OFED for Linux
Status and Plans

Authors: Robert J Woodruff; Rupert Dance

Date: 3/17/2015

#OFADevWorkshop
Agenda

• General Goals and Charter
• EWG – OFED releases last year
• Process Review and Changes
• EWG Roadmap
• OFED H/W and S/W validation
• How to contribute
The charter of the EWG working group is to provide enterprise ready distributions of the Open Fabrics code for 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
- Provides platform for experimental OFA technologies
- Used for interoperability testing and OFA logo program
EWG – OFED Linux Status Update

• Production releases done since last year:
  – OFED 3.12
  – OFED 3.12-1
OFED 3.12

- Based on kernel.org 3.12 kernel
- OSes supported
  - RHEL EL 6.4, RHEL EL 6.5
  - SLES 11 SP 3
  - Kernel.org 3.12
- Main new features:
  - Mellanox Connect-IB (mlx5) support
  - Emulex RoCe NIC (ocrdma) support
  - Updated user-space packages
    - dapl-2.0.40, ibsim-0.6, infiniband-diags-1.6.4, infinipath-psm-3.2-2_ga8c3e3e_open, libcxgb4-1.3.2, libibmad-1.3.11, libibumad-1.3.9 libipathverbs-1.2.1, libnes-1.1.4, librdmacm-1.0.18, mstflint-3.5.0 opensm-3.3.17, perftest-2.0-0.80.g54c73c6, srptools-1.0.2
OFED 3.12-1

• Based on kernel.org 3.12 kernel
• OSes supported
  – RHEL EL 6.5, RHEL EL 6.6, RHEL 7.0
  – SLES 12 and SLES 11 SP 3
  – Kernel.org 3.12
• Main new features:
  – Added support for Intel® Xeon Phi™ to OFED mainline as an experimental feature
  – Updated user-space packages (added or updated)
    • dapl-2.1.2, ibacm-1.0.9, ibpd-1.0.0-2, infinipath-psm-3.3-2, libcxgb4-1.3.5, libibscif-1.0.0, libibverbs-1.1.8, libmlx4-1.0.6, libocrdma-1.0.4, librdmacm-1.0.19.1, mstflint-3.7.1-1.22, opensm-3.3.18, perftest-2.3-0.12
OFA Process Changes

• Starting with OFED-3.5, the code is based on upstream kernel and user-space packages.
  – Only fixes accepted upstream are included
  – Changes were designed to allow better alignment with distro and upstream code base to reduce fragmentation
  – Companies can provide experimental features based on OFED.

• Initially we had experimental features added as separate branch OFED releases, e.g., OFED-3.5-x-MIC.
  • Got feedback this was causing more fragmentation, not less
  – Changed the process to allow experimental features back in the OFED mainline, but not enabled by default
  • Added Intel® Xeon Phi™ support in OFED-3.12-1
OFA Process change – cont’d

• Distro integration and alignment
  – Most production level users can probably just use what is in the distros
    • Only users needing bleeding edge or experimental features that are not yet upstream and in the distros need to run the latest OFED
  – OFED content and distro OFA content are now more closely aligned than ever
    • SLES 12 and OFED-3.12 based on same kernel code and user-space packages
    • RHEL EL 7.x base on kernel.org 3.10 base with some newer patches backported and upstream user-space packages
  – Are there other changes and improvements we can make?

• End Users
  – Are there other changes and improvements we can make?
OFA Process change – cont’d

• Coordination with new working groups
  – OFIWG
    • Open Fabrics Interfaces (OFI) working group
    • libfabric included in OFED 3.18
  – OFVWG
    • Enhancing the verbs interface to meet the needs of future systems and applications
    • As features get developed/released, they will feed into OFED using the existing OFA development processes
OFED Roadmap.

• OFED-3.18
  – Added support for new distros, RHEL EL 7.1
  – Added OFI 1.0 (libfabric) from the OFIWG
  – Mellanox ConnectX4 ready
  – Updated user-space packages and cherry picked bug-fixes from later upstream kernel.org
  – Currently under development and at RC1
  – Likely will release in Q2, 2015

• What’s after OFED-3.18 ? OFED-4.xx
  – Likely based on kernel.org 4.0 or 4.1
  – Support for RoCE v2
  – Next OFI 1.x package
  – Other features TBD
  – Likely release in 2H of 2015
OFED – SW & HW validation

- **OpenFabrics Interoperability Logo Group (OFILG)**
  - **Purpose**: validate OFED functionality, test ULPs and verify interoperability in a heterogeneous environment
  - **Members**: Chelsio, DDN, Emulex, IBM, Intel, Mellanox and NetApp

- **OFA Cluster at UNH-IOL**
  - **Servers**: iWARP 12 hosts, InfiniBand 18 hosts, RoCE 10 hosts
  - **InfiniBand HW**: 12 HCAs, 5 switches, 5 SRP targets, 1 gateway
  - **iWARP HW**: 6 RNICs, 2 switches
  - **RoCE HW**: 8 RCA, 2 switches

- **OFED versions Tested**
  - 1.5.x, 3.5.x, 3.12, 3.12-1 and 3.18

- **PXE Boot environment available**
  - RHEL 5.x, 6.x, 7.x, SLES 11, Ubuntu 10.04 and 12.04
  - OFED 1.4.x, 1.5.x, 3.5.x, 3.12, 3.12-1

- **Protocols Tested**
  - Fabric Init, IPoIB, Link Init, NFSoRDMA, Open MPI, RDMA Utilities, RSocket, SM failover, SRP, uDAPL
  - Tests executed – approximately 8,049
Ethernet – transport

iWARP Logo Event Topology

Chelsio T4 CX4 (10G) leda
Chelsio T4 CX4 (10G) hyperion
Chelsio T4 SFP+ (10G) mimas
Chelsio T4 SFP+ (10G) rhea
Chelsio T5 SFP+ (10G) elara

Fujitsu XG2000C 10G Switch

Intel NE020 CX4 (10G) polydeuces
Intel NE020 CX4 (10G) prometheus
Intel NE020 CX4 (10G) fenrir
Intel NE020 CX4 (10G) anthe

RoCE Logo Event Topology

Mellanox Switch SX1012
12-port 40Gb/s

Arista Networks Switch 7050 QX
32-port 40Gb/s

IBM Power 7 R2 w/ Mellanox MT26448 SFP+ 10Gb/s

Mellanox Switch MCX312B-XCCT
SFP+ 10Gb/s

Mellanox MCX314A-BCCT QSFP+ 40Gb/s atlas

Emulex OCF14102 SFP+ 10Gb/s

Mellanox MCX314A-BCCT QSFP+ 40Gb/s

Emulex OCF14401 QSFP+ 40Gb/s

March 15 – March 18, 2015
#OFADevWorkshop
If You Want to Help….

- Developing code:
  - Including back-ports in Linux
  - Reviewing code submitted to Linux kernel/libs
- Doing QA and testing
- Performance tuning
- Sending patches and comments to the mailing lists:
  - OFED for Linux: ewg@lists.openfabrics.org
  - General Linux development: linux-rdma@vger.kernel.org
  - OFIWG Working Group: ofiwg@lists.openfabrics.org
  - OFVWG Working Group: ofvwg@lists.openfabrics.org
  - Maintainers and git trees: http://www.openfabrics.org/downloads/MAINTAINERS
- Participate in EWG meetings
- Opening bugs in Bugzilla (http://bugs.openfabrics.org/)
  - When opening a new bug you can choose OpenFabrics Linux
Thank You