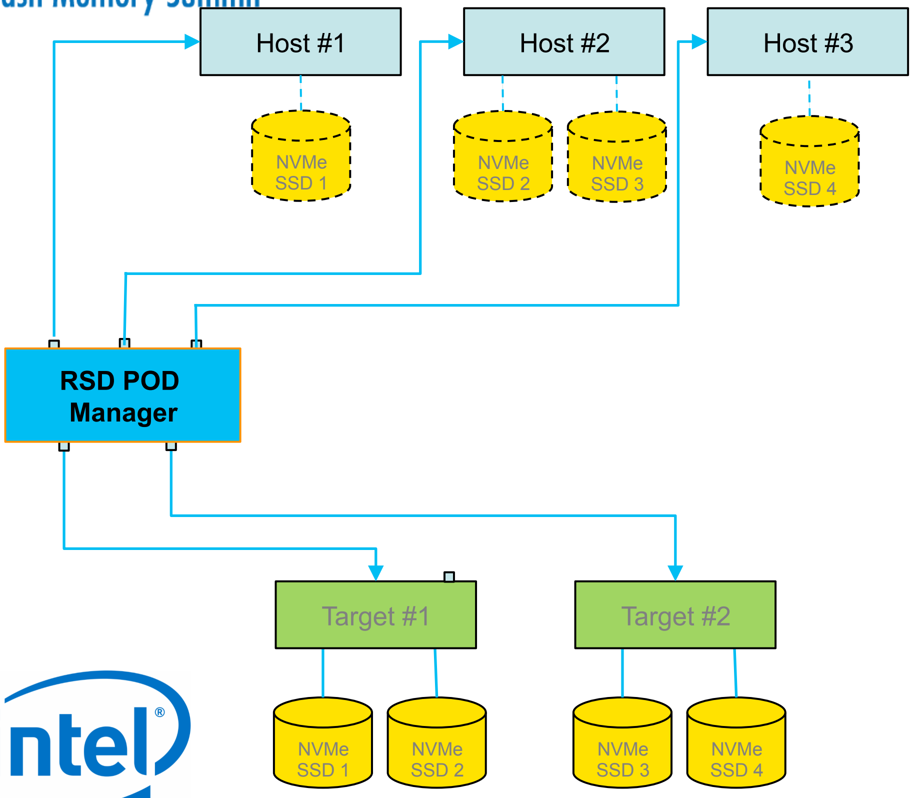
- More information available at: http://www.intel.com/IntelRSD

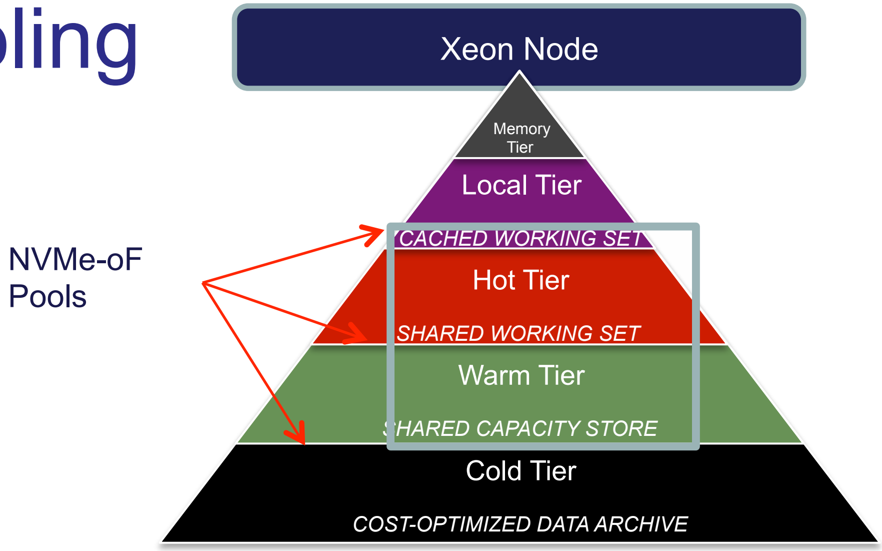
NVMe-over-Fabric Overview

- Export NVMe Drives to remote systems
- Appears as NVMe drive/namespace to remote application
- Transport NVMe Command sets over a Fabric
  - Low latency, efficient transport architecture
  - Defines use of RDMA as a transport



RSD Storage Pooling





- Support disaggregation of storage of varying performance using NVMe-oF
- Support cached working set, Hot Tier, Warm Tier of Storage solutions
- Dynamically Compose systems from storage pools

NVMe-oF is great but….

How do I provision the Target?

Where are my NVMe-oF Targets?

How do I create volumes?

What kind of drives do they have?

How do I compose a system?

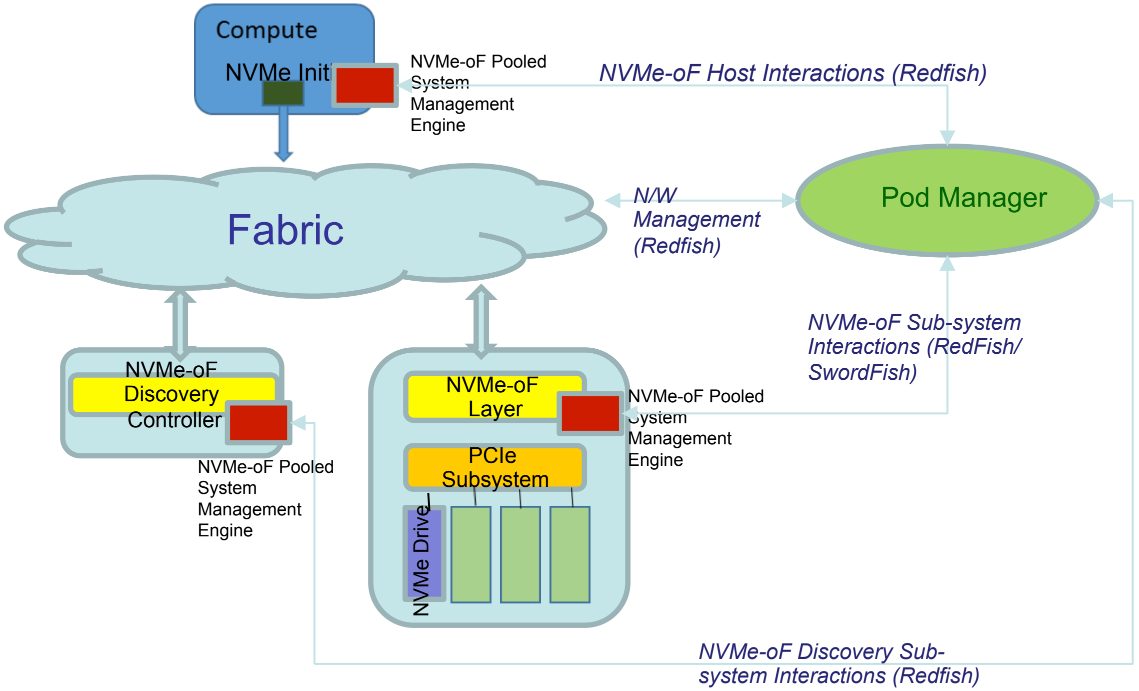
How do I provision storage for my compute nodes?

How do I know my Targets are healthy?

Enter Intel® *RSD…..*

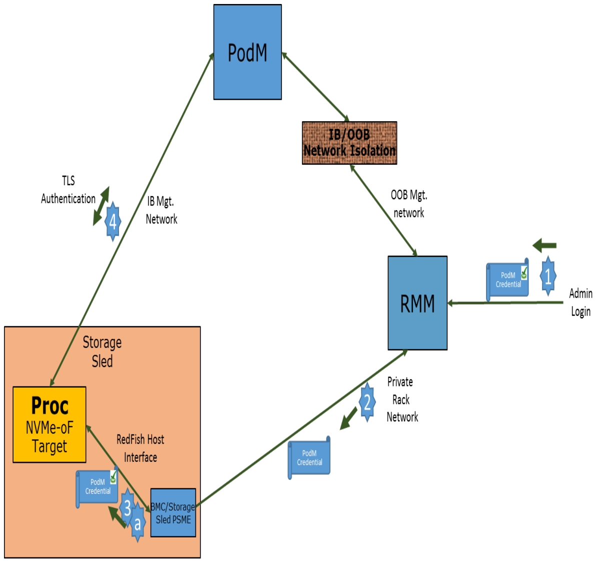
Provide management for

- Secure Discovery and Provisioning of NVMe-of Storage pools, NQN, Network
- Storage services configuration including Volume mgmt. and Access Control
- Telemetry


Compute
NVMe Init
NVMe-oF Pooled System Management Engine
NVMe-oF Host Interactions (Redfish)
Fabric
N/W Management (Redfish)
Pod Manager
NVMe-oF Sub-system Interactions (RedFish/ SwordFish)
NVMe-oF Discovery Controller
NVMe-oF Pooled System Management Engine
NVMe-oF Layer
NVMe-oF Pooled System Management Engine
PCIe Subsystem
NVMe Drive
NVMe-oF Discovery Sub-system Interactions (Redfish)

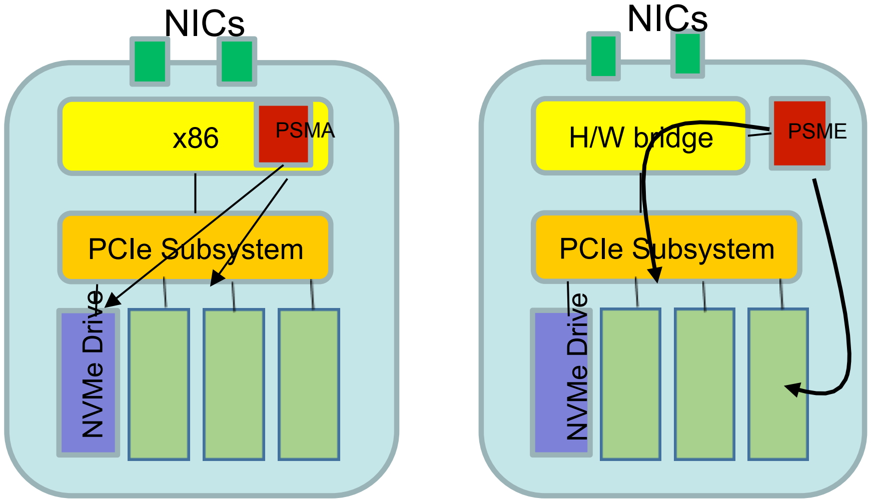
How do I Discover and Provision Targets?

- Isolated private OOB mgmt. network in each rack with a Rack Management Module (RMM)
- Admin provisions RMM with PodM and PSME credentials
- RMM distributes credentials
- PSME and PodM establish authenticated channel on the mgmt. network
- Target PSME reports its role and configuration
- Pod Manager provisions the Target with NVMe-oF parameters (NQNs, network configuration)



What kind of storage do I have?

- Target PSME is responsible for collecting drive and volume information
  - Enumerate NVMe drives, namespaces and volumes directly if a SW Target
  - Obtain from NVMe-oF Bridge or OOB on the platform
- Drive Information such as….
  - Capacity
  - Firmware version
  - Media capabilities
  - Health
  - Telemetry
- Report all the above to Pod Manager



How do I allocate storage to a system?

- Storage allocation can be done
  - via Composition of logical systems
  - via hot-add of new storage resources to an existing system
- Specify storage properties to PodM
  - capacity, performance, endurance
- Pod Manager chooses NVMe-oF targets that meet criteria
  - Create volume(s) on the target
  - Associate volume(s) with the host
  - Inform host (or discovery service) of new storage availability

How do I know my Targets are healthy?

- Targets report telemetry to PodM
- Pod Manager aggregates telemetry from multiple systems to provide rack-level health
- Targets report telemetry at various layers
  - NVMe-oF Protocol
  - Volume
  - NVMe Drive
  - Platform Telemetry

Manage NVMe-oF using RSD

How do I provision the Target? √

How do I create volumes? √

How do I compose a system? √

Where are my NVMe-oF Targets? √

What kind of drives do they have? √

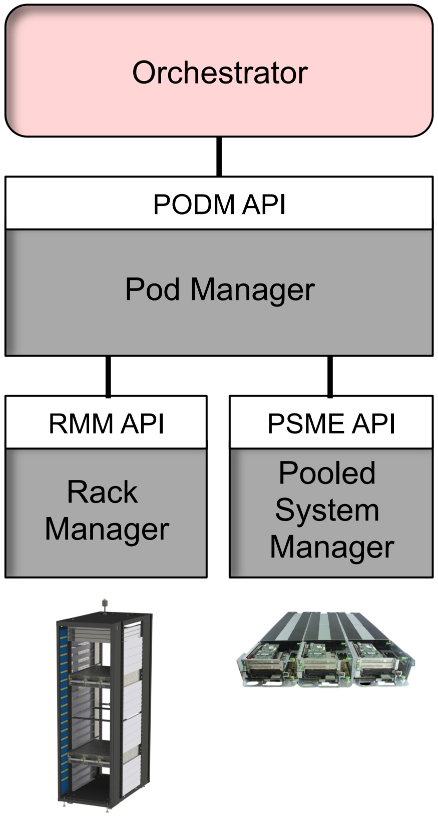
How do I provision storage for my compute nodes? √

How do I know my Targets are healthy? √

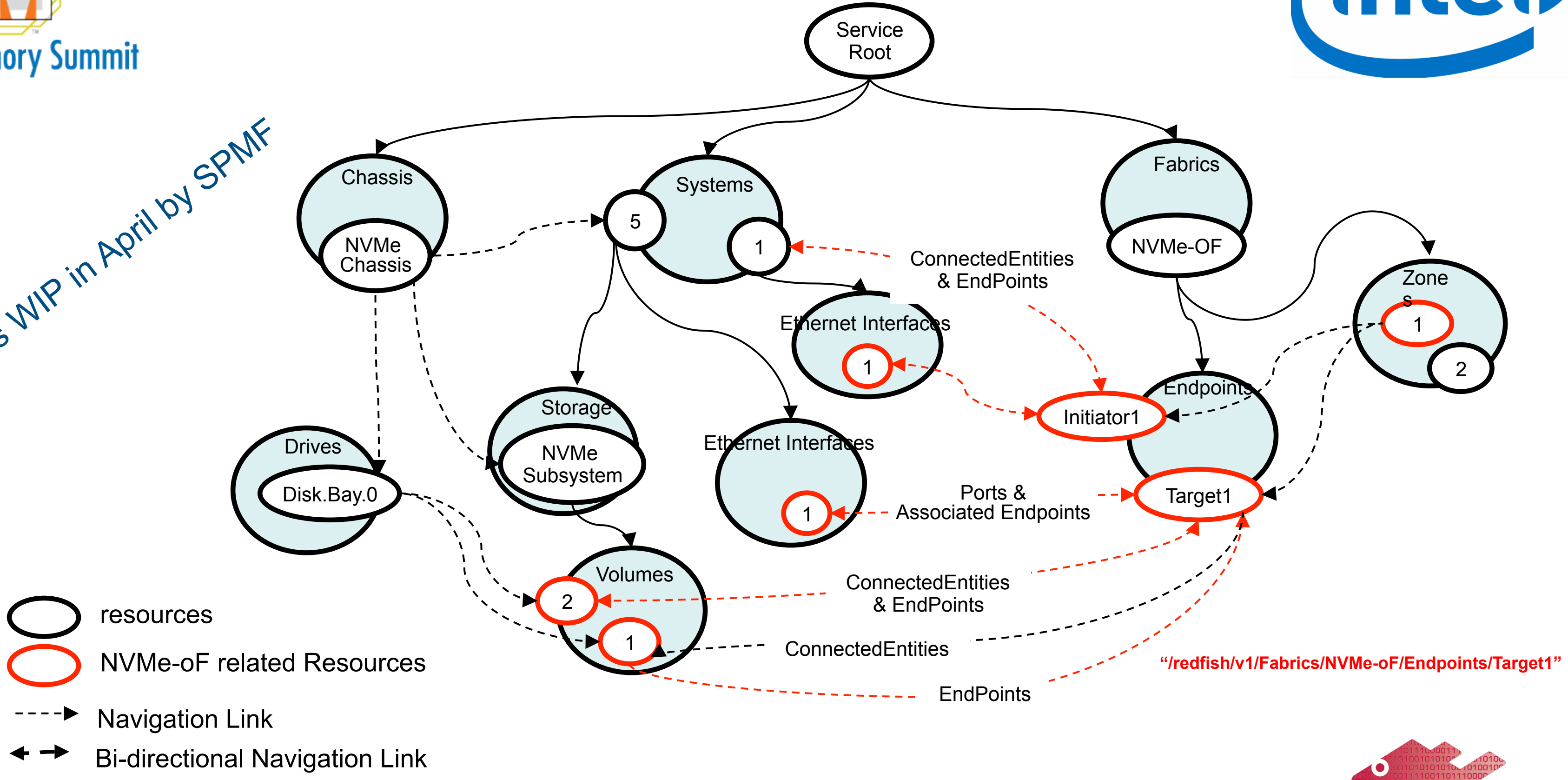
Closer look at NVMe-oF Management Model

RSD Manageability Standards

- Intel® Rack Scale Design manageability interfaces are based on Redfish™
  - Pod Manager (PODM) API
  - Rack Manager (RMM) API
  - Pooled System Manager (PSME) API
- Redfish™ has two parts
  - Interface specification (HTTP, JSON, OData)
  - Resource models for manageability
- Manageability Models
  - DMTF – physical platform, compute
  - SNIA – networked storage and Storage Service (Swordfish)
  - IETF – Ethernet Switches (YANG-to-Redfish)



NVMe-oF Redfish Common Fabric Model



Released as WIP in April by SPMF

Flash Memory Summit

Target End point JSON response

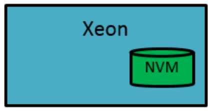
```
{
   "@odata.context": "/redfish/v1/$metadata#Endpoint.Endpoint",
   "@odata.id": "/redfish/v1/Fabrics/NVMe-oF/Endpoints/Target1",
   "@odata.type": "#Endpoint.v1_1_0.Endpoint",
   "Id": "Target1",
   "Name": "NVMe Drive 1 Volumes",
   "Description": "Two volumes created within the NVMe Drive in NVMeChassis 1 Bay 0",
   "EndpointProtocol": "NVMeOverFabrics",
   "Identifiers":  [ {
         "DurableName": "nqn.corp.com:nvme:nvm-subsys-sn-7642",
         "DurableNameFormat": "NQN"
   } ],
   "ConnectedEntities": [ {
         "EntityType": "Volume",
         "EntityRole": "Target",
         "EntityAccessMode": "Read",
         "EntityLink": { "@odata.id": "/redfish/v1/Systems/5/Storage/NVMeSubsystem/Volumes/1“ }
   } ],
   "Transports": [ {
         "TransportType": "Ethernet",
         "TransportProtocol": "RDMA",
         "TransportDetails": [ {
               "IPv4Address": { "Address": "10.3.5.132" },
               "Port": 13244,
               "RDMAType": "RoCEv2"
         } ]
   } ],
   "Links": {
      "Ports": [ { "@odata.id": "/redfish/v1/Systems/5/EthernetInterfaces/1 } ]
   }
}
```

DMTF

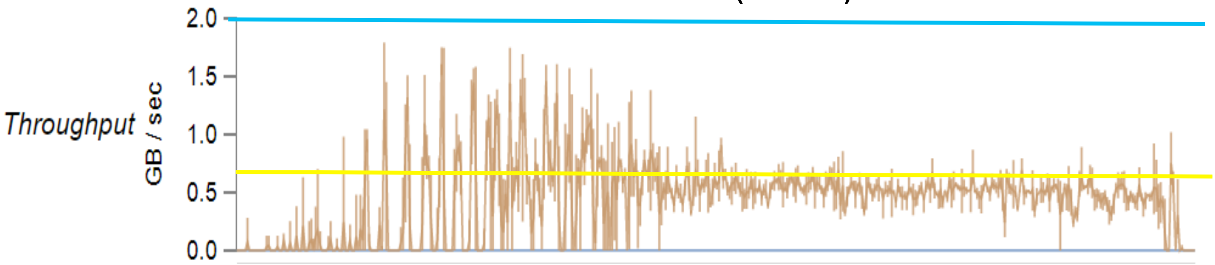
Now that we solved the NVMe-oF management problem…..

……..what are the benefits of Pooling?

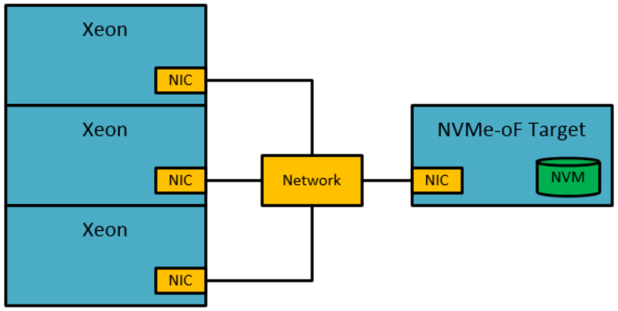
Leave no Bandwidth behind!

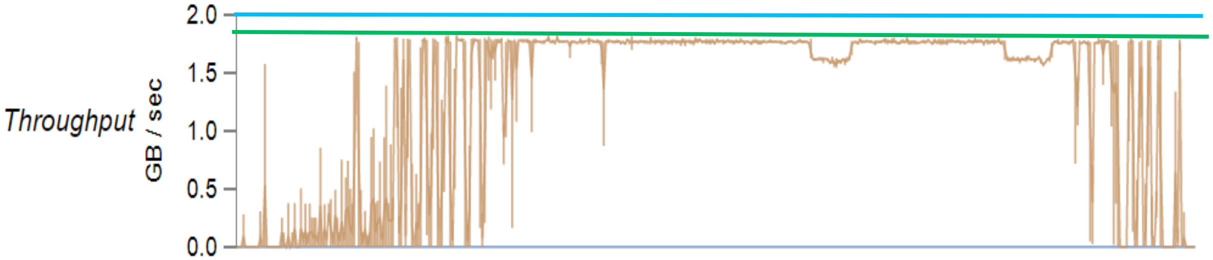


Somatic Mutation Finder (SMuFin) from BSC



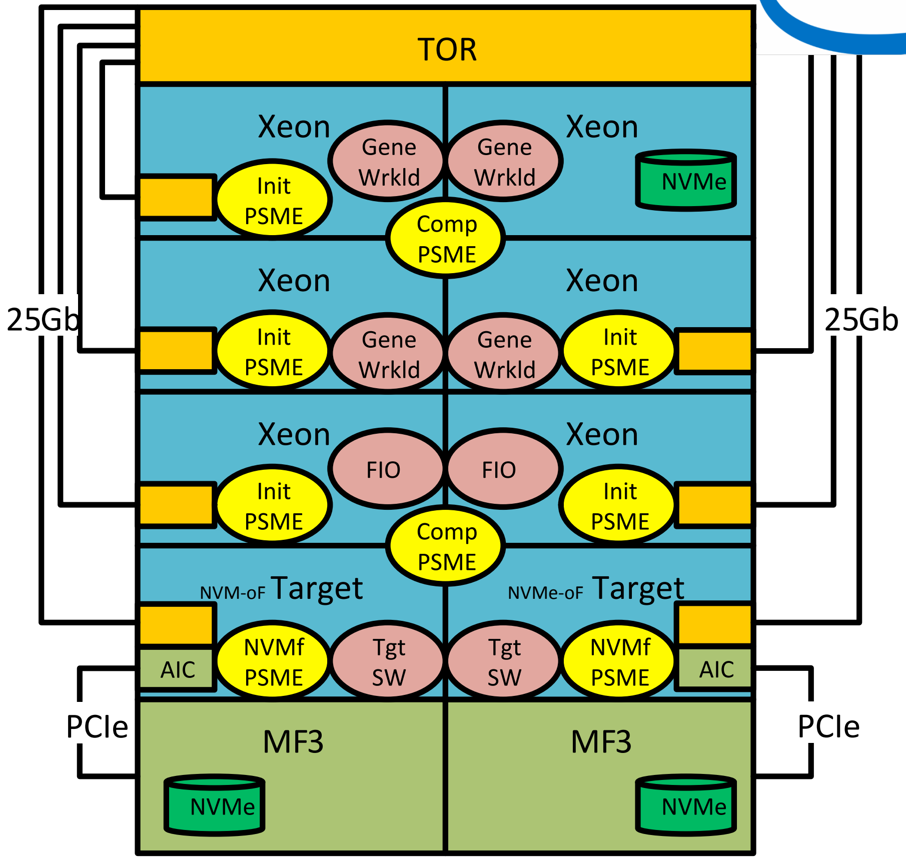
3X improved utilization with NVMe-oF





Demo at Intel Booth

- See RSD Management of NVMe-oF in action
- Witness the power of NVMe-oF pooling with SMuFin
- Booth 745A



Call to Action

- Drive increased efficiency in your Data Center using NVMe-over-Fabric to disaggregate and pool storage
- Manage NVMe-over-Fabric in the Data Center using Intel® RSD with standards-based management
- Get involved and provide feedback on NVMe-oF management model in DMTF and SNIA

Learn more at: http://intel.com/intelrsd

Flash Memory Summit

intel

Q&A

Notices & Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at intel.com.

No computer system can be absolutely secure.

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

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

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

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

BACKUP

Example JSON response

Simple properties

Complex properties

Subordinate resources

Associated resources

Actions

```
{
   "@odata.context": "/redfish/v1/$metadata#ComputerSystem.ComputerSystem",
   "@odata.id": "/redfish/v1/Systems/CS_1",
   "Id": "CS_1",
   "Name": "My Computer System",
   "SystemType": "Physical",
   "AssetTag": "free form asset tag",
   "Manufacturer": "Manufacturer Name",
   "Model": "Model Name",
   "SerialNumber": "2M220100SL",
   "PartNumber": "",
   "Description": "Description of server",
   "UUID": "00000000-0000-0000-0000-000000000000",
   "HostName": "web-srv344",
   "IndicatorLED": "Off",
   "PowerState": "On",
   "BiosVersion": "P79 v1.00 (09/20/2013)",
   "Status": { "State": "Enabled", "Health": "OK", "HealthRollup": "OK" },
   "Boot":                { . . . },
   "ProcessorSummary": { . . . },
   "MemorySummary":  { . . . },
   "TrustedModules":    [ { . . . } ],
   "Processors":                    { "@odata.id": "/redfish/v1/Systems/CS_1/Processors" },
   "Memory":              { "@odata.id": "/redfish/v1/Systems/CS_1/Memory" },
   "EthernetInterfaces":  { "@odata.id": "/redfish/v1/Systems/CS_1/EthernetInterfaces" },
   "SimpleStorage":       { "@odata.id": "/redfish/v1/Systems/CS_1/SimpleStorage },
   "LogServices":         { "@odata.id": "/redfish/v1/Systems/CS_1/LogServices" },
   "SecureBoot":          { "@odata.id": "/redfish/v1/Systems/CS_1/SecureBoot" },
   "Bios":                { "@odata.id": "/redfish/v1/Systems/CS_1/Bios" },
   "PCIeDevices":         [ {"@odata.id": "/redfish/v1/Chassis/CS_1/PCIeDevices/NIC"} ],
   "PCIeFunctions":       [ {"@odata.id": "/redfish/v1/Chassis/CS_1/PCIeDevices/NIC/Functions/1" }],
   "Links": {
      "Chassis":          [ { "@odata.id": "/redfish/v1/Chassis/Ch_1" } ],
      "ManagedBy":        [ { "@odata.id": "/redfish/v1/Managers/Mgr_1" } ],
      "Endpoints":        [ { "@odata.id": "/redfish/v1/Fabrics/PCIe/Endpoints/HostRootComplex1" } ],
   },
   "Actions": {
      "#ComputerSystem.Reset": {
         "target": "/redfish/v1/Systems/CS_1/Actions/ComputerSystem.Reset",
             "@Redfish.ActionInfo": "/redfish/v1/Systems/CS_1/ResetActionInfo"
      }
   }
}
```