OFA Logo Program Developments

#OFADevWorkshop
Presented by Bob Noseworthy, Technical Sherpa
University of New Hampshire’s InterOperability Laboratory (UNH-IOL)
Outline

• Evolution and Benefits of the OFA Logo Program
  – Overview
  – Recent updates
• Using the OFA Logo List
• Feedback needed:
  – Continued evolution
    • Virtualization, NVMe, IPv6, Increased Distro involvement
  – Additional ULPs of Interest
• Next challenges
UNH-IOL / OFA

• UNH-IOL:
  – Host of the OpenFabrics Interoperability Logo Group
  – >25 Years providing Interop/Conformance Test
  – 2013: 12th Recipient of the IEEE-SA Corporate Award

• Your Speaker: Bob Noseworthy
  – Technical Sherpa / Chief Engineer
  – 18 Years in Ethernet conformance test
    • (10Mbps-100Gbps)
  – Expertise in IEEE 802.3, 802.1
  – Oversaw IOL’s brief iWARP Consortium
Today’s OFA Logo Program
OpenFabrics Interop Logo Group

- OpenFabrics Interoperability Logo Group (OFILG)
  - **Purpose**: Validate OFED functionality, test ULPs and verify interoperability in a heterogeneous environment
  - **Current Members**:
    - Chelsio, DDN, Emulex, IBM, Intel, Mellanox and NetApp
- Validating IB, RoCE and iWARP
- Current Upper Layer Protocols Tested
  - Fabric Init, IPoIB, Link Init, NFSoRDMA, Open MPI, RDMA Utilities, RSocket, SM failover, SRP, uDAPL (& optional iSER and RDS)
- OFA Logo Test Plan defined by OFA-IWG
  - driven by OFA member contribution

March 30 – April 2, 2014
#OFADevWorkshop
OFILG (2)

• OFA Cluster hosted at UNH-IOL
  – **Servers**: iWARP 12 hosts, InfiniBand 18 hosts, RoCE 15 hosts
  – **InfiniBand HW**: 12 HCAs, 4 switches, 5 SRP targets, 1 gateway
  – **iWARP HW**: 9 RNICs, 1 switch
  – **RoCE HW**: 6 RCA, 1 switch

• OFED versions Tested
  – 1.5.x, 3.5.x, 3.12

• PXE Boot environment available
  – Centos 6.x
  – Past Distros: RHEL 5.x and 6.x, SLES 11, Ubuntu 10.04 and 12.04
  – OFED 1.4.x, 1.5.x, 3.5.x, 3.12

• Highly scripted test environment
  – January 2014 Logo Event Tests executed – approximately 8,049
OFILG – Recent test topologies

InfiniBand Logo Event Topology

RoCE Logo Event Topology

iWARP Logo Event Topology
OFA Logo Program (1/3)

- [iol.unh.edu/ofilp](iol.unh.edu/ofilp)
- P&P doc (Policy & Procedures)
- What can be on the Logo List?
OFA Logo Program (2/3)

• What can be on the Logo List? Additions of v1.16/v1.17 in bold
  – InfiniBand HCA
  – Ethernet R-NIC
  – RoCE Adapter (RCA)
  – InfiniBand Switch with Subnet Manager
  – InfiniBand Switch with no Subnet Manager
  – Ethernet Switches
  – Ethernet DCB & Fabric Switches
  – SRP Target/Server over InfiniBand
  – iSER Target/Server (over InfiniBand or iWARP or RoCE)
  – NFS-RDMA Client/Server (over InfiniBand or iWARP or RoCE)
  – Gateways (InfiniBand-to-Ethernet, InfiniBand-to-FibreChannel)
OFA Logo Program (3/3)

• What can be on the Logo List? Additions of v1.16/v1.17 in bold
  – Server Systems using InfiniBand HCA and running OFA software
  – Server Systems using InfiniBand HCA and running non-OFA software
    • such as for Sun Solaris, Apple Mac, HPUX, IBM AIX and other operating systems
  – Server Systems using Ethernet R-NIC and running OFA software
  – Server Systems using Ethernet R-NIC and running non-OFA software
    • such as for Sun Solaris, Apple Mac, HPUX, IBM AIX and other operating systems
  – Server Systems using RoCE Adapters and running OFA software
  – Server Systems using RoCE Adapters and running non-OFA software
    • such as for Sun Solaris, Apple Mac, HPUX, IBM AIX and other operating systems
Recent additions - RoCE

- First RCAs added May 2013
- End stations tested
  - RCA providers
  - System providers
- Anticipate return of iSER targets and additional system provider participation this year
- Bridges / Ethernet Fabrics
  - More on this in upcoming slides
The OFA Logo List (iol.unh.edu/ofilglist)

Example from 2014 list:

### InfiniBand Switches With Managers

Features Tested:
- Link Initialization
- Fabric Initialization
- IPoIB
- SM Failover
- SRP
- RDMA Operations
- RDMA Stress
- MPI

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Model Number</th>
<th>Ports</th>
<th>Speed</th>
<th>HW</th>
<th>FW</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel</td>
<td>12200-CH01</td>
<td>36</td>
<td>QDR</td>
<td>3</td>
<td>7.2.0.1.1</td>
<td>Report</td>
</tr>
<tr>
<td>Mellanox</td>
<td>SX-6036</td>
<td>36</td>
<td>FDR</td>
<td>X2</td>
<td>9.1.9470</td>
<td>Report</td>
</tr>
<tr>
<td>Mellanox</td>
<td>IS-5030</td>
<td>36</td>
<td>QDR</td>
<td>X2</td>
<td>7.4.220</td>
<td>Report</td>
</tr>
</tbody>
</table>

### InfiniBand Switches Without Managers

Features Tested:
- Link Initialization
- Fabric Initialization
- IPoIB
- SM Failover
- SRP
- RDMA Operations
- RDMA Stress
- MPI

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Model Number</th>
<th>Ports</th>
<th>Speed</th>
<th>HW</th>
<th>FW</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mellanox</td>
<td>SX-6025</td>
<td>36</td>
<td>FDR</td>
<td>X2</td>
<td>9.2.0</td>
<td>Report</td>
</tr>
</tbody>
</table>

### SRP Targets

Features Tested:
- IB Link Initialization
- IB Fabric Initialization
- SM Failover
- SRP

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Model Number</th>
<th>Ports</th>
<th>Speed</th>
<th>HW</th>
<th>FW</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>DataDirect Networks</td>
<td>SFA10000</td>
<td>8</td>
<td>QDR</td>
<td>NA</td>
<td>2.1.1.2.18319</td>
<td>Report</td>
</tr>
<tr>
<td>DataDirect Networks</td>
<td>SFA12000</td>
<td>4</td>
<td>FDR</td>
<td>NA</td>
<td>2.1.1.2.18319</td>
<td>Report</td>
</tr>
<tr>
<td>DataDirect Networks</td>
<td>SFA7700</td>
<td>4</td>
<td>FDR</td>
<td>NA</td>
<td>2.1.1.2.18319</td>
<td>Report</td>
</tr>
<tr>
<td>Netapp</td>
<td>Pikes Peak (E5400)</td>
<td>4</td>
<td>QDR</td>
<td>1.0</td>
<td>07.86</td>
<td>Report</td>
</tr>
<tr>
<td>Netapp</td>
<td>Soyuz (E5500)</td>
<td>4</td>
<td>QDR</td>
<td>1.0</td>
<td>07.86</td>
<td>Report</td>
</tr>
</tbody>
</table>
OFA Logo List Reports

- [http://iol.unh.edu/ofilglist](http://iol.unh.edu/ofilglist)
- Most recent Logo List includes summary report of all testing
Using the OFA Logo List

• How do you use it?
• Do you also use the IBTA Integrator’s List?
  – For Cables?
  – For Equipment (HCAs, Switches)?
• Is the product you are evaluating on the list?
• What limits your usage?
  – Validation Depth?
  – Validation Breadth?
    • More ULPs covered, different Distros validated
  – Participating Vendors/Product variety?
Tomorrow’s OFA Logo Program?
Logo Program Next Steps

• Input needed
  – Set priorities
  – End User Stories / Needs

• Any OFA Member may participate in the OFA’s Interoperability Working Group (OFA-IWG):
  – e.g:
    • End Users;
    • Application Developers;
    • Purchasers and Evaluators;
  – share your needs and desires with the OFA-IWG

• Only OFILG Members can receive logos for their products

• The following slides are my speculation and are NOT active work within the OFA-IWG, if you support some of the ideas, please get involved !!!
RoCE Bridge Testing

- Ethernet Bridges tested
  - None formally, yet play a critical role in CE
  - Emulex provided an Arista Networks 40GbE switch for current RoCE testing

- Put the “Fabric” in Ethernet Fabric validation

- Ethernet Fabrics are real
  - Congestion Avoidance (Data Center Bridging)
  - Well-beyond Multiple Spanning Tree
    - Proprietary examples: Qfabric, MLAG, FabricPath
    - Standards-based: IETF TRILL, IEEE SPB
Virtualization

• Support for hardware pass-thru for RDMA
• Which Hypervisor solutions?
  – Must be OFILG member driven
• Which OSes within the VM
  – Again: OFILG member driven

• Virtualization providers encouraged to directly participate in OFILG
Distro/OS provider Involvement

- Today using Scientific Linux, or Centos
- Distro/OSes tested in previous Interop events:
  - Canonical’s Ubuntu
  - Microsoft (WinOfed)
  - RedHat
  - SUSE

- Distro/OS providers encouraged to directly participate in OFILG

- Distro/OS choice is at the direction of OFILG and OFA-IWG members
  - (as is the entirety of the OFA Logo Test Plan)
Speed

• IB at FDR / EDR
• Ethernet at 40 / 100 Gbps
  – 4x10Gbps or 10x10Gbps
  – Soon 4x25Gbps
• Testing at Higher Speeds / lower latencies critical for validation of some ULPs
• Requires
  – PCIe gen3 / gen4
  – Continuously challenging OFA test-cluster resources
  – Credit to AMD, HP, Intel for their contributions to the cluster over the years to support ever faster technologies.
Speed begets Performance

• Should OFA Logo Program expand to include Performance requirements?
  – Discussed at length during formation process
  – Currently Logo Program validates functionality and interoperability, but does not restrict based on performance
    • The Market is currently left to make that determination
  – Should a “minimum bar” be set for some of the validation
    • e.g: A fixed number or some percentage below a moving average of performance numbers from past events
NVMe

- Further pushing the envelop of speed
- UNH-IOL NVMe Consortium
  - Working in concert with the NVM Express Organization
- Integrator’s List
  - https://www.iol.unh.edu/services/testing/NVMe/integratorslist.php
- Validation of storage protocols / solutions at NVMe speeds
  - eg: EMC XtremeIO, Fusion-IO, DDN Storage Fusion Xcelerator (SFX) solutions, etc.
IPv6

• UNH-IOL IPv6 Consortium / IPv6 Forum
• IPv6 Ready Logo List
  – https://www.iol.unh.edu/services/testing/ipv6/logoholders.php
• USGv6 Certification
  – https://www.iol.unh.edu/services/testing/ipv6/usgv6tested.php
• Not done today:
  – Validation of IPv6 connectivity in an RDMA OFED environment
    • IPoIB
    • IPoCE
    • iWARP-o-v6 ?

March 30 – April 2, 2014

#OFADevWorkshop
Still more growth options

• Scalability
  – Today, only functional testing is done (homogeneously and heterogeneously)

• Improved Fail-Over validation
  – HCA to HCA Failover / RCA-to-RCA Failover
    • requested via OFA RDMA Programming & System Admin Classes

• Additional ULPs
  – Lustre

• Or…?
  – Contribute your ideas to the OFA-IWG discussion
Interop vs Conformance

• OFILG focus:
  – Interoperability validation of OFED and ULPs with underlying RDMA transports
  – Conformance validation is currently not required to participate in the logo testing.
## Interop v Conformance (2)

<table>
<thead>
<tr>
<th></th>
<th>Interop</th>
<th>Conformance</th>
<th>Logo / I.L./Cert.</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>IB</td>
<td>OFA</td>
<td>IBTA</td>
<td>OFA Logo, IBTA I.L.</td>
<td></td>
</tr>
<tr>
<td>iWARP</td>
<td>OFA</td>
<td>None</td>
<td>OFA Logo</td>
<td>Past UNH-IOL Conformance test</td>
</tr>
<tr>
<td>RoCE</td>
<td>OFA</td>
<td>None</td>
<td>OFA Logo</td>
<td>Future IBTA I.L.?</td>
</tr>
<tr>
<td>DCB</td>
<td>UNH-IOL</td>
<td>UNH-IOL</td>
<td>None</td>
<td>OFA Logo: Add subset?</td>
</tr>
<tr>
<td>TRILL</td>
<td>UNH-IOL</td>
<td>UNH-IOL</td>
<td>None</td>
<td>OFA Logo: Add subset?</td>
</tr>
<tr>
<td>SPB</td>
<td>UNH-IOL</td>
<td>UNH-IOL</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>IPv6</td>
<td>UNH-IOL</td>
<td>UNH-IOL</td>
<td>IPv6 Ready / USGv6</td>
<td>OFA Logo: Add subset?</td>
</tr>
<tr>
<td>NVMe</td>
<td>UNH-IOL</td>
<td>UNH-IOL</td>
<td>UNH-IOL I.L.</td>
<td>OFA Logo: Add subset?</td>
</tr>
<tr>
<td>PCIe</td>
<td>UNH-IOL</td>
<td>PCI-SIG/ UNH</td>
<td>PCI-SIG Workshops</td>
<td></td>
</tr>
<tr>
<td>40/100GE</td>
<td>UNH-IOL</td>
<td>UNH-IOL</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Cables (IB)</td>
<td>None</td>
<td>IBTA</td>
<td>IBTA I.L.</td>
<td></td>
</tr>
<tr>
<td>Cables (10-100GE)</td>
<td>UNH-IOL</td>
<td>UNH-IOL</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>
Next Steps: Much to discuss

• Far more to consider than this brief presentation and discussion can touch on

• Continue the discussion with the:  
  *Open Fabrics Interoperability Working Group*  
  (OFA-IWG)

• Survey
  – [https://docs.google.com/forms/d/1qwGuWe4rsFvYM6nyJVwa4q12_bWDiLlfRFi3p0EGV_w/viewform](https://docs.google.com/forms/d/1qwGuWe4rsFvYM6nyJVwa4q12_bWDiLlfRFi3p0EGV_w/viewform)
  – Same link as above: [http://tinyurl.com/2014OFADev](http://tinyurl.com/2014OFADev)
How to get involved…

• OFA Interoperability Working Group (IWG)
  – Every other Tuesday, 1pm ET
  – **Tuesday, April 08, 2014, 10:00 AM US Pacific Time**
    916-356-2663, 8-356-2663, Bridge: 1, Passcode: 7004471

• Join the OFA-IWG Mailing List
  – [https://www.openfabrics.org/index.php/working-groups/wg-mail-subscription.html?view=form](https://www.openfabrics.org/index.php/working-groups/wg-mail-subscription.html?view=form)

• Links:
  – *OFA Interop Program Overview:*
    [https://www.openfabrics.org/images/docs/LinkedDocs/UNH_IOL_OFA_11-11_MR_FINAL.pdf](https://www.openfabrics.org/images/docs/LinkedDocs/UNH_IOL_OFA_11-11_MR_FINAL.pdf)
  – *OFA Logo List:*  [http://iol.unh.edu/ofilglist](http://iol.unh.edu/ofilglist)
  – *OFA Logo Program:*  [http://iol.unh.edu/ofilp](http://iol.unh.edu/ofilp)
  – *OFA Test Plan:*  [http://iol.unh.edu/ofatestplan](http://iol.unh.edu/ofatestplan)
Thank You

- Contact:
  Bob Noseworthy (ren@iol.unh.edu)
  UNH-IOL OFA Group (ofalab@iol.unh.edu)
And in other news: Training

Next OFA Training courses @ UNH-IOL

- May 19-20\textsuperscript{th}: Writing Application Programs for RDMA using OFA Software
  - [https://www.openfabrics.org/index.php/resources/training/training-offerings.html](https://www.openfabrics.org/index.php/resources/training/training-offerings.html)

- May 21-22\textsuperscript{rd}: Infiniband Fabric Administration

- On-site training can also be arranged
  - please contact Rupert Dance at [rsdance@soft-forge.com](mailto:rsdance@soft-forge.com)