Scalable High Performance on NVMe/FC

Jayamohan Kallickal
Broadcom
Legal Disclaimer

All or some of the products detailed in this presentation may still be under development and certain specifications, including but not limited to, release dates, prices, and product features, may change. The products may not function as intended and a production version of the products may never be released. Even if a production version is released, it may be materially different from the pre-release version discussed in this presentation.

Nothing in this presentation shall be deemed to create a warranty of any kind, either express or implied, statutory or otherwise, including but not limited to, any implied warranties of merchantability, fitness for a particular purpose, or non-infringement of third-party rights with respect to any products and services referenced herein.

Broadcom, the pulse logo, Connecting everything, Avago Technologies, Avago, the A logo, Brocade, Emulex, ExpressLane, LightPulse, and OneCommand are among the trademarks of Broadcom and/or its affiliates in the United States, certain other countries, and/or the EU. Other marks may belong to third parties.
Safe Harbor Statement

This presentation contains forward-looking statements (including within the meaning of Section 21E of the United States Securities Exchange Act of 1934, as amended, and Section 27A of the United States Securities Act of 1933, as amended) concerning Broadcom. These statements include, but are not limited to, statements that address our expected future business and financial performance and other statements identified by words such as “will”, “expect”, “believe”, “anticipate”, “estimate”, “should”, “intend”, “plan”, “potential”, “predict”, “project”, “aim”, and similar words, phrases or expressions. These forward-looking statements are based on current expectations and beliefs of the management of Broadcom, as well as assumptions made by, and information currently available to, such management, current market trends and market conditions and involve risks and uncertainties, many of which are outside the Company’s and management’s control, and which may cause actual results to differ materially from those contained in forward-looking statements. Accordingly, you should not place undue reliance on such statements.

Particular uncertainties that could materially affect future results include risks associated with: our acquisition of CA, including (1) potential difficulties in employee retention, (2) unexpected costs, charges or expenses, and (3) our ability to successfully integrate CA’s business and achieve the anticipated benefits of the transaction; any loss of our significant customers and fluctuations in the timing and volume of significant customer demand; our dependence on contract manufacturing and outsourced supply chain; our dependency on a limited number of suppliers; any other acquisitions we may make, including integrating acquired companies with our existing businesses and our ability to achieve the benefits, growth prospects and synergies expected by such acquisitions; our ability to accurately estimate customers’ demand and adjust our manufacturing and supply chain accordingly; our significant indebtedness, including the additional indebtedness that we incurred in connection with the CA acquisition and the need to generate sufficient cash flows to service and repay such debt; dependence on and risks associated with distributors of our products; dependence on senior management; quarterly and annual fluctuations in operating results; global economic conditions and concerns; the amount and frequency of our stock repurchases; cyclicality in the semiconductor industry or in our target markets; our competitive performance and ability to continue achieving design wins with our customers, as well as the timing of any design wins; prolonged disruptions of our or our contract manufacturers’ manufacturing facilities or other significant operations; our ability to improve our manufacturing efficiency and quality; our dependence on outsourced service providers for certain key business services and their ability to execute to our requirements; our ability to maintain or improve gross margin; our ability to protect our intellectual property and the unpredictability of any associated litigation expenses; compatibility of our software products with operating environments, platforms or third-party products; our ability to enter into satisfactory software license agreements; sales to our government clients; availability of third party software used in our products; use of open source code sources in our products; any expenses or reputational damage associated with resolving customer product warranty and indemnification claims; our ability to sell to new types of customers and to keep pace with technological advances; market acceptance of the end products into which our products are designed; our ability to protect against a breach of security systems; fluctuations in foreign exchange rates; our overall cash tax costs, legislation that may impact our overall cash tax costs and our ability to maintain tax concessions in certain jurisdictions; and other events and trends on a national, regional and global scale, including those of a political, economic, business, competitive and regulatory nature.

Our filings with the SEC, which you may obtain for free at the SEC’s website at http://www.sec.gov, discuss some of the important risk factors that may affect our business, results of operations and financial condition. Actual results may vary from the estimates provided. We undertake no intent or obligation to publicly update or revise any of the estimates and other forward-looking statements made in this presentation, whether as a result of new information, future events or otherwise, except as required by law.

2019 Storage Developer Conference. © Broadcom Inc. All Rights Reserved.
Agenda

1. NVMe Specification and Roadmap
2. NVMe/FC Functionality
3. What’s New?
4. Advantages
5. Techniques to improve Performance
6. Results
NVMe Specifications and Feature Roadmap
NVM Express™ Organization

http://www.nvmexpress.org/

Drivers for Windows, Linux, VMware, Solaris, FreeBSD and UEFI

NVM Express, Inc.
Includes more than 75 firms from across the industry

Promoter Group
Led by 13 elected companies

Source: NVM Express™ organization
**NVMe™ Specification Roadmap**

<table>
<thead>
<tr>
<th>Year</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVMe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVMe-oF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVMe-MI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

**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 June’19**
- Enclosure Management
- In-band Mechanism
- Storage Device Extension

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

**NVMeoF-1.1**
- In 45-day Review

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

**NVMe Transport Spec(s)**
- Released NVMe specification
- Planned release
- Current release

Source: NVM Express™ organization
Three New NVM Express 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.

Source: NVM Express™ organization
NVMe/FC Functionality
NVMe Over FC Protocol Layers

Source: NVMe Over Fabrics Specification
NVM ExpressTM over Fabrics Revision 1.0a
1. FC Name Server Points to NVMe Discovery Controller(s)

2. FC-NVMe Initiator Connects to NVMe Discovery Controller(s)

3. FC-NVMe Initiator connects to NVMe Subsystem(s) to begin data transfer
Queue Creation Using Fabric Connect Command

- Create a fabric-dependent transport connection
- Send a Command Capsule with Fabric Connect Operation
- Send Authentication Fabrics command.
- Admin Queue or I/O Queue Ready for NVMe Commands

Source: NVMe Over Fabrics Specification
NVM ExpressTM over Fabrics Revision 1.0a
The NVMe over FC association is created when the NVMe host makes a request to establish relationship to a particular controller of an NVMe subsystem.

Initiator port initiates the creation of the association by transmitting a Create Association NVMe_LS request to the target NVMe_Port.
- Request Identifies the NVMe-host by
  - Host NQN and Host ID
- Request Identifies the NVMe-Subsystem by
  - Subsystem NQN and Controller-ID

If the target NVMe_Port and NVM subsystem allow the communication relationship to be created, the target NVMe_Port transmits a Create_Association accept payload to the initiator NVMe_Port. The payload contains
- Association Identifier
- Connection Identifier
NVMe Over FC - Connection

- An NVMe over FC connection corresponding to an I/O Queue is created when the NVMe host makes a request to establish the transport connection for an I/O Queue for a particular controller.

- Initiator port initiates the creation of the transport connection by transmitting a Create I/O Connection NVMe_LS request from the initiator NVMe_Port to the target NVMe_Port.

- If the target NVMe_Port and NVMe controller accept the request, the target NVMe_Port transmits a Create I/O Connection accept payload to the initiator NVMe_Port. The payload contains
  - Connection Identifier
What’s New?
NVMe™ 1.4 – I/O Determinism

Service isolation region

- Enables hosts to treat an SSD as many small sub-SSDs
  - process host threads I/O independently in small sub-SSDs without blocking from other thread I/Os
- Can reduce average read latency significantly
  - for higher performance and for better Quality of Service (QoS)
  - due to the parallel execution of I/Os without any conflict to the media
NVMe™ 1.4 – Multipathing and Namespace Sharing

Technical Term: Asymmetric Namespace Access (ANA)

- **NVMe™ Multipathing I/O** refers to two or more completely independent PCI Express paths between a single host and a namespace.

- **Namespace sharing** enables two or more hosts to access a common shared namespace using different NVM Express controllers.

Both multi-path I/O and namespace sharing require that the NVM subsystem contain two or more controllers.

Source: NVM Express™ organization
FC-NVMe-2 Sequence Level Error Recovery

Detect and recover errors before they reach the protocol layer

- **FC-NVMe-2**
  - adds Sequence Level Error Recovery at the Fibre Channel transport level to avoid connection and association termination
  - allows errors (missing or corrupt frames) to be detected and recovered at FC frame level at the transport layer before the protocol layers knows
  - frames that are missing will timeout and would be retransmitted (HBA gets asked to resend the sequence)
    - new FC Basic Link Service for fast recovery
  - Upper layers does not know anything happened

- Sequence level recovery might also be implemented for FCP at a later point
Advantages
For Enterprise Storage, the choice is clear
- Most of today’s AFAs are deployed on Fibre Channel
- Robust fabric services
- Deterministic fabric behavior based on credit-based flow control

Ethernet fabrics Options
- RoCEv2 requires “lossless” Ethernet and “RNICs”
- iWARP has limited vendors supporting
- NVMe over FCoE comes along with FC, but niche usage only. Requires lossless Ethernet
- NVMe over TCP is recently standardized, just coming out

Source: Brocade
Fibre Channel

- Not RDMA-based
- Fibre Channel was already a lossless, topology-agnostic fabric, NVMe is just a new “upper layer protocol”
- Natural extension to leverage fabrics for shared storage arrays
  - FC being the predominant fabric for storage (70% of flash storage is FC)
- Will run on exiting GEN5/6 fabrics
- Broad vendor support
  - Broadcom(Emulex), Marvell (QLogic), Brocade, CISCO

Source: Broadcom
Ecosystem and Interop Testing

Test Track 5  GEN6, GEN5 FC and FC-NVMe Dual Fabric HA Large Fabric Build

Implement pair wise zone for each I-T

Concurrent FC and FC-NVMe outlined in black

Analyzers inserted inline between switches and T328 and VIAVI cascaded; Initiator and Target separation enables ISL visibility to all I-T traffic.

Performance Improvement Techniques
Server Hardware Architecture

Processor 1

PCIe RC

HBA

Processor 2

PCIe RC

UPI
Processor Architecture

L3 Cache

PCI Root Complex

UPI Agent

DDR Channels

UPI Channels
CPU Affinity

**EQ Per Core**

- Per-CPU WQ/CQ (a “Hardware Queue”)
- Interrupt vector/EQ per CPU
- Interrupt vector/EQ per CPU

**EQ Per Socket**

- One Interrupt Vector/EQ per Socket
Sharing Adapter Resources

- FC exchanges
  - Adapter has a fixed number
  - Needed for SCSI and NVMe
  - Exchange assigned to each IO for the duration of the IO
  - Partitioning per CPU resulted in few resources per CPU, thus lots of IO “busying”
  - Solve by pools per Hardware Queue with resources migrating between Hardware Queues on as-needed basis
Interrupt Handling

- Interrupt Handling:
  - Disassociate EQ from CQ
    - EQ must be serviced by ISR
    - CQ serviced by Independent Thread
- CQ Processing Tenancy
  - How much work you do while in the thread
  - Large limits put in. If limit reached and work remains, re-schedule
- Periodic Queue Pointer Updates to Hardware
- Interrupt Rate Management
  - Interrupt re-enablement
    - Use architecture-specific re-arming to reduce interrupt rate
  - Interrupt delay largely left “immediate”
  - Exception: CPU shared by Interrupt Vectors or HWQs
Results
NVMe Lancer G6 & Prism 2-ports IOPs Trend

NVMe SLES 12 SP3 Lancer G6 & Prism IOps for 12.0.x to 12.4.x with Prism target

- Emulex G7 2 Port 512 byte
- Emulex G6 2 Port 512 Byte
FCP vs. NVMe/FC Performance

Single Node A700s ONTAP 9.4 Oracle 12c SLOB FCP vs. NVMe/FC (75% read / 25% update)

https://www.suse.com/media/presentation/TUT1083_NVMe_Over_Fabrics_High_Performance_IO_With_Existing_Fabrics.pdf
TPC-C Microsoft SQL 2017 for Linux FCP vs NVMe

Transactions Per Minute

<table>
<thead>
<tr>
<th>TPM</th>
<th>LPe32002_FCP</th>
<th>LPe32002_NVMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>876001</td>
<td>143934</td>
</tr>
<tr>
<td>250000</td>
<td>500000</td>
<td>750000</td>
</tr>
<tr>
<td>1000000</td>
<td>1250000</td>
<td>1500000</td>
</tr>
<tr>
<td>1500000</td>
<td>1750000</td>
<td>2000000</td>
</tr>
</tbody>
</table>

New Orders Per Minute

<table>
<thead>
<tr>
<th>NOPM</th>
<th>LPe32002_FCP</th>
<th>LPe32002_NVMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19061</td>
<td>31219</td>
</tr>
<tr>
<td>500000</td>
<td>600000</td>
<td>800000</td>
</tr>
</tbody>
</table>

IO Wait%

<table>
<thead>
<tr>
<th>IO WAIT</th>
<th>LPe32002_FCP</th>
<th>LPe32002_NVMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5.12</td>
<td>1.68</td>
</tr>
<tr>
<td>2</td>
<td>4.52</td>
<td>2.14</td>
</tr>
<tr>
<td>4</td>
<td>3.93</td>
<td>3.78</td>
</tr>
<tr>
<td>6</td>
<td>3.54</td>
<td>4.61</td>
</tr>
</tbody>
</table>

CPU Utilization%

<table>
<thead>
<tr>
<th>CPU UTIL %</th>
<th>LPe32002_FCP</th>
<th>LPe32002_NVMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14.89</td>
<td>21.39</td>
</tr>
<tr>
<td>20</td>
<td>17.92</td>
<td>25.53</td>
</tr>
<tr>
<td>40</td>
<td>21.39</td>
<td>28.94</td>
</tr>
</tbody>
</table>

CPU Efficiency

<table>
<thead>
<tr>
<th>TPM PER CPU%</th>
<th>LPe32002_FCP</th>
<th>LPe32002_NVMe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>58831.5</td>
<td>67290.5</td>
</tr>
<tr>
<td>20</td>
<td>56000.00</td>
<td>64000.00</td>
</tr>
<tr>
<td>40</td>
<td>60000.00</td>
<td>68000.00</td>
</tr>
<tr>
<td>60</td>
<td>62000.00</td>
<td>70000.00</td>
</tr>
<tr>
<td>80</td>
<td>64000.00</td>
<td>78000.00</td>
</tr>
<tr>
<td>100</td>
<td>66000.00</td>
<td>86000.00</td>
</tr>
</tbody>
</table>
Oracle OLTP TPC-C Performance

- **Transactions Per Minute**
- **New Orders Per Minute**
- **IO CPU (%SYS)**
- **TPM per CPU**
- **CPU %IOWAIT**

Comparison between NVMe over FC and FC/SCSI:
NVMe-Fabrics in vSphere
ESXi

Shoby Cherian
VMware Inc
Disclaimer

This presentation may contain product features or functionality that are currently under development.

This overview of new technology represents no commitment from VMware to deliver these features in any generally available product.

Features are subject to change, and must not be included in contracts, purchase orders, or sales agreements of any kind.

Technical feasibility and market demand will affect final delivery.

Pricing and packaging for any new features/functionality/technology discussed or presented, have not been determined.
Agenda

- NVMe Next-Gen in ESXi
- New NVMe Driver Model
- IO Flow
- Management / UI
- Leveraging NVMe-oF - POCs
NVMe-oF – Pick Your Transport!

NVMe Fabrics Core

- NVMe FC
  - NVMe FC Driver
  - FC

- NVMe RDMA (RoCE v2 & iWARP)
  - RDMA Stack
  - RDMA Driver

- NVMe TCP
  - TCP/IP Stack
  - Ethernet Driver

- Ethernet
Transition to NVMe Next-Gen Stack in ESXi

Current

- NVMe PCIe Driver
  - SCSI/NVMe Translation
  - Controller Configuration
  - Namespace Discovery
  - PCIe Specific Things

Next-Gen

- NVMe PCIe Driver 1
- NVMe PCIe Driver 2
- NVMe-oF Layer
  - NVMe RDMA Driver
  - FC-NVMe Driver
  - NVMe TCP Driver
- RDMA Stack
  - RoCE Driver
  - iWARP Driver
- FC Driver
- TCP/IP Stack

Storage Stack
- NVMe-SCSI Translation
- NVMe-PSA
- NVMe Common Layer

Next Generation Storage Stack
- ESXCLI
- vmkctl
- plugins

Kernel Space
User Space
New NVMe Driver Model

- Implements most of common functions defined in NVMe base specification and NVMe-oF specification that are needed for VMware ESXi.
- Common user interface for NVMe device management.
- Transport agnostic driver interface for PCIe and Fabrics based NVMe driver development.
- Supports auto discovery/connect of NVMe-oF controllers for NVMe/FC.
- Supports persistent connections of NVMe-oF controllers.
- Supports legacy SCSI based storage stack and new NVMe native storage stack.
- Much simpler way of implementing NVMe transport device driver.
NVMe-oF – Control and Data Flows
## vSphere UI – FC-NVMe

### Storage Adapters

<table>
<thead>
<tr>
<th>Adapter</th>
<th>Type</th>
<th>Status</th>
<th>Identifier</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>vmhba32</td>
<td>PCIE</td>
<td>Unknown</td>
<td>--</td>
<td>0</td>
</tr>
<tr>
<td>vmhba33</td>
<td>Fibre Channel</td>
<td>Unknown</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>vmhba1</td>
<td>Fibre Channel</td>
<td>Online</td>
<td>20:00:00.90:fe:dd:59.e6 10:00:00.90:fe:dd:59.e6</td>
<td>0</td>
</tr>
</tbody>
</table>

### Properties

<table>
<thead>
<tr>
<th>Name</th>
<th>Subsystem</th>
<th>Transport Type</th>
<th>Fuse Support</th>
<th>Model</th>
<th>Firmware Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>ngn.1992-08.com.netapp.san.c0a5fc3d0681f994a00a098dcd37a.subsystem.prme_fcoe_007_subsystem#201900a098dcd4a7c-201a00a098dcd4a7c</td>
<td>ngn.1992-08.com.netapp.san.c0a5fc3d0681f994a00a098dcd37a.subsystem.prme_fcoe_007_subsystem#201900a098dcd4a7c-201b00a098dcd4a7c</td>
<td>fibreChannel</td>
<td>true</td>
<td>NetApp ONTAP Controller</td>
<td>FFFFFFFF</td>
</tr>
<tr>
<td>ngn.1992-08.com.netapp.san.c0a5fc3d0681f994a00a098dcd37a.subsystem.prme_fcoe_007_subsystem#201900a098dcd4a7c-201a00a098dcd4a7c</td>
<td>ngn.1992-08.com.netapp.san.c0a5fc3d0681f994a00a098dcd37a.subsystem.prme_fcoe_007_subsystem#201900a098dcd4a7c-201b00a098dcd4a7c</td>
<td>fibreChannel</td>
<td>true</td>
<td>NetApp ONTAP Controller</td>
<td>FFFFFFFF</td>
</tr>
</tbody>
</table>
### vSphere UI – NVMe RDMA

![Screen Shot of vSphere UI showing NVMe RDMA configurations](image)

#### Storage Adapters

<table>
<thead>
<tr>
<th>Adapter</th>
<th>Type</th>
<th>Status</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>vmhba1</td>
<td>SCSI</td>
<td>Unknown</td>
<td>--</td>
</tr>
<tr>
<td>vmhba33</td>
<td>RDMA</td>
<td>Unknown</td>
<td>--</td>
</tr>
</tbody>
</table>

#### Properties

- **Target**: 1
- **Devices**: 1
- **Paths**: 1

---

2019 Storage Developer Conference. © VMware Inc. All Rights Reserved.
```
[root@localhost:~] esxcli nvme adapter list
<table>
<thead>
<tr>
<th>Adapter</th>
<th>Adapter Qualified Name</th>
<th>Transport Type</th>
<th>Driver</th>
<th>Associated Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>vmhba32</td>
<td>aqn:nvme_pcie:nqn.2014-08.org.nvmeexpress15ad15adVMware_NVMe-0000VMware_Virtual_NVMe_Disk</td>
<td>PCIe</td>
<td>nvme_pcie</td>
<td></td>
</tr>
<tr>
<td>vmhba33</td>
<td>aqn:brcmnvmeffc:100000090fa94892f</td>
<td>FC</td>
<td>brcmnvmeffc</td>
<td></td>
</tr>
<tr>
<td>vmhba34</td>
<td>aqn:brcmnvmeffc:100000090fa948930</td>
<td>FC</td>
<td>brcmnvmeffc</td>
<td></td>
</tr>
<tr>
<td>vmhba35</td>
<td>aqn:nvmerdma:24-8a-07-b4-34-32</td>
<td>RDMA</td>
<td>nvmerdma</td>
<td>vmrdma0, vmnic0</td>
</tr>
</tbody>
</table>

[root@localhost:~] esxcli nvme controller list
<table>
<thead>
<tr>
<th>Name</th>
<th>Controller Number</th>
<th>Adapter</th>
<th>Transport Type</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td>nqn.2014-08.org.nvmeexpress_15ad_VMware_Virtual_NVMe_Disk</td>
<td>256</td>
<td>vmhba32</td>
<td>PCIe</td>
<td>true</td>
</tr>
<tr>
<td>nqn.2014-08.org.sanblaze:virtualun.prme-hwe-drv-sanblaze-002.0.0#vmhba33#200200110de23a00:200400110de23a00</td>
<td>259</td>
<td>vmhba33</td>
<td>FC</td>
<td>true</td>
</tr>
<tr>
<td>nqn.2014-08.org.sanblaze:virtualun.prme-hwe-drv-sanblaze-002.1.0#vmhba34#200300110de23b00:200500110de23b00</td>
<td>264</td>
<td>vmhba34</td>
<td>FC</td>
<td>true</td>
</tr>
<tr>
<td>nqn.2010-06.com.purestorage:flasharray.4d4baf0f03558e0f#vmhba35#10.20.54.101</td>
<td>266</td>
<td>vmhba35</td>
<td>RDMA</td>
<td>true</td>
</tr>
<tr>
<td>nqn.2010-06.com.purestorage:flasharray.4d4baf0f03558e0f#vmhba35#10.20.54.102</td>
<td>268</td>
<td>vmhba35</td>
<td>RDMA</td>
<td>true</td>
</tr>
</tbody>
</table>

[root@localhost:~] esxcli nvme namespace list
<table>
<thead>
<tr>
<th>Name</th>
<th>Controller Number</th>
<th>Namespace ID</th>
<th>Block Size</th>
<th>Capacity in MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>t10.NVMe_Vmware_Virtual_NVMe_Disk</td>
<td>256</td>
<td>000000001</td>
<td>512</td>
<td>40960</td>
</tr>
<tr>
<td>eui.600110d003e23b0004010000ac07d235</td>
<td>264</td>
<td>1</td>
<td>512</td>
<td>10240</td>
</tr>
<tr>
<td>eui.600110d003e23b0004010000ac07d236</td>
<td>264</td>
<td>2</td>
<td>512</td>
<td>8192</td>
</tr>
<tr>
<td>eui.600110d002e23a0003000000c5728fa4</td>
<td>259</td>
<td>1</td>
<td>512</td>
<td>2048</td>
</tr>
<tr>
<td>eui.600110d002e23a0003000000c5728fa5</td>
<td>259</td>
<td>2</td>
<td>512</td>
<td>8192</td>
</tr>
<tr>
<td>eui.600110d002e23a0003000000c5728fa6</td>
<td>259</td>
<td>3</td>
<td>512</td>
<td>61440</td>
</tr>
<tr>
<td>eui.00d80b8bccc7e4324a9374a00011fc6</td>
<td>266</td>
<td>73670</td>
<td>512</td>
<td>10240</td>
</tr>
<tr>
<td>eui.00d80b8bccc7e4324a9374a00011fc7</td>
<td>266</td>
<td>73671</td>
<td>512</td>
<td>61440</td>
</tr>
<tr>
<td>eui.00d80b8bccc7e4324a9374a00011fc6</td>
<td>268</td>
<td>73670</td>
<td>512</td>
<td>10240</td>
</tr>
<tr>
<td>eui.00d80b8bccc7e4324a9374a00011fc7</td>
<td>268</td>
<td>73671</td>
<td>512</td>
<td>61440</td>
</tr>
</tbody>
</table>
```
NVMe-oF (RoCE) based vSAN Disaggregation

Proof Of Concept

Compute Only ESXi/vSAN Nodes

JBOF Resource Pool - A

JBOF Resource Pool - B

NVMe-oF

RDMA Over Converged Ethernet (RoCE)

NVMe-oF Benefits for vSAN

• Scale storage and compute independently

• Retain the simplicity of HCI management for provisioning and disaggregation workflows

Storage Node – Intel® Storage System
RAF1000JSP with Intel® P4500 ‘Ruler’ Form Factor SSDs
NVMe-oF (TCP/IP) based vSAN Disaggregation

Proof Of Concept: The benefits of HCI and disaggregation Combined!

**Customer Value**

- Same integrated vSAN management and user experience
- Fully-featured Enterprise grade storage solution
- Scale Storage and Compute Independently
- Heterogeneous clustering of vSAN nodes with local drives or diskless
- NVMe Performance
- Enhanced SSD endurance
- Thin Provisioning, Compression, RAID
In Summary

- NVMe-oF has heralded very exciting time in Storage industry
- Spurring innovation and changes in OS, DAS and SAN Models
- Customers are the ultimate winners!