Linux NFS/RDMA Roadmap
Chuck Lever, Oracle

#OFADevWorkshop
Today’s Topics

- Linux NFS/RDMA community accomplishments
- Plans for NFS/RDMA in mainline Linux

This is not a discussion about any Oracle product or feature
Community Mission

• Revive NFS/RDMA capability in the mainline Linux NFS client and server

• Ensure capability is interoperable, robust, and performant
Community Mission

- Provide a platform for prototyping and validating relevant protocol specifications and updates
- Demonstrate a storage ULP that works well on whole family of RDMA fabrics
2014 In Review

• Stabilize mainline client and server
• Build out test infrastructure
• Identify a server-side maintainer
2014 In Review

- Three remaining memory registration modes
  - FRWR: Fast, but not supported everywhere
  - FMR: Still important in some environments
  - PHYSICAL: “Always works” fallback
Client Interoperability Goals

- Existing implementations
  - Linux server
  - Solaris server
- Expected soon
  - Ganesha
pNFS In 30 Seconds

pNFS client

NFS or block protocols

pNFS Data Stores

pNFS layouts describe to clients where file data resides

- Point each client at a unique data store
- Each file can reside on a different server
- Server can expose its backend storage to clients
- Split namespace across servers
- Stripe file contents across servers
- Data migration, replication, and load balancing
Enabling pNFS on RDMA

- pNFS requires NFSv4.1

- NFSv4.1 requires a backchannel
  - Sidecar
  - Bi-directional RPC/RDMA
Standards Work

- RFC 5666: Remote Direct Memory Access Transport for Remote Procedure Call
  - Clarify how to encode NFS COMPOUND
  - Clarify how to encode RDMA_NOMSG calls
  - Bi-directional RPC/RDMA?
Standards Work

- RFC 5667: Network File System (NFS) Direct Data Placement
  - Update NFSv4.0 and NFSv4.1 RPC/RDMA bindings
  - Create bindings for pNFS layouts and NFSv4.2
The Year To Come

- NFSv4.1 on RDMA
- NFS/RDMA in domU, containers, kvm
- Close out interoperability issues
- Continue to improve performance and efficiency
Longer Term

- Scaling number of mount points per client
- HCA hot-plug
- Persistent memory file systems on NFS servers
- Reducing whole stack latency
Is It Ready?

• It Depends.
  • Your distribution
  • Your adapters and fabric
  • Your requirements
  • Client only, or client and server?
Thank You