# pNFS Status

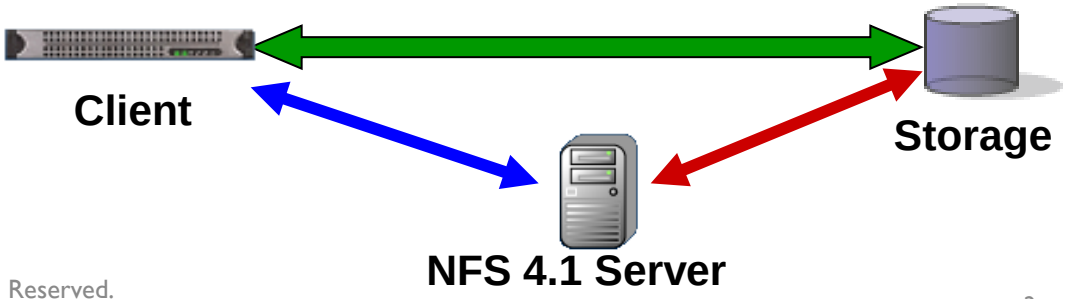
SDC September, 2010

Brent Welch, Panasas

# The pNFS Standard



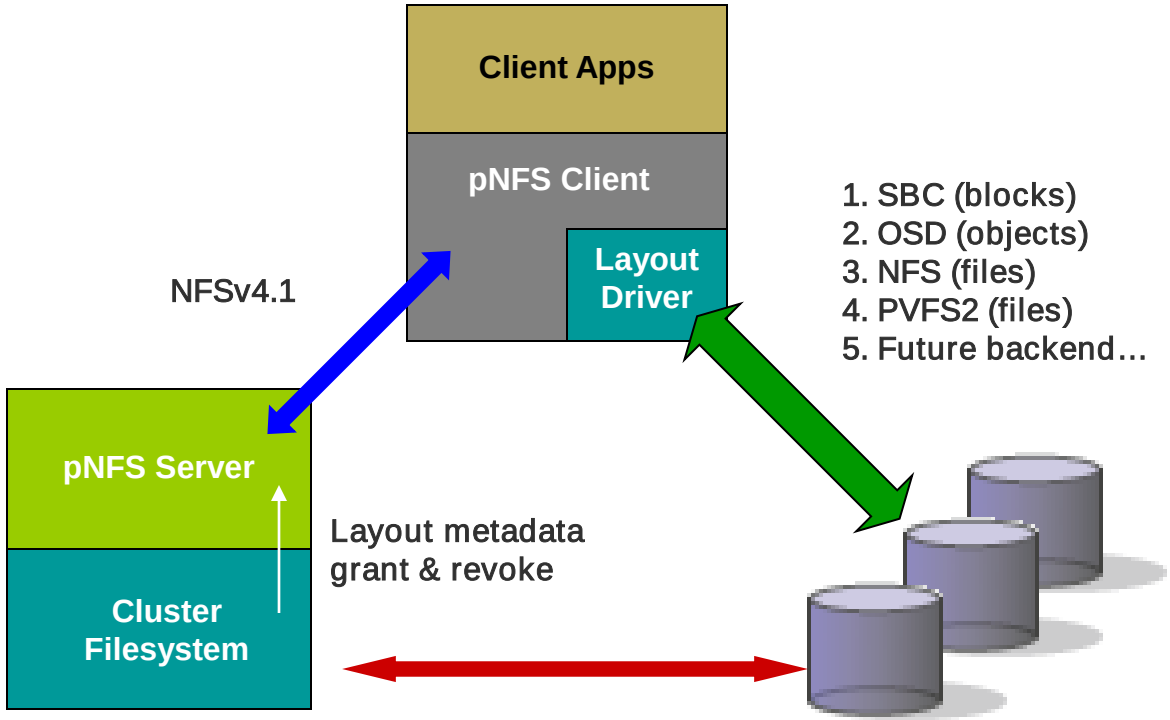
- The **pNFS** standard defines the NFSv4.1 protocol extensions between the **server** and **client**
- The **I/O** protocol between the **client** and **storage** is specified elsewhere, for example:
  - SCSI **Block** Commands (**SBC**) over Fibre Channel (**FC**)
  - SCSI **Object**-based Storage Device (**OSD**) over iSCSI
  - Network **File** System (**NFS**)
- The **control** protocol between the **server** and **storage** devices is also specified elsewhere, for example:
  - SCSI **Object**-based Storage Device (**OSD**) over iSCSI



# pNFS Client



- Common client for different storage back ends
- Wider availability across operating systems
- Fewer support issues for storage vendors



# Key pNFS Participants









- Panasas (Objects)
- Network Appliance (Files over NFSv4)
- IBM (Files, based on GPFS)
- EMC (Blocks, HighRoad MPFSi)
- Sun/Oracle (Files over NFSv4)
- U of Michigan/CITI (Files over PVFS2)

# Standards process milestone



- 2003 First pNFS meeting among vendors
- 2005 First IETF drafts
- 2008 Approval of drafts for standard track
- 2010 RFC status achieved!
  - 5661: NFSv4.1 protocol
  - 5662: NFSv4.1 XDR Representation
  - 5663: pNFS Block/Volume Layout
  - 5664: pNFS Objects Operation

# pNFS Availability



- pNFS is part of the IETF NFSv4 minor version 1 standard
  - RFCs issued in January 2010 after 10 month review period
- Linux pNFS implementation available "out of tree" from the pNFS developers
  - Git tree hosted at open-osd.org (sponsored by Panasas)
  - RedHat generates experimental RPMs from this tree
- Steady rate of patch adoption into main Linux source tree
  - Details on subsequent slides

# pNFS Implementation



- NFSv4.1 mandatory features have priority
  - RPC session layer giving reliable at-most-once semantics, channel bonding, RDMA
  - Server callback channel
  - Server crash recovery
  - Other details
- EXOFS object-based file system (file system over OSD)
  - In kernel module since 2.6.29 (2008)
  - Export of this file system via pNFS server protocols
  - Simple striping (RAID-0), mirroring (RAID-1), and now RAID-5 in progress
  - "Most stable and scalable implementation"
- Files (NFSv4 data server) implementation
  - Server based on GFS
  - Layout recall not required due to nature of underlying cluster file system
- Blocks implementation
  - Server in user-level process, FUSE support desirable
  - Sponsored by EMC

# Calibrating My Predictions



- 2006
  - "TBD behind adoption of NFS 4.0 and pNFS implementations"
- 2007 September
  - Anticipate working group "last call" this October
  - Anticipate RFC being published late Q1 2008
  - Expect vendor announcements after the RFC is published
- 2008 November (SC08)
  - IETF working group last call complete, area director approval
  - *(Linux patch adoption process really just getting started)*
- 2009 November (SC09)
  - Basic NFSv4.1 features 2H2009
  - NFSv4.1 pNFS and layout drivers by 1H2010
  - Linux distributions shipping supported pNFS in 2010, 2011

# Linux Development 2008



- January
  - pNFS patches are against 2.6.18
  - Linux head-of-line is 2.6.24
  - Benny Halevy (Panasas) assumes defacto gatekeeper role
- June
  - In rhythm with merges and forward porting pNFS patches (2.6.25)
  - iSCSI/OSD patches in active review
- December
  - iSCSI/OSD patches submitted for 2.6.29 merge window
  - EXOFS implementation underway

# Linux Release Cycle 2009



- 2.6.30
  - Merge window March 2009
  - RPC sessions, NVSv4.1 server, OSDv2 rev5, EXOFS
- 2.6.31
  - Merge window June 2009
  - NFSv4.1 client, sans pNFS
- 2.6.32
  - Merge window September 2009
  - 130 server-side patches add back-channel
- 2.6.33
  - Merge window December 2009, released Feb 2010
  - 43 pNFS patches

# Linux Release Cycle 2010



- 2.6.34
  - Merge window February 2010, Released May 2010
  - 21 NFS 4.1 patches
- 2.6.35
  - Merge window May 2010, release August? 2010
  - 1 client and 1 server patch (4.1 support)
- 2.6.36
  - Merge window August 2010
  - 16 patches accepted into the merge
- 2.6.37 preparations
  - 290 patches represent pNFS functionality
  - Working on strategy to review and merge
  - Finalizing patches before October Bake-a-thon testing session

# Linux Release Cycle 2011



- 2.6.37
  - Merge window November? 2010
  - Files pNFS client and server
- 2.6.38
  - Merge window February? 2011
  - Object pNFS client and server
- 2.6.39
  - Merge window May? 2011
  - Blocks client and server?

# How to use pNFS today



- Benny's git tree <bhalevy@panasas.com>:

  git://linux-nfs.org/~bhalevy/linux-pnfs.git

- The rpms <steved@redhat.com>:

  http://fedorapeople.org/~steved/repos/pnfs/i686

  http://fedorapeople.org/~steved/repos/pnfs/x86_64

  http://fedorapeople.org/~steved/repos/pnfs/source/

- Bug database <pnfs@linux-nfs.org>

  https://bugzilla.linux-nfs.org/index.cgi

- OSD target

  http://open-osd.org/

# Thank you for supporting pNFS!



- pNFS benefits substantially from the support by ESSC/DoD
  - As a small company, Panasas uses its resources carefully
  - pNFS is a long range investment for the whole storage community
    - pNFS is not identical to Panasas proprietary protocols
  - Their support has made it possible to continue our efforts toward pNFS adoption by the broader market