NFS-RDMA Status

James Lentini
Network Appliance
jlentini@netapp.com
11/15/2007

NFS-RDMA is...

- A binding to an RPC-layer protocol, RPC-RDMA, that allows NFS to use RDMA networks (such as Infiniband and iWARP)
- A transparent solution for applications, NFS protocol features, and NFS users
- A significant performance boost to clients
  - Reduces client CPU overhead
  - Utilizes high-bandwidth, low-latency fabrics
- The storage solution for single-wire RDMA clustering

Linux Client and Server (1)

- Client in Linux 2.6.24
  - Requires mount.nfs command supporting the "string" mount API (nfs-utils-1.1.1 or greater)
- Server expected in Linux **soon** (2.6.25?)
  - Development trees at http://linux-nfs.org/cgi-bin/gitweb.cgi
  - Discussion on nfs@lists.sourceforge.net
- Step-by-step guide to Linux NFS-RDMA at: http://nfs-rdma.sourceforge.net/Documents/README

Linux Client and Server (2)

- Full implementation supporting NFSv2, v3, v4
- Excellent performance results obtained
  - Full IB DDR bandwidth, see Helen Chen's presentation at OFA Sonoma 2007
  - NFS-RDMA pushes the performance limits of Linux, see Batsakis, et al. *Enhancing the Linux Memory Architecture to Support File Systems over Heterogeneous Devices* at the 2007 Linux Storage & Filesystem Workshop

OFED NFS-RDMA Support

- Tom Tucker created a supplemental NFS-RDMA package for OFED 1.2, see http://www.opengridcomputing.com
- Jeff Becker (Jeffrey.C.Becker@nasa.gov) has volunteered to create NFS-RDMA packages for OFED 1.3
- NFS-RDMA inclusion in mainline Linux simplifies OFED support

Standardization

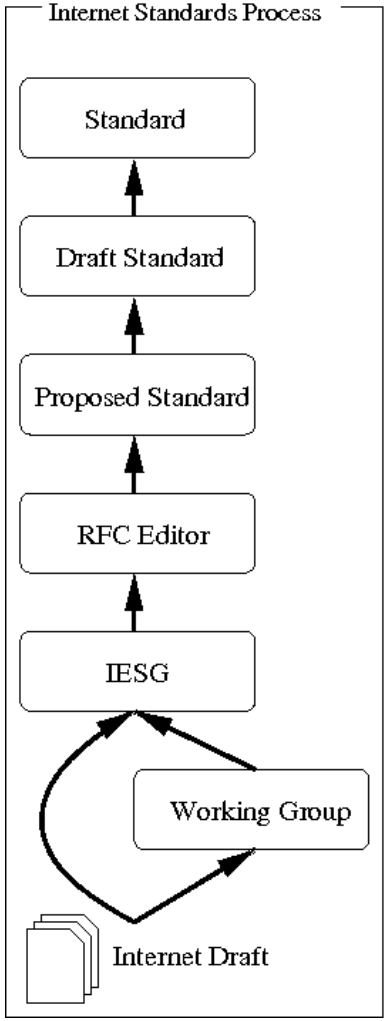
- **Protocol Standardization**: three documents in the IETF NFSv4 Working Group:

  http://www.ietf.org/html.charters/nfsv4-charter.html

  - NFS RDMA Problem Statement
  - RDMA Transport for ONC RPC: Describes the RPC-RDMA protocol for sending RPC messages on an RDMA transport (IB, iWARP)
  - NFS Direct Data Placement: Describes the NFSv2/v3/v4 mapping to RPC-RDMA operations

  All documents are in IESG review

- **Multiple Interoperable Implementations**: Linux client/server interoperate with OpenSolaris client/server



Further Reading (1)

- Talpey: Tutorial at FAST '07
- Nwokah: Deployment at Low Latency Technical Forum 2007
- Chen: Performance at OFA Sonoma 2007
- Lentini and Tucker: Protocol and Implementation at OFA Sonoma 2006

Further Reading (2)

- R. Noronha, L. Chai, S. Shepler and D. K. Panda, *Enhancing the Performance of NFSv4 with RDMA*, Int'l Workshop on Storage Network Architecture and Parallel I/Os (SNAPI'07), Septenber 2007.
- R. Noronha, L. Chai, T. Talpey and D. K. Panda, *Designing NFS With RDMA For Security, Performance and Scalability*, Int'l Conference on Parallel Processing, Xian, China, September 2007.