

Lawrence Livermore National Laboratory

Open Fabrics Requirements

26 March 2012



Ira Weiny

Lawrence Livermore National Laboratory, P. O. Box 808, Livermore, CA 94551
This work performed under the auspices of the U.S. Department of Energy by
Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344

LLNL-PRES-542791

OFED vs Open Fabrics

Open Fabrics Enterprise Distribution (OFED) was supposed to be the vehicle to wide spread acceptance, and upstream adoption, of a standard reference implementation of “RDMA”.

Open Fabrics is the organization who facilitated that development.



OFED Successes

- Now in upstream kernel
- Accepted by most major distros
 - RHEL
 - SuSe
 - Ubuntu (many packages are old)



OFED Improvements 2011

- New Kernel development process
 - OFED 3.2
- Upstream XRC
 - work has been accepted in the kernel
 - Not yet accepted for upstream libibverbs
- Web page looks much better



Future OFED

- OFED no longer “enterprise”
- Make OFED -- “Tech Preview”
- Break up user space and kernel support
 - OFED is most valuable as backports from upstream to the distros
 - User space packages should be made to stand on their own.
 - I know my packages are 2 of the worst offenders. :-)



Final comment on OFED

As long as Open Fabrics continues to advertise OFED as the “standard” to RDMA (high speed networking) they undermine the goal of becoming the standard.

Only inclusion in the upstream kernel and the distros makes you the **standard**.



So what should Open
Fabrics do now?



Champion “RDMA” and alternate networking protocols

- Maintain documentation
 - Web page
 - Training
 - Mailing list presence
- Help standardize other protocols through the web site and advertising?



Software Development

- Git trees
- Development mailing list
- Kernel backports
- Continued user space package development
 - Separate from “OFED”
 - Make them stand on their own



Software Development

- Develop “tech preview” features for testing.
- Provide the single point of access to whats coming.

We have a good community here. We should foster that.



High speed software interfaces

- For Users
 - MPI scalability issues
 - Alternatives to MPI?
 - CCI
 - PGAS/SHMEM



Management

- For System Administrators
 - Better SM
 - Better diag tools
 - Better documentation



Management

- Open Flow?



LLNL Requirements

“Exascale”, yea!



Exascale

- Scalability
- Scalability
- Scalability
- ...



Exascale

- LLNL needs a scalable, low latency, high bandwidth, **commodity** network
- What is “commodity” (definition)
 - ...goods or services [that have lost it's] market ... differentiation across its supply base, often by the diffusion of the intellectual capital necessary to acquire or produce it efficiently... [ref: wikipedia]



Exascale

- But at the same time we want
 - Adaptive routing
 - Collective Offload
 - Enhanced hardware integration
 - GPGPU, memory, PCIe bypass
 - Anything which enhances scalability
 - Alternate topologies
- How do we resolve these conflicting requirements?



LLNL's requirements

LLNL will use what ever is “best value”. This includes ease of operation as well as ultimate performance.



LLNL's requirements

Finally, we are committed to Open Source.

Open Source solutions provide the best return on investment without vendor lock in using commodity hardware.

