OPENFABRICS ALLIANCE

OFED for Linux & Windows Status and Plans

Authors:
Linux: Robert J Woodruff; Tziporet Koren
Windows: Eric Lantz; Uri Habusha

Date: 4/4/2011

Agenda

- General Goals and Charter
- EWG – OFED for Linux status update
- WWG – OFED for Windows status update
- How to contribute

EWG and WWG General Goals and Charter

- The charter of the EWG and WWG working groups is to provide enterprise ready distributions of the Open Fabrics code for Windows and Linux
  - Includes providing backports to support several Linux kernel versions and Linux distributions
  - Includes comprehensive testing, validation, and hardening of the code
  - Includes software packaging, release notes, and software installers to allow for easy installation
  - Includes processes for bug tracking and problem resolution

EWG: OFED for Linux

EWG – OFED Linux Status Update

- Releases done in last year:
  - OFED 1.5.1
  - OFED 1.5.2
  - OFED 1.5.3
- 2011 Plans
  - OFED 1.6

OFED 1.5.1

- Released on 3/25/2010
- Main new features:
  - Added RoCEE support
  - Added enhanced atomic operations to ConnectX (kernel only).
  - Updated Open MPI to rev 1.4.1-2ofed
  - Updated MVAPICH2 to rev 1.4.1
  - Updated User space libraries:
    - DAPL, libnes, librdmacm
  - Removed tvflash RPM
  - Fixed IPv6 support and IPv4 routing corner cases for RDMA CM

OFED 1.5.2

- Released on 9/21/2010
- Last release supporting RHEL 4.x
- Main new features:
  - Added RAW Ethernet QP support (nes & mlx4)
  - Added multicast support in performance tests
  - RoCE in GA level
  - Added new package: ibacm
  - Updated user space packages
  - Management: Moved to OpenSM to rev 3.3.7
  - MPI: updated Open MPI, MVAPICH2 and MPI tests
  - SDP: Improved Zero copy stability and performance
  - NFS-RDMA: Removed from default installation

OFED 1.5.3

- Released on 3/10/2011
- Main new features:
  - Added RHEL 5.6 & 6 Support
  - Updated user space libraries
  - Update MVAPICH2 package
  - Improved RAW Ethernet QP support for mlx4 and enhanced mlx4_en steering mode
  - Improved SDP latency by usage of inline

OFED 1.6 Schedule

- Ongoing work on backports - during Q2
- First RC - middle of June
  - RCs every 2 weeks
- GA - middle of September

OFED 1.6 Features

- Kernel base: 2.6.38
- Remove MPI packages from OFED
  - EWG will agree on common packages for testing
- OSes support:
  - The latest OSes will be supported:
    - RHEL 5.6 & 5.7
    - RHEL 6 & 6.1
    - SLES 11 SP1 & SP2
  - Can we drop the support of SLES 10?
- FDR support
- User space FMRs
- CMA APM support

OFED 1.6 Features – Cont.

- SRIOV support for mlx4 with CX2 & CX3
  - Will be supported with KVM
- NFSRDMA will be supported with limited backports
  - RHEL6, SLES11sp1
- OpenSM main improvements:
  - Torus-2QoS routing engine
  - Performance Manager improvements: improved redirection and extended counters support
  - Additional port balancing options for routing
  - SRIOV support
  - Extended link speed support
- New Hardware support
  - Chelsio's T4 adapter (iw_cxgb4/cxgb4).
  - Mellanox's ConnectX3 support

WWG: OFED for Windows

OFED For Windows 2.3

- **Current OFW version: 2.3**
- **Release Date: December 28, 2010**
- Supports: Windows 7, Server 2008 R2 and HPC, Server 2008, Server 2003
- Logo-Tested source
  - Significant coverage: NetworkDirect, MPI, Benchmarks, Commercial Apps
  - Only hardware suppliers can logo binaries
- NDIS 6.1 IPoIB for Server 2008/R2, Windows 7,Vista
- NDIS 5.1 IPoIB for Server 2003 & XP
- NetworkDirect (Windows RDMA) provider

OFED For Windows 2.3 (cont.)

- OFED verbs library enables easy porting of Linux OFED applications into the OFED for Windows environment.
- uDAT/DAPL is now a common code base with OFED uDAT/DAPL version 2.0.30.
- Bug fixes for stability in IBcore, ConnectX, NetworkDirect, VNIC, IPoIB, DAT/DAPL
- OpenSM upgraded to version 3.3.6 (umad vendor).
- Installation Methods Supported (no change)
  - “Double-Click, Install Wizard”= MSI Package (WIX 3.0 compliant)
  - Driver injection in Windows Deployment Service (OS Imaging)

OFED for Windows Results

- Efficiency:
  - 90.1% cluster efficiency (HPL) on 512-core, 64-node cluster of Nehalem cores with QDR (40Gb/s) Infiniband
- Scalability:
  - 5 Windows machines in the TOP100
    - All used WinOFED-derived drivers
  - Achieved >1 PetaFLOPS with current WinOFED driver
    - Tokyo Institute of Technology (TSUBAME 2)
    - Result: 1.127PF on 1296 GPU-enabled nodes (1 of 7 >1PF systems)
    - GPGPU now table stakes
      - » Kazushige Goto & Laurent Visconti created/tuned our hybrid xhpl
  - Two of the Top 5 Green500 use WinOFED driver
    - #3 – TSUBAME 2.0, Tokyo Tech
    - #5 – Jazz, CASPUR consortium (Italy)

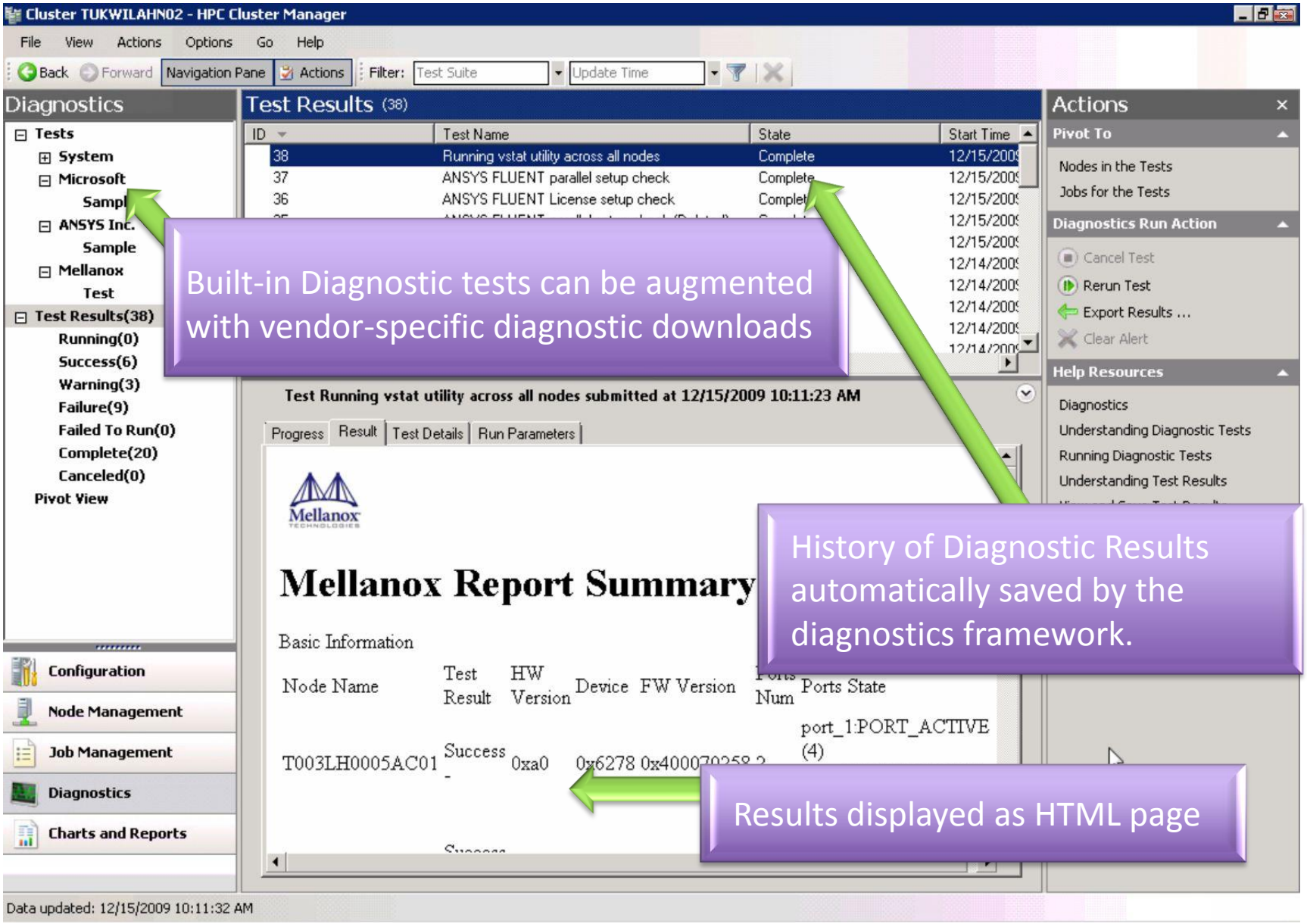
Next Release: OFED for Windows 3.0

- Dates:
  - freeze Q2'11
  - release target Q3'11
  - HPC Server V3 SP2 June ‘11
- “Working” Feature List
  - ROCE support
  - NetworkDirect v2 provider (likely postponed)
  - OpenSM 3.3.9 (vendor umad)
  - IPoIB CM (Connected Mode) support.
  - DAT/DAPL 2.0.35
  - Fourteen Data Rate (FDR) support

OFED for Windows wiki pages

- Temporarily Offline
- Back online by Sept 2011

HPC Server 2008 R2 Diagnostics Framework



Built-in Diagnostic tests can be augmented with vendor-specific diagnostic downloads

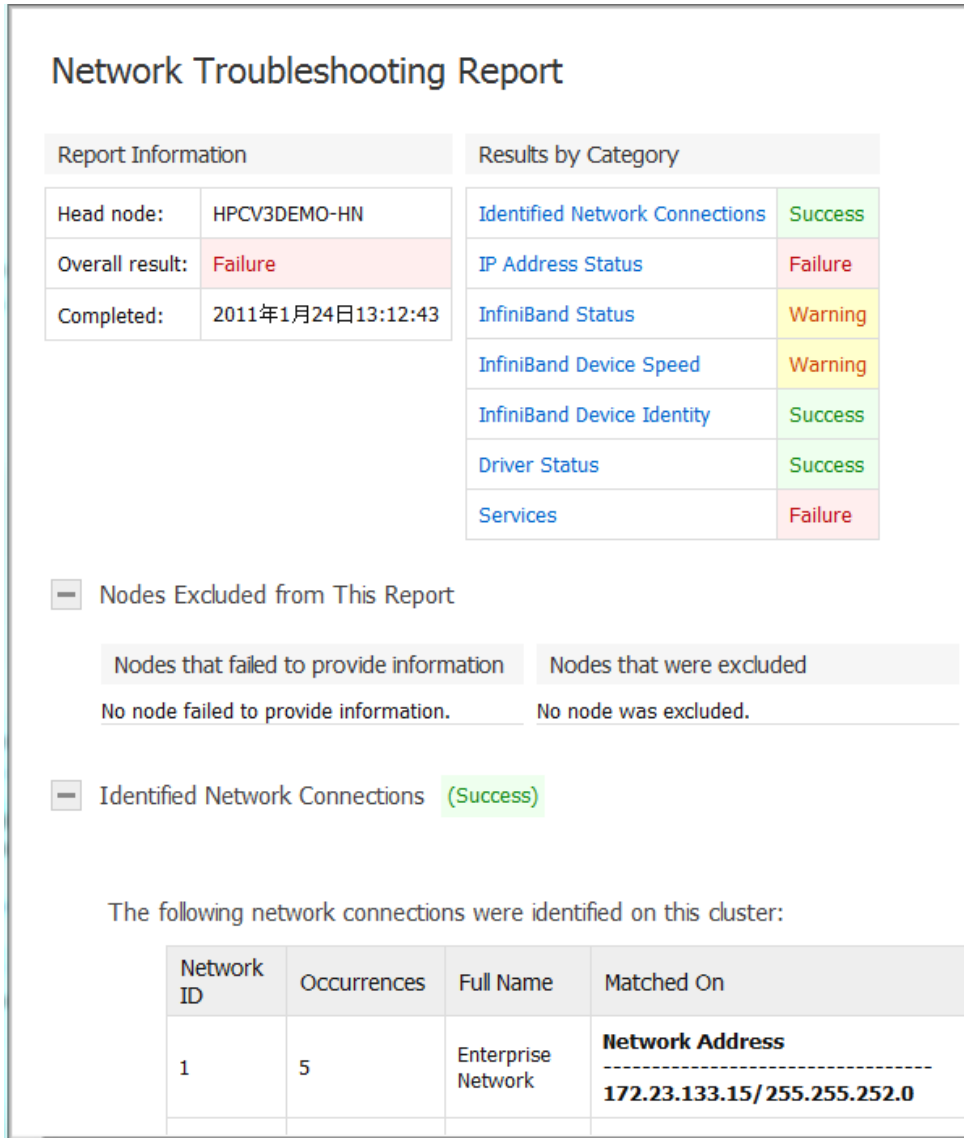
History of Diagnostic Results automatically saved by the diagnostics framework.

Results displayed as HTML page

New Network Troubleshooting Plug-In

300 DOWNLOADS IN FIRST 3 MONTHS

- No-cost download at MS.com
- Cluster-wide config/perf consistency checking
- Identifies and reports:
  - NIC configuration (NOT ibdiagnet)
  - all subnets in a cluster
  - HTML output



Network Troubleshooting Report

Report Information	
Head node:	HPCV3DEMO-HN
Overall result:	Failure
Completed:	2011年1月24日13:12:43

Results by Category	
Identified Network Connections	Success
IP Address Status	Failure
InfiniBand Status	Warning
InfiniBand Device Speed	Warning
InfiniBand Device Identity	Success
Driver Status	Success
Services	Failure

Nodes Excluded from This Report

Nodes that failed to provide information	Nodes that were excluded
No node failed to provide information.	No node was excluded.

Identified Network Connections (Success)

The following network connections were identified on this cluster:

Network ID	Occurrences	Full Name	Matched On
1	5	Enterprise Network	**Network Address** ---------- **172.23.133.15/255.255.252.0**

If You Want to Help….

- Developing code:
  - Including back-ports in Linux
- Doing QA and testing
- Performance tuning
- Sending patches and comments to the mailing lists:
  - **OFED for Windows: ofw@lists.openfabrics.org**
  - **OFED for Linux: ewg@lists.openfabrics.org**
  - **General Linux development: linux-rdma@vger.kernel.org**
- Participate in EWG/WWG meetings
- Opening bugs in Bugzilla (https://bugs.openfabrics.org/)
  - When opening a new bug you can choose OpenFabrics Windows or OpenFabrics Linux

OPENFABRICS ALLIANCE

Thank You!