pNFS Birds-of-Feather
FAST 2010: February 24

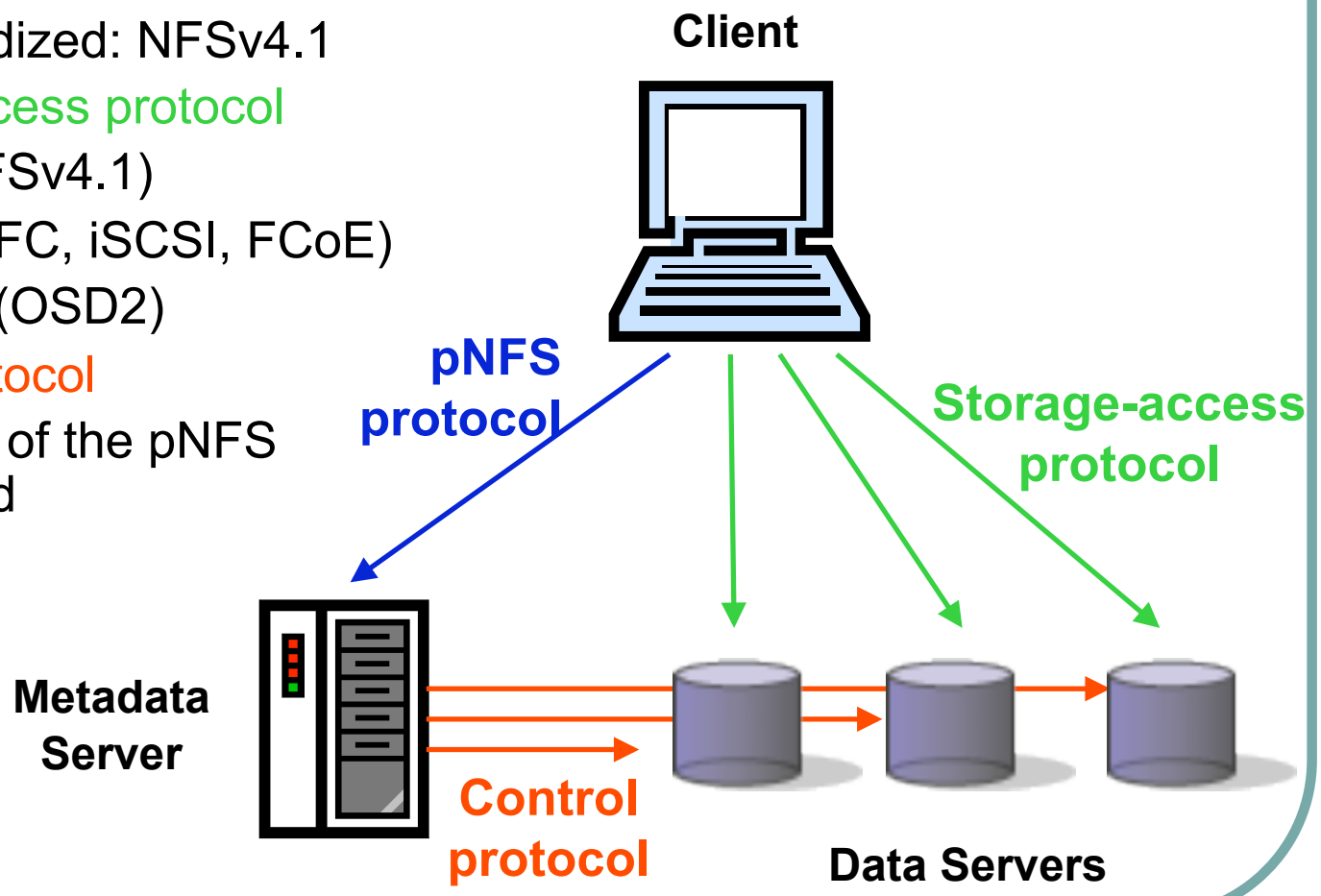
- Sorin Faibish, **EMC**
- **and pNFS friends**

Outline

- What is pNFS?
- pNFS Timeline
- Standards Status
- Industry Support
- Q&A

What is pNFS?

- pNFS protocol
  - standardized: NFSv4.1
- Storage-access protocol
  - files (NFSv4.1)
  - blocks (FC, iSCSI, FCoE)
  - objects (OSD2)
- Control protocol
  - Outside of the pNFS standard

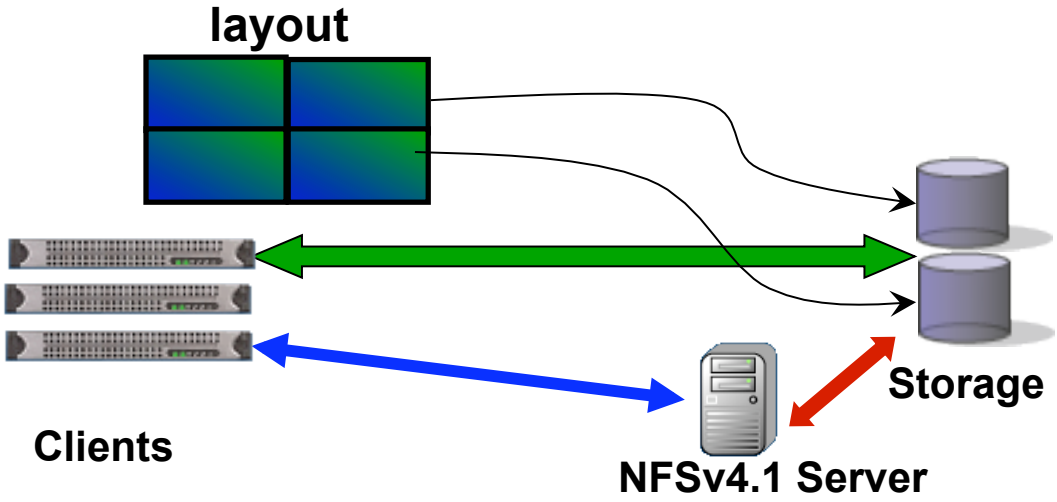


pNFS Value Proposition

- Distributes data across storage cluster
- Eliminates or reduces load and capacity balancing
- And yes: can accelerate I/O

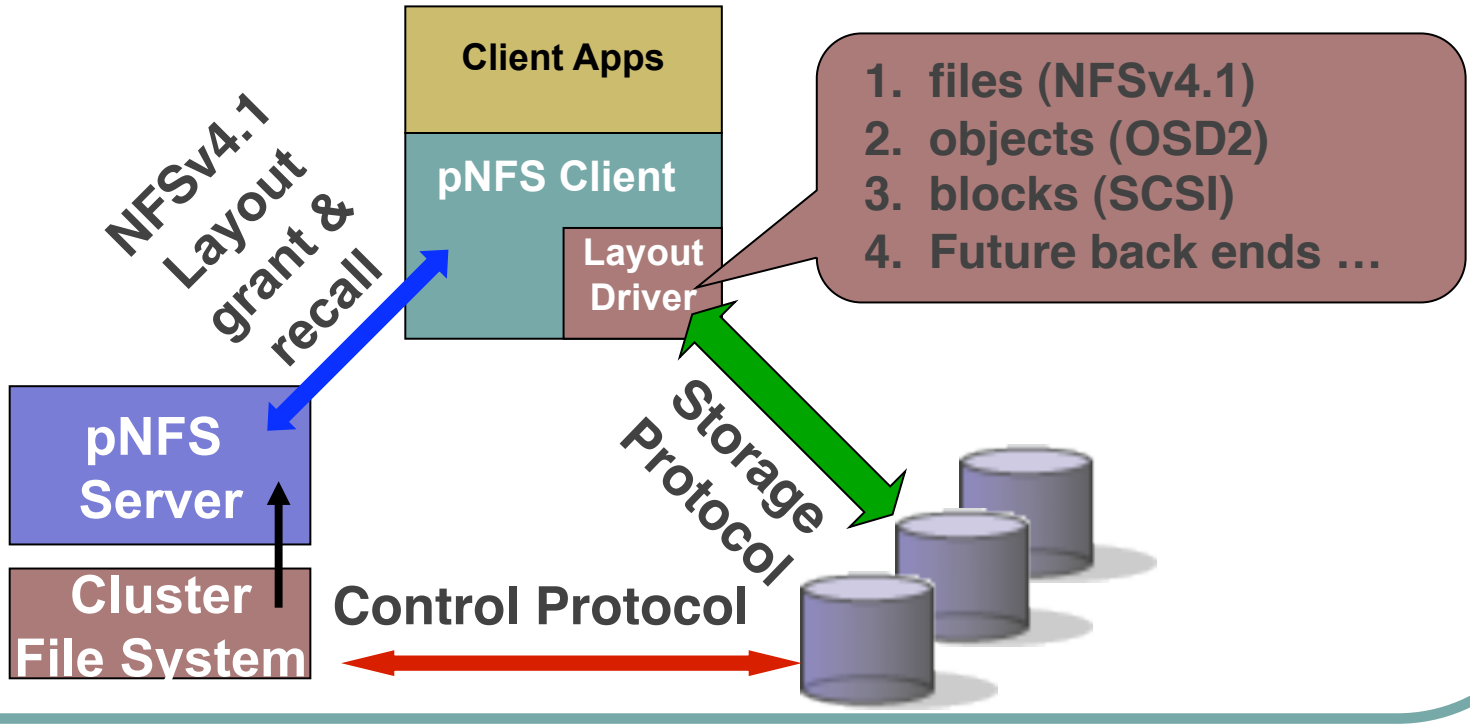
pNFS Layouts

- Client gets a *layout* from the NFSv4.1 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 NFSv4.1 I/O



Linux pNFS Client

- Transparent to applications
- Common client for different storage back ends
- Fewer support issues for storage vendors
- Normalizes access to clustered file systems



NFSv4.1 – OpenSource Status

- Two OpenSource Implementations
  - OpenSolaris and Linux (file, osd and block)
- OpenSolaris Client and Server
  - Support only file-based layout
  - S
    upport for multi-device striping already present (NFSv4.1 + pNFS)
  - “Simple Policy Engine” for policy-driven layouts also in the gate
- Linux Client and Server
  - Support files (NFSv4.1)
  - Support in progress blocks (SCSI), objects (OSD T10)
  - Cli
    e
    nt consists of generic pNFS client and “plug ins” for “layout drivers
- Windows NFSv4.1 Client from CITI - NEW

Timeline

- 2004 – CMU, NetApp and Panasas draft pNFS problem and requirement statements
- 2005 – CITI, EMC, NetApp and Panasas draft pNFS extensions to NFS
- 2005 – NetApp and Sun demonstrate pNFS at Connectathon
- 2005 – pNFS added to NFSv4.1 draft
- 2006 - 2008 – specification baked
  - Bake-a-thons, Connectathons
  - 26 iterations of NFSv4.1/pNFS spec
- 2009 – RFC submitted (680 pages)
- 2010 – RFC published
- 2010 – Fedora 10 includes pNFS server/client gits and rpms (did you triy it yet?)

pNFS Standards Status

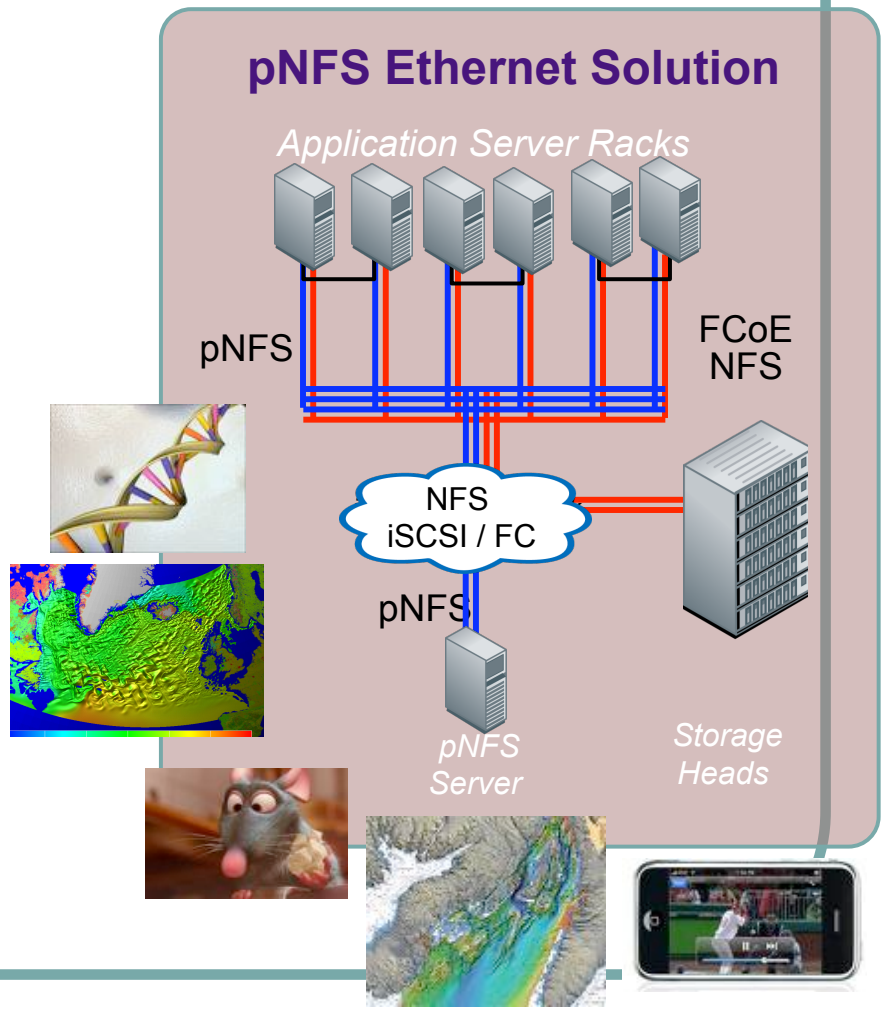
- NFSv4.1/pNFS were standardized at IETF
  - NFSv4 working group (WG)
- All done including RFC 5661,3,4:
  - WG last call (DONE)
  - Area Director review (DONE)
  - IETF last call (DONE)
  - IESG approval for publication (DONE)
  - IANA review (DONE)
  - RFC publication (2010)
- Consists of several documents:
  - RFC 5661 - NFSv4.1/pNFS/file layout
  - RFC 5662 - NFSv4.1 protocol description for IDL (rpcgen) compiler
  - RFC 5663 - blocks layout
  - RFC 5664 - objects layout
  - RFC 5665 - netid specification for transport protocol independence (IPv4, IPv6, RDMA)

Industry Contributors to pNFS Standard

- BlueArc
- CITI
- CMU
- EMC
- IBM
- LSI
- NetApp
- Ohio SuperComputer Center
- Panasas
- Seagate
- StorSpeed
- Sun Microsystems

Traditional HPC Use Cases

- Seismic Data Processing / Geosciences' Applications
- Broadcast & Video Production
- High Performance Streaming Video
- Finite Element Analysis for Modeling & Simulation
- HPC for Simulation & Modeling
- Data Intensive Searching for Computational Infrastructures



Questions – Client Technology

- When will a valid pNFS client be available?
- How does the parallel NFS block layout work?
- What's an object layout, how do objects compare to blocks/files?
- How do current cluster file systems work with pNFS client and server?
- How is client access to data servers coordinated and controlled?
- Attendee Questions Encouraged…

Questions – Files, Blocks and Objects

- How many layout types can there be?
- How does pNFS make managing a lot of systems easier?
- Can my application control how its data is striped?
- When can we expect to see real pNFS performance, not vendor claims based on older technology?
- How can I use pNFS in an enterprise environment; would it support non-disruptive (NDU) upgrades?
- Attendee Questions Encouraged…

Questions – Data Management

- Can I retain data management practices I use today based on Snapshots and Volume replication?
- Can File, Object and Block layouts co-exist in the same storage network?
- Can a client use volumes accessed via each layout concurrently?
- I'm deploying a Unified Ethernet Fabric; how do I secure data access – files, blocks & objects?
- Attendee questions encouraged…

Questions – What else

- NFSv4.1 and pNFS what are the expectations compared to 4.0?
- What additional function do you want to see added to 4.1:
  - FedFS
  - Copy server side
- What would you like to see in NFSv4.2 additional to what is proposed:
  - Sparse files
  - Access Permissions