

# pNFS Status

Brent Welch, Panasas

HEC FSIO Conference

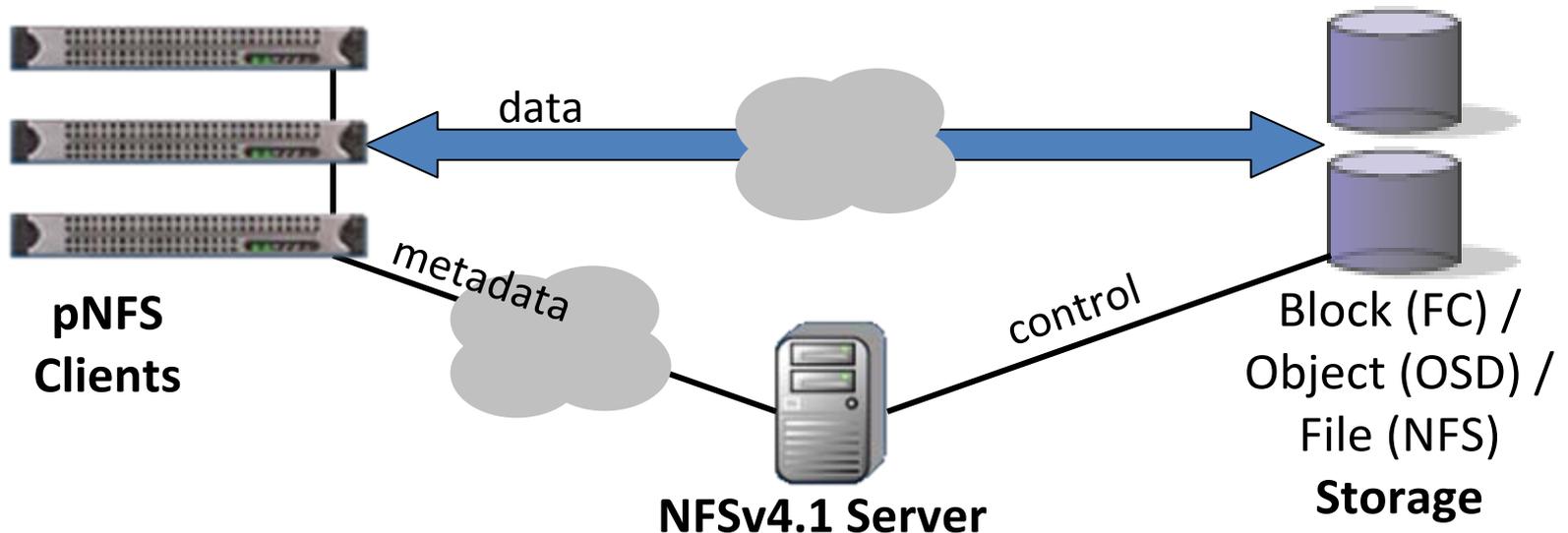
August 10, 2009

# pNFS Background

- pNFS adds parallel I/O to the NFS protocol
  - Eliminate file server bottleneck
  - Provide parallel data paths, even for a single file
- pNFS is part of NFSv4.1
  - Standard approved by IETF Dec, 2008
  - RFC number awaiting RFC editorial review

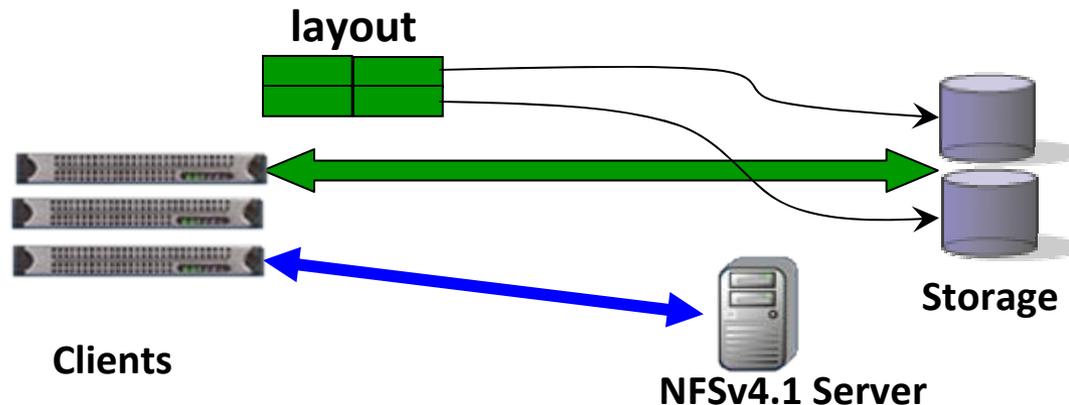
# pNFS: Standard Storage Clusters

- pNFS is an extension to the Network File System v4 protocol standard
- Allows for parallel and direct access
  - From Parallel Network File System clients
  - To Storage Devices over multiple storage protocols
  - Moves the Network File System server out of the data path



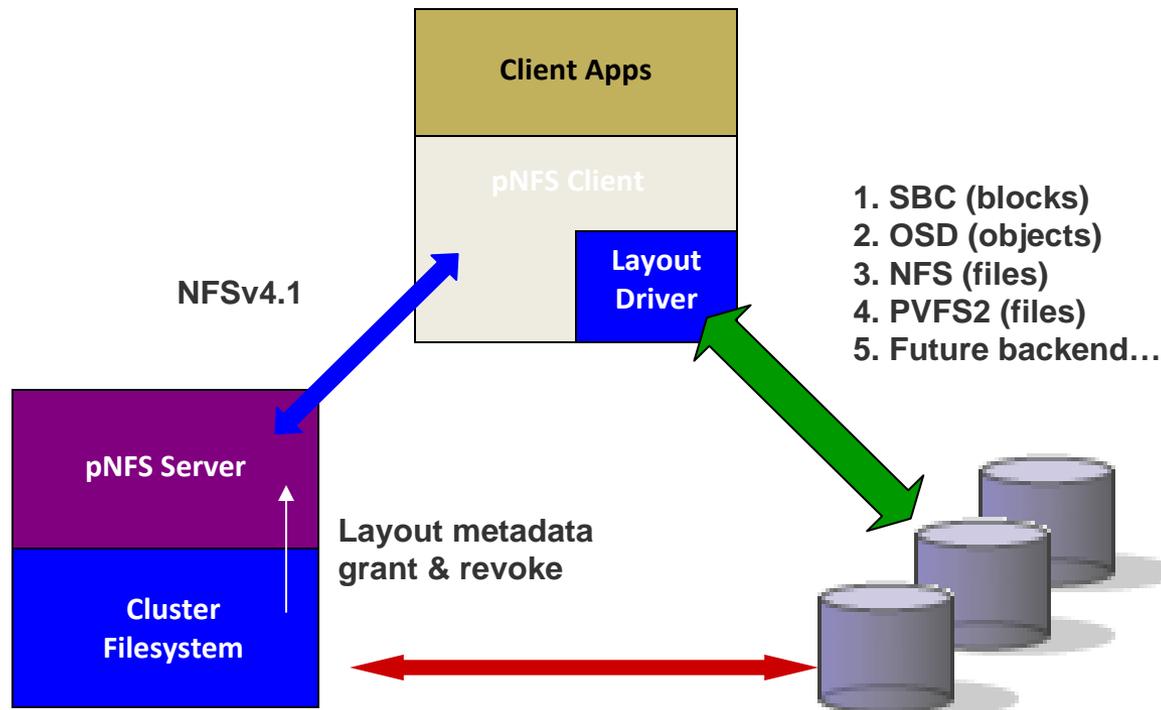
# pNFS Layouts

- Client gets a *layout* from the NFS Server
- The layout maps the file onto storage devices and addresses
- The client uses the layout to perform direct I/O to storage
- At any time the server can recall the layout
- Client commits changes and returns the layout when it's done
- pNFS is optional, the client can always use regular NFSv4 I/O



# pNFS Client

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



# pNFS Status

- Implementation interoperability continues
  - San Jose Connect-a-thon March '06, February '07, May '08, June '09
  - Ann Arbor NFS Bake-a-thon September '06, October '07
  - Dallas pNFS inter-op, June '07, Austin February '08, Sept '08, **October '09**
- Server vendors waiting for Linux client
  - Sun, NetApp, EMC, IBM, Panasas, ...
  - 2.6.30 – exofs object storage file system (local) and iSCSI/OSDv2
  - 2.6.31 is in stabilization.
    - most of nfsv4.1: sessions, 4.1 as an option, no pnfs yet
    - server back channel is absent
  - 2.6.32 merge, finish nfsv4.1 including server callbacks
    - Hopefully some of the internal APIs for pNFS.
    - Merge window opens in September
  - Goal to complete patch adoption by Q1 2010

# How to use pNFS today

- Up-to-date GIT tree from Linux developers
  - [bhalevy@panasas.com](mailto:bhalevy@panasas.com) manages the source trees
- RedHat fedora RPMs that include pNFS
  - [steved@redhat.com](mailto:steved@redhat.com) builds experimental packages
- pNFS mailing list, [pnfs@linux-nfs.org](mailto:pnfs@linux-nfs.org)
- <http://open-osd.org>
  - Useful to get to OSD target, the user level program
  - Exofs uses kernel initiator, need the target

# How to use pNFS today

- Benny's git tree:  
`git://linux-nfs.org/~bhalevy/linux-pnfs.git`
- The the kernel rpms can be found at:  
`http://fedorapeople.org/~steved/repos/pnfs/i686`  
`http://fedorapeople.org/~steved/repos/pnfs/x86_64`
- The source rpm can be found at:  
<http://fedorapeople.org/~steved/repos/pnfs/source/>
- Bug database  
<https://bugzilla.linux-nfs.org/index.cgi>
- OSD target  
`http://open-osd.org/`

# Key pNFS Participants



- Panasas (Objects), funding from ORNL for Linux work
- Network Appliance (Files over NFSv4)
- IBM (Files, based on GPFS)
- EMC (Blocks, HighRoad MPFSi)
- Sun (Files over NFSv4)
- U of Michigan/CITI (Linux maintainers, EMC and Microsoft contracts)
- LSI – open source block-based server
- DESY – Java-based implementation