12th ANNUAL WORKSHOP 2016

INTEL® OMNI-PATH™ FABRIC MANAGEMENT AND TOOLS FEATURES

James Wright

Intel

[April 7th, 2016]

- **Outline**
  - Overview of Software Components and usage model
  - Management Software Details
    - Management Agents, Administration
    - Fabric Diagnostic and Debug Features
    - Scalable Fabric management GUI
  - Command Line Tools Details
    - Monitoring & Diagnostics
    - Configuration and Management
  - Usage Examples
    - Host verification
    - Performance analysis

Familiar fabric management components

- ### Fabric management stack

  - New fabric manager. Management of Omni-Path Architecture (OPA) fabric/features
    - Runs on OPA-connected management nodes
      - or switch embedded processor
    - Initializes, configures and monitors the fabric's routing, QoS, security, and performance features
  - New management agents
    - Runs on OPA-managed host and switch nodes
    - In-band management datagrams

- ### Host Command Line (CLI) tools

  - New tools. Monitoring/control of local/fabric OPA resources
  - Includes updated FastFabric toolkit for configuration, monitoring, diags, and repair

- ### Fabric Management Graphical User Interface (GUI)

  - New interface. Exposes OPA features
  - Provides interactive GUI access to Fabric Management features (configuration, monitoring, diagnostics, element management drill down)

INTEL® OMNI-PATH™ FABRIC MANAGEMENT AND TOOLS FEATURES
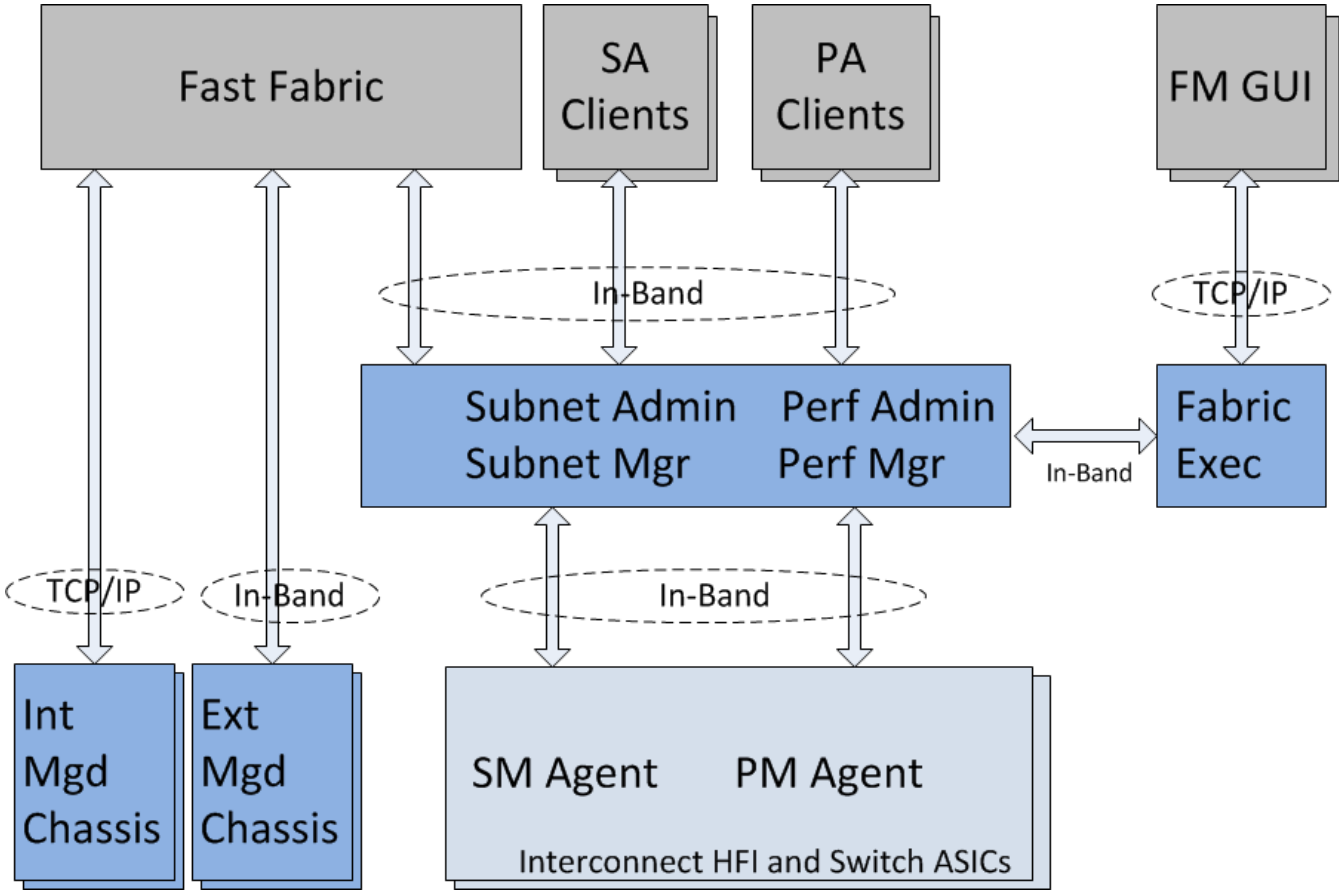
Fabric Management Software
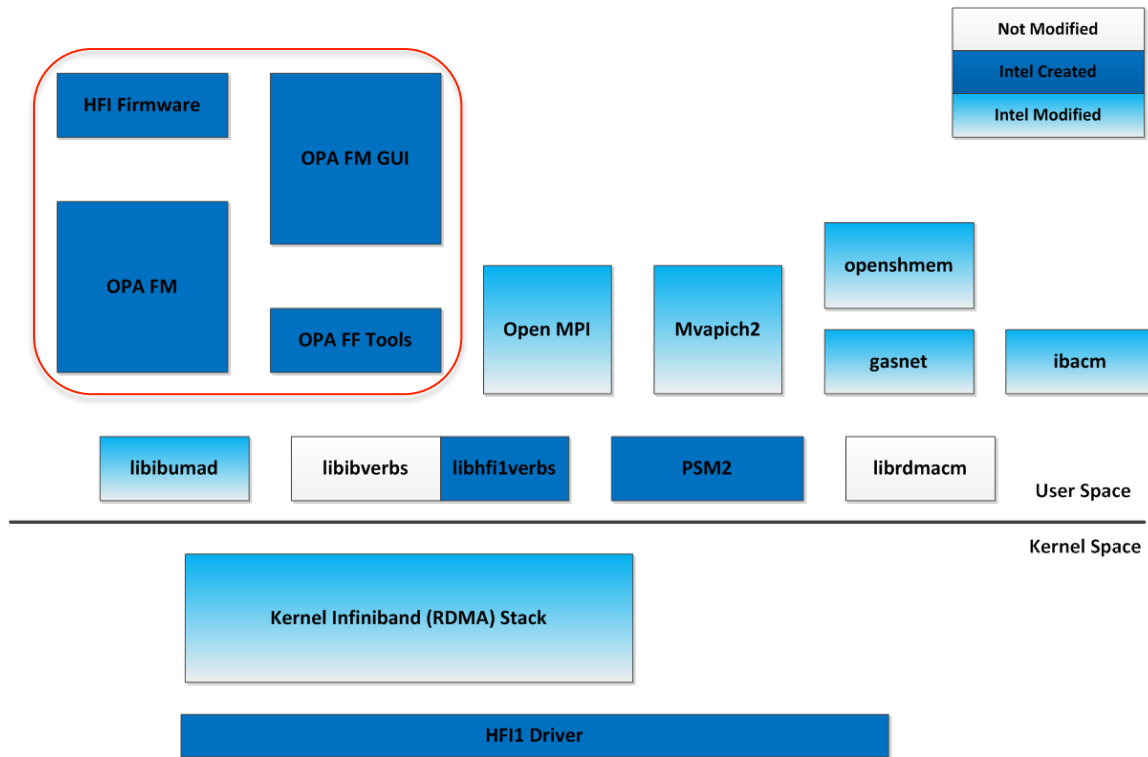
- **Leverages existing stacks for each type of management**
- **Supports 3rd party unified management consoles**
- **Provides a scalable centralized fabric management framework**

Intel® Omni-Path integrates with the OpenFabrics stack.

Fabric Management components

Fabric Manager's Subnet Management

- **Traditional responsibilities**
  - Initialization, QoS/policy definitions, topology discovery, LID assignment, routing, change monitoring, etc

- **Traditional exchanges of SMI packets with SMA management agents**
  - Directed route, LID-route, hybrids

- **Traditional run time database support for SA queries**
  - 'opasaquery', 'opareport'

Some OPA differences

- **Scaling efficiencies**
  - 2k, aggregates, multi-port (MADs)

- **Topology verification**

- **Advanced traffic features**

- **vFabrics**

SMA

Subnet Manager (SM)

SMA

SMA

SM communicates with
Subnet Manager Agents (SMA)
on each node
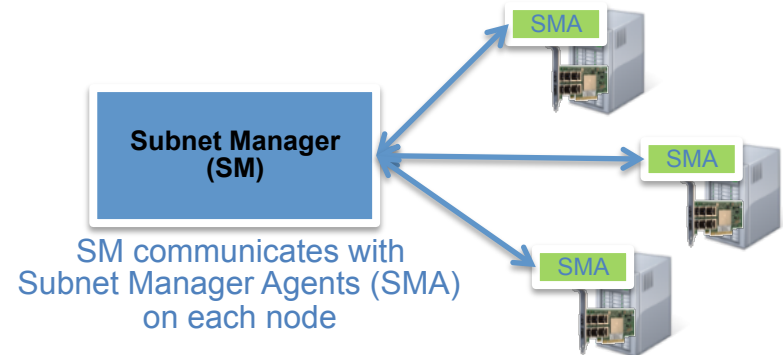
The FM implements a complete Intel® Omni-Path Architecture-compliant Subnet Manager (SM)

- **FM Topology verification**
  - From install spreadsheet (or known good snapshot)
  - Create punch-lists during install
  - Monitor for unauthorized changes during operations
  - FM can warn, quarantine incorrect links
  - FastFabric online and offline topology analysis

- **Extensive cable FRU management support built-in**
  - **Link Quality Indicator**
    - "5 bars" instantaneous view of link quality
    - In every HW port, monitored by FM, FastFabric Tools, FM GUI
  - **Port type information**
    - QSFP/Standard, Fixed/Backplane, Variable, Disconnected, …
  - **QSFP CableInfo**
    - shows all key cable /transceiver info
    - Vendor, model, length, technology, date, etc
    - Fully integrated into FM, FastFabric tools, FM GUI
  - **Link Down Reason**
    - LinkDownReason and NeighborLinkDownReason – most recent reason link went down

Addressing the sysadmin challenge of managing cable FRUs
- **In a typical cluster, the majority of fabric FRUs are cables**

INTEL® OMNI-PATH™ FABRIC MANAGEMENT AND TOOLS FEATURES

Some OPA differences - Advanced Traffic Features

	Description	Benefits
Traffic Flow Optimization	▪ Optimizes Quality of Service (QoS) in mixed traffic environments, such as storage & MPI ▪ Transmission of lower-priority packets can be paused so higher priority packets can be transmitted	▪ Ensures high priority traffic is not delayed →Faster time to solution ▪ Deterministic latency → Lowers run-to-run timing inconsistencies
Packet Integrity Protection	▪ Allows for rapid recovery of transmission errors on an Intel® OPA link with low latency for both corrupted and uncorrupted packets ▪ Resends 1056-bit bundles rather than entire packet	▪ Fixes happen at the link level rather than end-to-end level ▪ Much lower latency than Forward Error Correction (FEC) defined in the InfiniBand specification[1]
Dynamic Lane Scaling	▪ Maintain link continuity in the event of a failure of one of more physical lanes ▪ Operates with the remaining lanes until the failure can be corrected at a later time	▪ Enables a workload to continue to completion. ▪ Enables service at appropriate time.

[1] Lower latency based on the use of InfiniBand with Forward Error Correction (FEC) Mode A or C in the public presentation titled "Option to Bypass Error Marking (supporting comment #205)," authored by Adee Ran (Intel) and Oran Sela (Mellanox), January 2013. Link: www.ieee802.org/3/bi/public/jan13/ran_3bi_01a_0113.pdf

OpenFabrics Alliance Workshop 2016

- **vFabrics**
  - Permit multiple applications to be run on the same fabric at the same time with limited interference
  - Administrator controls the degree of isolation

- **A vFabric consists of**
  - A group of applications that run on a group of devices.
  - For each vFabric the operational parameters of the vFabric can be selected

- **Using vFabrics**
  - Defaults to simple Admin, Default vFabrics
  - Administrator may further slice up the physical fabric into many overlapping virtual fabrics

- **Typical usage**
  - Separating a cluster into multiple vFabrics so that independent applications can run with minimal or no effect on each other
  - Separating classes of traffic. For example, compute, storage, mgmt sharing same wire for all networking
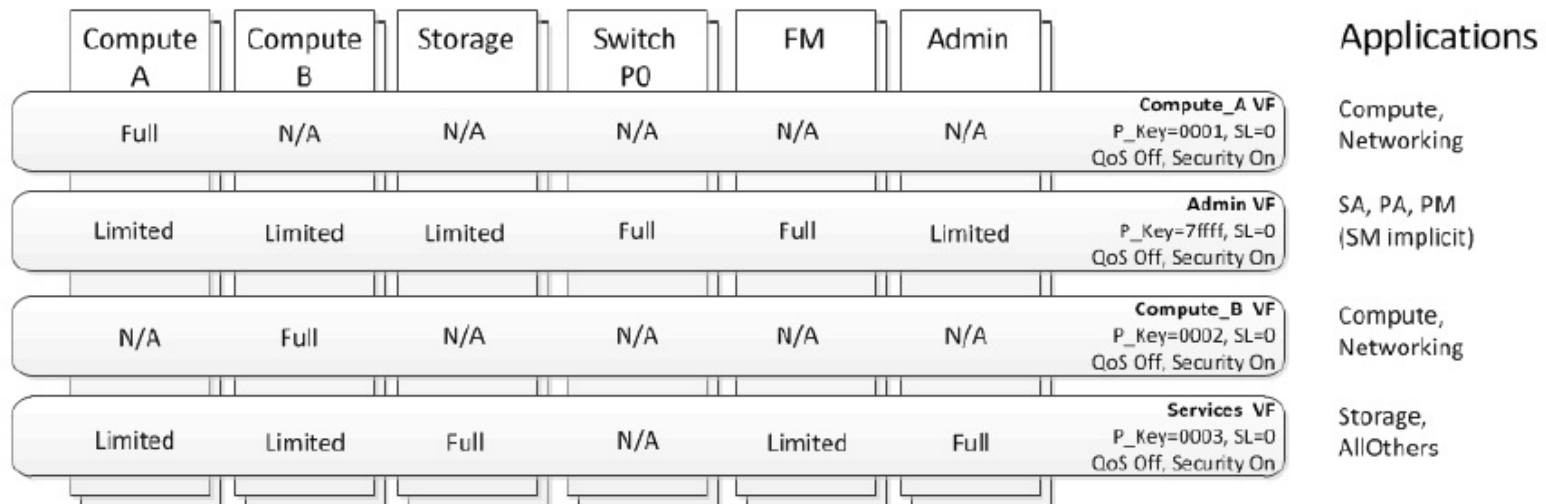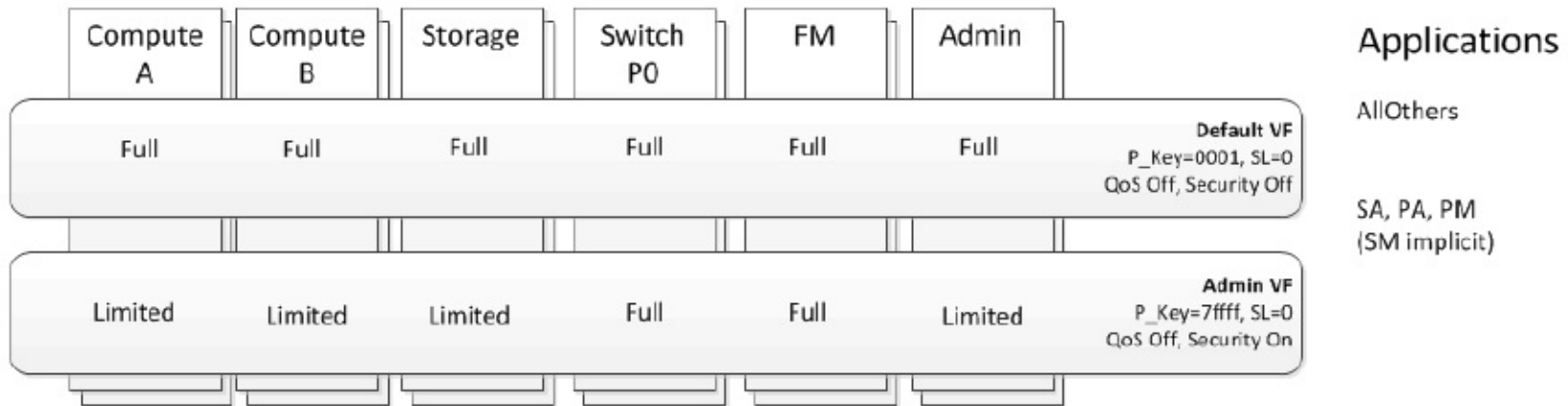
OpenFabrics Alliance Workshop 2016

Fabric Manager's Performance Manager

- **Traditional responsibilities**
  - Gathers port counter info
    - Data counters
    - Error counters
    - Per port, per VL
    - Configurable
  - Monitors error levels against defined alarm
  - Archives for post-event data analysis
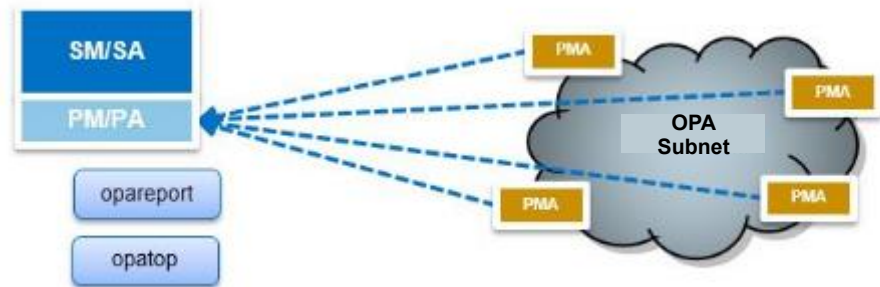
- **Traditional exchange of GSI packets with PMA management agents**
  - LID routed
- **Traditional run time database support for PA queries**
  - 'opapaquery', 'opareport', 'opatop'

Some OPA differences

- 'Short term' (24-hour) history, supports historical queries e.g. "what was happening 3 hours ago"
- Filter by user defined device-groups, sweep images, other criteria
- PM/PA database sync – PM data retained during FM failover

The FM implements a complete Intel® Omni-Path Architecture-compliant Performance Manager (PM)

Management GUI

- **Offers "big picture" of a fabric**
- **Queries SA, PA data through FE**

- **Subnet Summary**
  - Statistics
  - Status
  - Health trend
  - Worst nodes
- **Subnet performance**
  - Utilization group
  - Counter Categories group
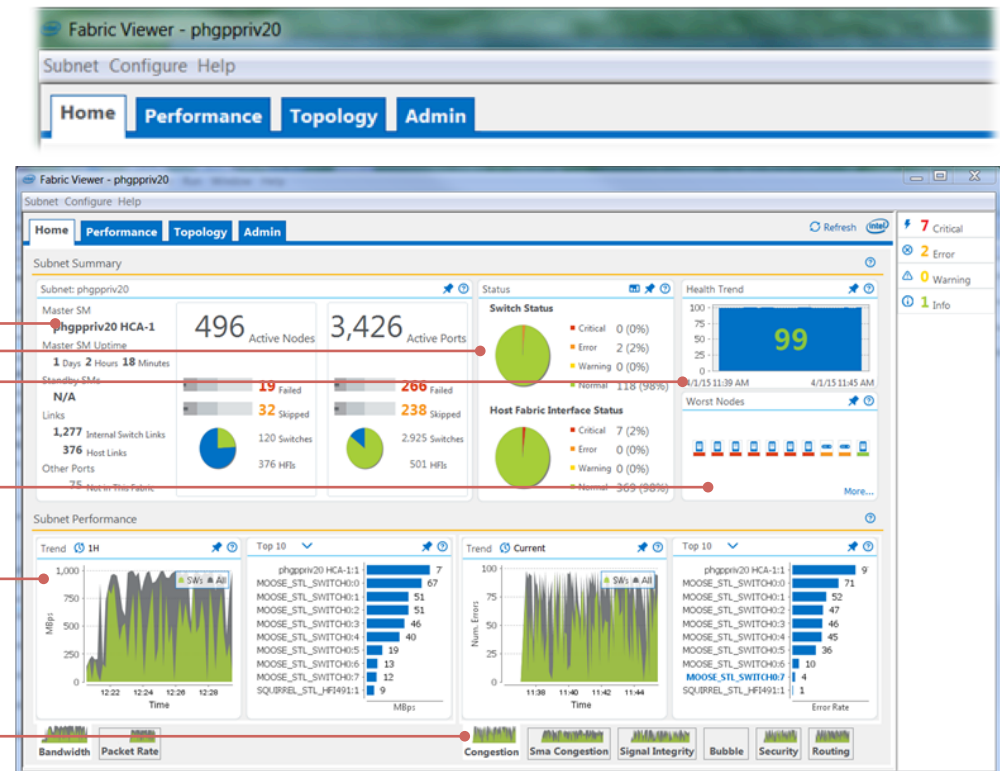
Fabric Management GUI accesses FM data through FE interface

- **Outline**
  - Overview of Software Components and usage model
  - Management Software Details
    - Management Agents, Administration
    - Fabric Diagnostic and Debug Features
    - Scalable Fabric management GUI
  - **Command Line Tools Details**
    - **Monitoring & Diagnostics**
    - **Configuration and Management**
  - Usage Examples
    - Host verification
    - Performance analysis

INTEL® OMNI-PATH™ FABRIC MANAGEMENT AND TOOLS FEATURES

Command Line Tools

- ## Familiar command scope
  - Local – Host fabric interface (HFI) ports/agents management
  - Platform – e.g. Edge/DCS switch element management
  - Fabric – query fabric's FM databases, filtered results

- ## OPA command line tools
  - Intel-supplied BASIC/IFS tools packaging for common distros (rpm)
    - Or build from source (github)
  - Similarities to IB commands, patterned after TrueScale, but OPA-specific
    - '--help'
    - 'man' pages
    - Install guides, user guides, command references (PDF)

IB	TRUE SCALE	OMNI-PATH
ibstat	iba_info	opainfo

INTEL® OMNI-PATH™ FABRIC MANAGEMENT AND TOOLS FEATURES

Extensive CLI tools

Category	CLIs
Monitoring and Diagnostics	
High Level Analysis/Monitoring	Intel® Omni-Path Fabric Suite FastFabric TUI Intel® Omni-Path Fabric Suite Fabric Manager GUI Intel® Omni-Path Fabric Suite FastFabric CLI: opatop, opalinkanalysis, opallanalysis (opachassisanalysis, opaesmanalysis, opahostsmanalysis, opafabricanalysis)
Mid-Tier CLI Analysis/Diagnosis	Intel® Omni-Path Fabric Suite FastFabric CLI: opareports, opareport, opaextractbadlinks, opaextractlink, opaextractsellinks, opaextractstat2, opafabricinfo
Lower Level CLI Diagnosis	Intel® Omni-Path Fabric Suite FastFabric CLI: opaextractlids, opaextracterror, opaextractperf, opaextractstat, opashowallports
Deep Level CLI Diagnosis/Data	Intel® Omni-Path Fabric Suite FastFabric CLI: opafequery, opafirmware, opahfirev, opapmaquery, opasmaquery, opaportinfo, opasaquery, opashowmc, opapaquery

Extensive set of command tools available from day 1

Extensive CLI tools

Management	
Multi-Switch External Management	Intel® Omni-Path Fabric Suite FastFabric CLI: opaswitchadmin, opagenswitches
Multi-Switch Internal Management	Intel® Omni-Path Fabric Suite FastFabric CLI: opachassisadmin, opacmdall, opapingall, opagenchassis, opagenesmchassis, opasetupssh
Multi-Host Management	Intel® Omni-Path Fabric Suite FastFabric CLI: opacmdall, opapingall, opadownloadall, opauploadall, opafindgood, opahostadmin, opascpall, opasetupssh, opaverifyhosts
Installation	Intel® Omni-Path Fabric Suite FastFabric CLI: opaconfig opa_config_ff, opa_config_fm
Link and Port Management	Intel® Omni-Path Fabric Suite FastFabric CLI: opaenableports, opadisableports, (opaextractbadlinks, opaextractsellinks generate input format for opaenable/disableports), opadisablehosts, opainfo, opaportinfo, opapmaquery, opasmaquery, opaportconfig, opaswdisableall
Fabric Manager (FM) Management	Intel® Omni-Path Fabric Suite FastFabric CLI: opafmcmd, opafmcmdall, opafmconfigcheck, opafmconfigdiff service opafm
Topology Analysis and Management	Intel® Omni-Path Fabric Suite FastFabric CLI: opareport -o verify*, opagentopology, opareport -o links, opaextractlink, opaextractsellinks
SRP IOUs	Intel® Omni-Path Fabric Suite FastFabric CLI: opareport -o ious
Link Issue Analysis	Intel® Omni-Path Fabric Suite FastFabric TUI Intel® Omni-Path Fabric Suite Fabric Manager GUI Intel® Omni-Path Fabric Suite FastFabric CLI: opatop, opaallanalysis, opaextractbadlinks, opareport (-o errors, slow* mis*), opashowallports

continued...

OpenFabrics Alliance Workshop 2016

INTEL® OMNI-PATH™ FABRIC MANAGEMENT AND TOOLS FEATURES

Extensive CLI tools

Category	CLIs
QoS Analysis	Intel® Omni-Path Fabric Suite FastFabric CLI: opareport (-o vfinfo, vfmember, bfrctrl), opasmaquery, opasaquery -o vfinfo, opasaquery -o path, other opasaqueries for sc, sl, vl tables, opareport -V -o comps -d 10 (dumps all QoS config)
Routing Analysis	Intel® Omni-Path Fabric Suite FastFabric CLI: opareport (-o portusage, treepathusage, pathusage, portgroups, validateroutes, validatepgs, validatecreditloops, linear, mcast)
Customer Support Data Gathering Host and Switch	Intel® Omni-Path Fabric Suite FastFabric CLI: opacaptureall, opacapture
Topology Generation and Conversions	Intel® Omni-Path Fabric Suite FastFabric CLI: opagentopology, opareport -o topology, opaxlattopology, opaxlattopology_cust, opaxmlgenerate
General XML File Utilities	Intel® Omni-Path Fabric Suite FastFabric CLI: opaxmlextract, opaxmlfilter, opaxmlindent
Verify Fabric (VF) query for launch integration	Intel® Omni-Path Fabric Suite FastFabric CLI: opagetvf, opagetvf_env
Distributive Subnet Administrator (SA)	Intel® Omni-Path Fabric Suite FastFabric CLI: opa_osd_dump, opa_osd_exercise, opa_osd_perf, opa_osd_query
Other	
Wireshark	Intel® Omni-Path Fabric Suite FastFabric CLI: opapacketcapture

OpenFabrics Alliance Workshop 2016

Ease of use

- **Textual User Interface (TUI) options available**
  - Stepwise installation, fabric verification assistance
    - opaconfig, opafastfabric (examples)

```
Intel OPA 10.0.1.0.21 Software

  1) Show Installed Software
  2) Reconfigure OFA IP over IB
  3) Reconfigure Driver Autostart
  4) Generate Supporting Information for Problem Report
  5) FastFabric (Host/Chassis/Switch Setup/Admin)
  6) Uninstall Software

  X) Exit
```

```
Intel FastFabric OPA Tools
Version: 10.0.1.0.21

  1) Chassis Setup/Admin
  2) Externally Managed Switch Setup/Admin
  3) Host Setup
  4) Host Verification/Admin
  5) Fabric Monitoring

  X) Exit
```

Of course –
Direct command line options available for the more expert user

OpenFabrics Alliance Workshop 2016

Fabric Scope, Local Scope

- **Fabric scope examples**
  - opareport, opatop, opaextractsellinks, opaextractlids, opasaquery, opapaquery, more
    - TrueScale lineage: iba_report, iba_top
      (similarities to ibdiagnet, iblinkinfo, ibnodes, ibhosts, ibswitches, sminfo)
    - Spreadsheet friendly 'CSV' output options available

- **Example:**
  - 'opaextractlids -F nodetype:SW'
    get LID info on all switches present in fabric, output in CSV format

```
[hds1fnb8301 ~]$ opaextractlids -F nodetype:SW
0x00117501026a5694;0;SW;hds1swba221;0x0003
0x00117501026a55c7;0;SW;hds1swba231;0x0006
0x001175010265bd0f;0;SW;hds1swb8241;0x0007
0x001175010265bb1d;0;SW;hds1swb8231;0x0008
0x00117501026a569a;0;SW;hds1swb9211;0x000d
0x00117501026a562f;0;SW;hds1swb9221;0x000e
0x001175010265bd24;0;SW;hds1swb9231;0x0013
0x001175010265bb24;0;SW;hds1swb9241;0x0014
0x00117501026a5676;0;SW;hds1swb9411;0x0015
```

Filtered query of the FM's SA

- ## opasaquery – detailed fabric SA info
  - Multiple otype options
    - e.g. systeguid, nodeguid, portguid, lid, desc, path, node, portinfo, sminfo, swinfo, link, slsc, scsl, vlarb, pkey, service, mcmember, inform, vfinfo, fabricinfo, quarantine, conginfo, bfrctrl, cableinfo, portgroup, etc.
  - Multiple query options
    - E.g. by lid, pkey, vfindex, serviceId, SL, type, guid, gid, name/description, etc.

```
[RHEL7.1 hds1fnb8301 20160328_2218 ~]# opasaquery -o fabricinfo
Number of HFIs: 131
Number of Switches: 9
Number of Links: 255
Number of HFI Links: 131          (Internal: 0   External: 131)
Number of ISLs: 124               (Internal: 0   External: 124)
Number of Degraded Links: 0       (HFI Links: 0   ISLs: 0)
Number of Omitted Links: 0        (HFI Links: 0   ISLs: 0)
```

Filtered query of the FM's PA

- **opapaquery – detailed fabric PA info**
  - Multiple otype options
    - e.g. portCounters, focusPorts, vfList, vfInfo, vfPortCounters, groupList, groupInfo, groupConfig, imageInfo, freezeImage, renewImage, releaseImage, etc.
  - Multiple query options
    - E.g. by lid, port, time delta, start, range, image number, etc.

```
# opapaquery -o vfList
Getting VF List...
 Number of VFs: 2
 VF 1: Default
 VF 2: Admin
opapaquery completed: OK

# opapaquery -o groupList
Getting Group List...
 Number of Groups: 3
 Group 1: All
 Group 2: HFIs
 Group 3: SWs
opapaquery completed: OK
```

```
# opapaquery -V Default -o vfInfo
Getting VF Info...
 VF name: Default
 NumPorts: 519 MinRate:  25g MaxRate: 100g MaxMiBps: 12500
 Internal utilization statistics:
   Util: Tot      0 Max      0 Min      0 Avg      0 MiB/s
   Util:   519     0    0    0    0    0    0    0    0    0
   Pkts: Tot      0 Max      0 Min      0 Avg      0 KiPps/s
 Internal Error Summary:
   Integrity      Max      0    Buckets:    519    0    0    0    0
   Congestion     Max      0    Buckets:    519    0    0    0    0
   SmaCongestion Max      0    Buckets:    519    0    0    0    0
   Bubble         Max      0    Buckets:    519    0    0    0    0
   Security       Max      0    Buckets:    519    0    0    0    0
   Routing        Max      0    Buckets:    519    0    0    0    0
   Utilization:        0.0%
   Discards:           0.0%
 Image Id:
   ImageNumber: 0x800000755f020006 Offset: 0
opapaquery completed: OK
```

Local Scope (basic) examples

- **Basic CLI commands, applicable to all hosts**
- **opainfo, opaportinfo, opaportconfig, opasmaquery, opapmaquery, more**
  - TrueScale lineage: iba_info, ibv_devinfo, iba_portinfo, iba_portconfig (similarities to ibstat, ibv_devinfo, ibstatus, ibportstate)
- **Example:**
  - 'opainfo' – basic sma/pma port and cable info at a glance

```
[RHEL7.1 hds1fnb8301 20160328_0300 ~]# opainfo
hfi1_0:1                          PortGID:0xfe80000000000000:001175010165add0
   PortState:      Active
   LinkSpeed      Act: 25Gb          En: 25Gb
   LinkWidth      Act: 4             En: 4
   LinkWidthDnGrd ActTx: 4  Rx: 4    En: 3,4
   LCRC           Act: 14-bit        En: 14-bit,16-bit,48-bit      Mgmt: True
   LID: 0x00000022-0x00000022        SM LID: 0x00000022 SL: 0
   QSFP: PassiveCu,   1m FCI Electronics   P/N 10131941-2010LF   Rev 2
   Xmit Data:              41758 MB Pkts:              12850115
   Recv Data:              46742 MB Pkts:              12858817
   Link Quality: 5 (Excellent)
```

OpenFabrics Alliance Workshop 2016

- **opasmaquery – detailed local SMA info**
  - Multiple otype options
    - e.g. bfrctrl, cableinfo, conginfo, desc, hficongcon, hficonglog, hficongset, linfdb, mcfdb, portgroup, nodeaffr, nodedesc, nodeinfo, portinfo, pstateinfo, pkey, slsc, scsl, scsc, scvlt, sminfo, swaggr, swconglog, swcongset, swinfo, swportcong, vlarb, ledinfo, etc.
  - Scriptable "grep" friendly output (with -g option)

```
[RHEL7.1 hds1fnb8301 20160328_1434 ~]# opasmaquery -o portinfo | more
Port 1 Info
  Subnet:    fe80000000000000
  LocalPort:      1                  PortState:         Active
  PhysicalState: LinkUp
  OfflineDisabledReason: None
  IsSMConfigurationStarted: True     NeighborNormal: True
  BaseLID:        0x00000022         SMLID:             0x00000022
  LMC:            0                  SMSL:              0
  PortType: Standard                 LimtRsp/Subnet:     32 us, 536 ms
  M_KEY:     0x0000000000000000      Lease:       0 s   Protect: Read-only
  LinkWidth    Act: 4               En: 4           Sup: 1,2,3,4
  LinkWidthDnGrd ActTx: 4  Rx: 4    En: 3,4         Sup: 1,2,3,4
  LinkSpeed    Act: 25Gb            En: 25Gb        Sup: 25Gb
  PortLinkMode Act: STL             En: STL         Sup: STL
  PortLTPCRCMode Act: 14-bit        En: 14-bit,16-bit,48-bit Sup: 14-bit,16-bit,48-bi
  NeighborMode  MgmtAllowed: Yes FWAuthBypass: On NeighborNodeType: Switch
  NeighborNodeGuid:   0x001175010265bb1d   NeighborPortNum:   1
  Capability:     0x00410022: CN CM APM SM
  Capability3:    0x0008: SS
  SM_TrapQP: 0x0  SA_QP: 0x1
  IPAddr IPV6/IPAddr IPv4:  ::/0.0.0.0
```

```
[RHEL7.1 hds1fnb8301 20160328_1429 ~]# opasmaquery -g -o portinfo | more
Port 1 Info
  LID.......................................................... : 0x22
  FlowControlMask.............................................. : 0x0
  VL.s2.Cap.................................................... : 0x8
  VL.HighLimit................................................. : 0x0
  VL.PreemptingLimit........................................... : 0x0
  VL.ArbitrationHighCap........................................ : 0x10
  VL.ArbitrationLowCap......................................... : 0x10
  PortStates.AsReg32........................................... : 0x300054
  PortStates.s.IsSMConfigurationStarted........................ : 0x1
  PortStates.s.NeighborNormal.................................. : 0x1
  PortStates.s.OfflineDisabledReason........................... : 0x0
  PortStates.s.PortPhysicalState............................... : 0x5
  PortStates.s.PortState....................................... : 0x4
  PortStates.s.LEDEnabled...................................... : 0x0
  PortPhyConfig.AsReg8......................................... : 0x4
  PortPhyConfig.s.PortType..................................... : 0x4
  MultiCollectMask.CollectiveMask.............................. : 0x0
  MultiCollectMask.MulticastMask............................... : 0x0
  s1.M_KeyProtectBits.......................................... : 0x0
  s1.LMC....................................................... : 0x0
  s2.MasterSMSL................................................ : 0x0
  s3.PartitionEnforcementInbound............................... : 0x0
  s3.PartitionEnforcementOutbound.............................. : 0x0
  s4.OperationalVL............................................. : 0x8
  P_Keys.P_Key_8B.............................................. : 0x0
  P_Keys.P_Key_10B............................................. : 0x0
  Violations.M_Key............................................. : 0x0
  Violations.P_Key............................................. : 0x0
  Violations.Q_Key............................................. : 0x0
  SM_TrapQP.QueuePair.......................................... : 0x0
  SA_QP.QueuePair.............................................. : 0x1
  NeighborPortNum.............................................. : 0x1
  LinkDownReason............................................... : 0x0
  NeighborLinkDownReason....................................... : 0x0
  LinkInitReason............................................... : 0x1
  Subnet.ClientReregister...................................... : 0x0
  Subnet.MulticastPKeyTrapSuppressionEnabled................... : 0x0
  Subnet.Timeout............................................... : 0x11
  LinkSpeed.Supported.......................................... : 0x2
  LinkSpeed.Enabled............................................ : 0x2
  LinkSpeed.Active............................................. : 0x2
  LinkWidth.Supported.......................................... : 0xf
  LinkWidth.Enabled............................................ : 0x8
  LinkWidth.Active............................................. : 0x8
  LinkWidthDowngrade.Supported................................. : 0xf
  LinkWidthDowngrade.Enabled................................... : 0xc
  LinkWidthDowngrade.TxActive.................................. : 0x8
```

INTEL® OMNI-PATH™ FABRIC MANAGEMENT AND TOOLS FEATURES

Local Scope Command Line Tools (pma)

- **opapmaquery
  detailed local PMA info**
  - Multiple otype options
    - e.g. getportstatus, getdatacounters, geterrorinfo, etc.
  - Includes options for per-VL details
    - Counters
    - Errors

```
VL Number    15
        Performance: Transmit
            Xmit Data                        512 MB
            Xmit Pkts                    1613837
        Performance: Receive
            Rcv Data                         722 MB
            Rcv Pkts                     1613843
        Performance: Congestion
            Xmit Wait                          0
            Congestion Discards                0
            Xmit Time Congestion               0
            Mark FECN                          0
            Rcv FECN                           0
            Rcv BECN                           0
        Performance: Bubbles
            Rcv Bubble                         0
            Xmit Wasted BW                     0
            Xmit Wait Data                     0
        Errors: Other
            Xmit Discards                      0
```

```
[RHEL7.1 hds1fnb8301 20160328_0303 ~]# opapmaquery -o getportstatus
Port Number              1
    VL Select Mask        0x00008001
    Performance: Transmit
        Xmit Data                        41759 MB
        Xmit Pkts                     12853526
        MC Xmt Pkts                         29
    Performance: Receive
        Rcv Data                         46746 MB
        Rcv Pkts                      12862230
        MC Rcv Pkts                       3096
    Performance: Congestion
        Xmit Wait                            0
        Congestion Discards                  0
        Xmit Time Congestion                 0
        Mark FECN                            0
        Rcv FECN                             0
        Rcv BECN                             0
    Performance: Bubbles
        Rcv Bubble                           0
        Xmit Wasted BW                       0
        Xmit Wait Data                       0
    Link Qual Indicator                      5 (Excellent)
    Errors: Signal Integrity
        Local Link Integ Err                 0
        Rcv Errors                           0
        Exc. Buffer Overrun                  0
        Link Error Recovery                  0
        Link Downed                          0
        Uncorrectable Errors                 0
        FM Config Errors                     0
    Errors: Security
        Xmit Constraint                      0
        Rcv Constraint                       0
    Errors: Other
        Rcv Sw Relay Err                     0
        Xmit Discards                        0
        Rcv Rmt Phys Err                     0
```

OpenFabrics Alliance Workshop 2016

- **Outline**
  - Overview of Software Components and usage model
  - Management Software Details
    - Management Agents, Administration
    - Fabric Diagnostic and Debug Features
    - Scalable Fabric management GUI
  - Command Line Tools Details
    - Monitoring & Diagnostics
    - Configuration and Management
  - **Usage Examples**
    - **Host verification - opareport**
    - **Performance analysis - opatop**

OpenFabrics Alliance Workshop 2016

'opareport' & 'opatop' – full-featured fabric diag tools

- **opareport**
  - Multiple otype options
    - e.g. comps, bromps, nodes, brnodes, ious, lids, links, extlinks, slowlinks, slowconfiglinks, slowconnlinks, misconfilinks, misconnlinks, errors, otherports, linear, mcast, portusage, pathusage, treepathusage, portgroups, quarantinednodes, validateroutes, validatepgs, validatecreditloops, vfinfo, vfmember, verifyfis, verifysws, verifynodes, verifysms, verifylinks, verifyextlinks, verifyall, all, route, bfrctrl, snapshot, topology, etc.
  - Multiple query options
    - e.g. live vs saved snapshot/topology, query FM or agents direct, output to screen or xml file, rich point syntax for filtered reports, more.

```
[hds1fnb8301 ~]$ opareport -o validatecreditloops
Getting All Node Records...
Done Getting All Node Records
Done Getting All Link Records
Done Getting All Cable Info Records
Done Getting All SM Info Records
Getting All FDB Tables...
Done Getting All FDB Tables
Validate Credit Loop Routes
Done Building All Routes
Fabric summary: 140 devices, 131 HFIs, 9 switches,
        510 connections, 18202 routing decisions,
        17030 analyzed routes, 0 incomplete routes
Done Building Graphical Layout of All Routes
Routes are deadlock free (No credit loops detected)
```

Opareport – viewing a mismatch configuration

Command: opareport –o verifysws –T topologyFile.xml
(Or opalinkanalysis verifysws)

Example of properly configured

```
Getting All Node Records...
Done Getting All Node Records
Done Getting All Link Records
Done Getting All Cable Info Records
Done Getting All SM Info Records
Parsing topology_sw-wip.xml...
SWs Topology Verification

SWs Found with incorrect configuration:
2 of 2 Fabric SWs Checked

SWs Expected but Missing or Duplicate in input:
2 of 2 Input SWs Checked

Total of 0 Incorrect SWs found
0 Missing, 0 Unexpected, 0 Duplicate, 0 Differen
```
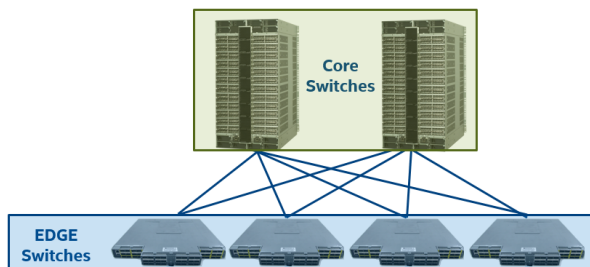
Example with missing cable

```
opareport -o verifylinks -T topology_sw-wip.xml
Getting All Node Records...
Done Getting All Node Records
Done Getting All Link Records
Done Getting All Cable Info Records
Done Getting All SM Info Records
Parsing topology_sw-wip.xml...
Links Topology Verification

Links Found with incorrect configuration:
2 of 2 Fabric Links Checked

Links Expected but Missing, Duplicate in input or Incorrect:
Rate MTU  NodeGUID           Port or PortGUID    Type Name
         Cable: CableLabel           CableLen    CableDetails
100g                         1                   FI phkpstl032 hfi1_0
<->                          37                  SW OmniPth<GUID-Value>
      Cable: 1032h-bd20swp37       1 m
Missing Link

3 of 3 Input Links Checked

Total of 1 Incorrect Links found
1 Missing, 0 Unexpected, 0 Misconnected, 0 Duplicate, 0 Different
```

OpenFabrics Alliance Workshop 2016

Command: opareport –o errors

- **Rapid fabric wide error analysis**
- **Quickly identify any bad cables/ ports**
- **Configurable error thresholds**
- **Point in time, or interval analysis**
- **Concise summary of bad link to check**
  - Name, port#, speeds, etc

```
[RHEL6.5 rh210 20150408_1443 sbin]# opareport -o errors
Getting All Node Records...
Done Getting All Node Records
Done Getting All Link Records
Done Getting All SM Info Records
Getting All Port Counters...
Done Getting All Port Counters
Links with errors > threshold Summary

Configured Error Thresholds:
    LinkQualityIndicator          3
    LinkErrorRecovery             3
    LinkDowned                    3
    PortRcvErrors                 100
    PortXmitDiscards              100
    PortXmitConstraintErrors      10
    PortRcvConstraintErrors       10
    LocalLinkIntegrityErrors      3
    ExcessiveBufferOverruns       3
2304 of 2304 Links Checked, 0 Errors found
--------------------------------------------------------
```

No Error thresholds exceeded on this fabric

'opatop' - Command line TUI for performance analysis

- **Top level view shows fabric, and per PA-group summaries**

- **Multiple levels of drill-down**
  - Study areas of interest, drill down to particular ports

- **Full access to PM on-line history**

- **Review data from the past**
  - Freeze/bookmark a timeslice (image) for study for as long as needed

```
S: Scroll forward (s*) or backward (S*) through port list
C: Select group/VF focus criteria forward (c*) or reverse (C*):
   Utilization (highest first)
   Utilization Packets (highest first)
   Utilization (lowest first)
   Integrity errors (highest first)
   Congestion errors (highest first)
   SmaCongestion errors (highest first)
   Bubble errors (highest first)
   Security errors (highest first)
   Routing errors (highest first)
Nn: Number of entries n in group/VF focus list
Pn: Select port index value n
Press Command Key: u=up (done help) s=scroll fwd S=scroll back <sp>=next line:
```

Can view several types of statistics

"u" key =up
0,1,2 = provides different views

```
opatop: Img:Thu May  8 02:10:57 2014, Live
Summary:  SW:       1 Ports: SW:     38  HFI:      35      Link:     36
          SM:       1 Node Fail:     0 Skip:      0 Port Fail:      0 Skip:      0
                        AvgMBps   MinMBps   MaxMBps   AvgKPps   MinKPps   MaxKPps
0 All           Int        0         0         0         0         0         0
      Integ:min  Congst:min  SmaCong:min  Bubble:min  Secure:min  Routing:min
1 HFIs          Snd        0         0         0         0         0         0
                Rcv        0         0         0         0         0         0
      Integ:min  Congst:min  SmaCong:min  Bubble:min  Secure:min  Routing:min
2 SWs           Int        0         0         0         0         0         0
                Snd        0         0         0         0         0         0
                Rcv        0         0         0         0         0         0
      Integ:min  Congst:min  SmaCong:min  Bubble:min  Secure:min  Routing:min

   Master-SM: LID: 0x0001 Port: 0   Priority: 0  State: Master
             Name: OmniPth00117501ff501ada
             PortGUID: 0x00117500FF501ADA
Secondary-SM: none



Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help |
sS Pmcfg Imginfo View 0-n:
```

```
opatop: Img:Thu May  8 23:35:41 2014, Live
Group Info Sel: All
Int NumPorts: 73  Rate Min:  any  Max: 100g
Ext NumPorts: 0
   Group BW Summary  (W)
   Group Err Summary (E)
   Group Config (C)
```

W,E,C provides different group views

opatop: Img:Fri Mar 26 07:11:22 2015, Bkmk Now:Fri Mar 26 07:32:23 2015
Group Err Stats:HFIs Criteria:Integ Number:10

Ext		Max	0+%	25+%	50+%	75+%	100+%
Integrity	27	2	1	0	0	0	
Congestion	0	3	0	0	0	0	
SmaCongest	0	3	0	0	0	0	
Security	0	3	0	0	0	0	
Routing	0	3	0	0	0	0	

Congest %:0   Discard %:0   Ineffic %:0   Adapt Rt %:0

1 hfi Port has Integrity Errors at least 25% of threshold

Can Drilldown into Details (D)

```
opatop: Img:Thu May  8 23:35:41 2014, Live
Group Info Sel: All
Int NumPorts: 73  Rate Min:  any  Max: 100g
Ext NumPorts: 0
  Group BW Summary (W)
  Group Err Summary (E)
  Group Config (C)
```

W,E,C provides different group views

```
opatop: Img:Fri Mar 26 07:11:22 2010, Bkmk  Now:Fri Mar 26 07:32:23 2010
Group Focus:HFIs  NumPorts:3  Number:10
 Ix LIDx      Port    Integrity    Node GUID 0x          NodeDesc
  0 0001       1          27         0011750000FF8F4C     Intel1A hfi1_0
<-> 0002      20           0         00066A00D900045F     MySwitchR7
  1 0003       1           0         00066A0098006F74     Intel1B hfi1_0
<-> 0002      19           0         00066A00D900045F     MySwitchR7
  2 0004       2           0         0002C90300000798     Intel1C hfi1_0
<-> 0002      18           0         00066A00D900045F     MySwitchR7

Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help | sS cC N0-n P0-n:
```

Node Intel1A has Integrity Errors

Can Drilldown into Port Stats

- **Release Notes and supporting documentation**
  https://downloadcenter.intel.com/download/25770/Intel-Omni-Path-Fabric-Software-Including-Host-Fabric-Interface-Driver-

- **The Intel® Omni-Path Architecture (Intel® OPA) Basic software**
  - Intel® Omni-Path Fabric Host Software
  - Intel® Omni-Path Architecture (Intel® OPA) User tools
  - Installation tools including Text User Interface (TUI)
  - Intel® Omni-Path Host Fabric Interface (Intel® OP HFI) components and drivers
  - Open MPI, MVAPICH2, and Intel® MPI Library optimized for Intel® OP HFIs

- **The Intel® OPA IFS software**
  - Intel® Omni-Path Fabric Suite
  - FastFabric tools including Text User Interface (TUI)
  - Fabric Manager tools

Additional information:
http://www.intel.com/content/www/us/en/high-performance-computing-fabrics/omni-path-architecture-fabric-overview.html

LEGAL DISCLAIMERS

OpenFabrics Alliance Workshop 2016

OPTIMIZATION NOTICE

Optimization Notice

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel.

Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804

12th ANNUAL WORKSHOP 2016

THANK YOU

James Wright

Intel