NFSv4.1 — Plan for a Smooth Migration

SNIA™ WEBCAST

Presented by:
Alex McDonald

Hosted by:
Gary Gumanow

HOSTED BY THE
ETHERNET STORAGE FORUM

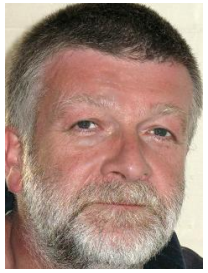
Ethernet Storage Forum Members

SNIA WEBCAST

The SNIA Ethernet Storage Forum (ESF) focuses on educating end-users about Ethernet-connected storage networking technologies.

Webcast Presenter

SNIA WEBCAST



Alex McDonald
Office of the CTO
NetApp

Alex McDonald joined NetApp in 2005, after more than 30 years in a variety of roles with some of the best known names in the software industry .

With a background in software development, support, sales and a period as an independent consultant, Alex is now part of NetApp's Office of the CTO that supports industry activities and promotes technology & standards based solutions, and is co-chair of the SNIA NFS Special Interest Group.

Gary Gumanow
Product Marketing,
Dell

Gary Gumanow, Dell Inc., is on the board of directors for the Ethernet Storage Forum, co-chair of the iSCSI SIG with SNIA, and has over 25 years of experience in IT management, systems integration, product management and strategic product planning.

Gary is currently responsible for product marketing of Dell's EqualLogic storage arrays. Gary holds two patents ad has authored many papers on storage, networking and server platform architecture.

SNIA's NFS Special Interest Group

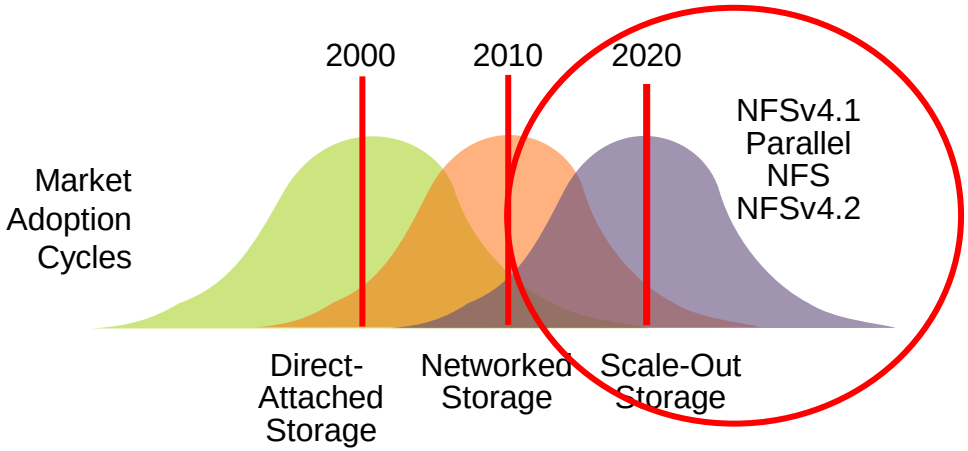
SNIA WEBCAST

- NFS SIG drives adoption and understanding of pNFS across vendors to constituents
  - Marketing, industry adoption, Open Source updates
- NetApp, EMC, Panasas and Sun founders
  - NetApp, EMC and Panasas act as co-chairs
- White papers on migration from NFSv3 to NFSv4
  - An Overview of NFSv4; NFSv4.0, NFSv4.1, pNFS, and proposed NFSv4.2 features
  - Migrating from NFSv3 to NFSv4
- Previous webcasts
  - 4 Reasons to Start Working with NFSv4 Now
  - Advances in NFS – NFSv4.1 and pNFS

NFS; Ubiquitous & Everywhere

SNIA WEBCAST

- NFS is ubiquitous and everywhere
- NFS doesn't stand still
  - NFSv2 in 1983, through NFSv4.1 in 2010
  - NFSv4.2 to be agreed at IETF shortly
  - Faster pace for minor revisions
- NFSv3 very successful
  - Protocol adoption is over time, and there have been no big incentives to change
- See White Papers, Tutorials and webcasts for NFSv4.x; details at www.snia.org



The Four Reasons for NFSv4.1

SNIA WEBCAST

	Functional	Business Benefit
Security	ACLs for authorization Kerberos for authentication	Compliance, improved access, storage efficiency, WAN use
High availability	Client and server lease management with fail over	High Availability, Operations simplicity, cost containment
Single namespace	Pseudo directory system	Reduction in administration & management
Performance	Multiple read, write, delete operations per RPC call Delegate locks, read and write procedures to clients Parallelised I/O	Better network utilization for all NFS clients Leverage NFS client hardware for better I/O

Agenda

SNIA WEBCAST

- We'll cover
  - Selecting the application for NFSv4.1
  - Planning;
    - Filenames and namespace considerations
    - Firewalls
    - Understanding statefulness
    - Security
  - Server & Client Availability
  - Where Next
    - Considering pNFS
- This is a high level overview
  - Use SNIA white papers and vendors (client & server) to help you implement

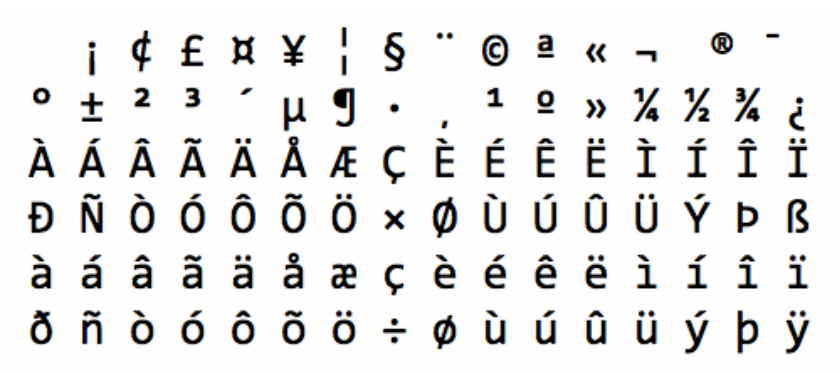
Selecting an Application

- First task; select an application or storage infrastructure for NFSv4.1 use
  - Home directories
  - HPC applications
- Don't select…
  - Oracle; use dNFS built in to the Oracle kernel
  - VMware & other virtualization tools; no support for anything other than NFSv3 as of this date
  - "Oddball" applications that expect to be able to internally manage NFSv3 "maps" with multiple mount points, or auxiliary protocols like `mountd`, `statd` etc; or requires O_DIRECT reads and writes
  - Any application that requires UDP; NFSv4 doesn't support anything except TCP

Planning; File Names

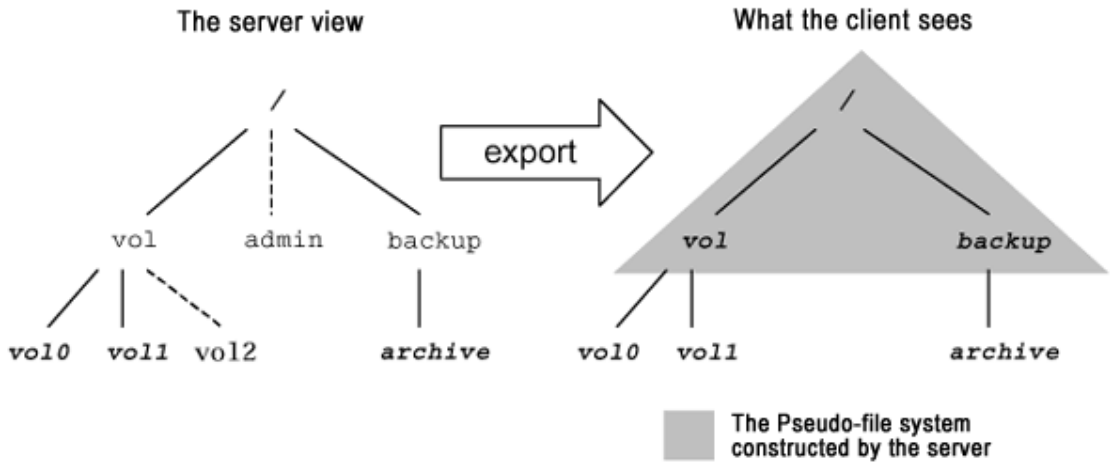
SNIA WEBCAST

- File Names
  - NFSv4 uses UTF-8
  - Check filenames for compatibility
    - NFSv3 file created with the name `René` contains an 8 bit ASCII
    - UTF-8 `é` indicates a multibyte UTF-8 encoding, which will lead to unexpected results
- Action
  - Review existing NFSv3 names to ensure that they are 7 bit ASCII clean
  - These aren't;



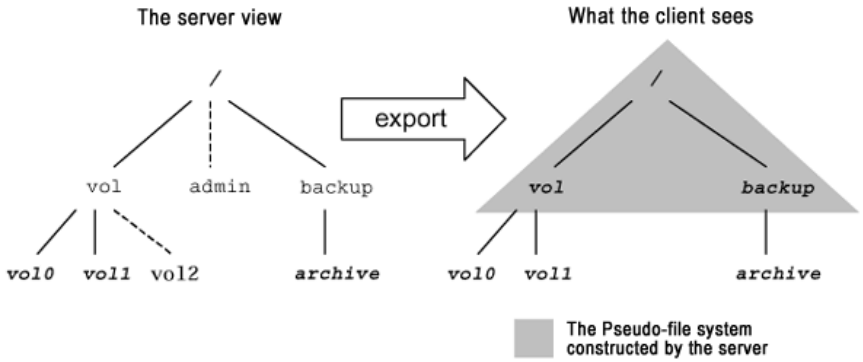
Planning; Namespace

- Uniform and "infinite" namespace
  - Moving from user/home directories to datacenter & corporate use
  - Meets demands for "large scale" protocol
  - Unicode support for UTF-8 codepoints
- No automounter required
  - Simplifies administration



Planning; Namespace

- Namespace Example
  - Server exports
    - `/vol/vol0`
    - `/vol/vol1`
    - `/backup/archive`



- Mount root `/` over NFSv3:
  - Allows the client to list the contents of `/vol/vol2`
- Mount root `/` over NFSv4:
  - `/vol/vol2` has not been exported and the pseudo filesystem does not contain it; **the directory is not visible**
  - An **explicit** mount of `vol/vol2` will be required.

Planning; Namespace

- Namespaces
- Action
  - Consider using the flexibility of pseudo-filesystems to permit easier migration from NFSv3 directory structures to NFSv4, without being overly concerned as to the server directory hierarchy and layout.
- However;
  - If there are applications that traverse the filesystem structure or assume the entire filesystem is visible, caution should be exercised before moving to NFSv4 to understand the impact presenting a pseudo filesystem
  - **Especially when converting NFSv3 mounts of / to NFSv4**

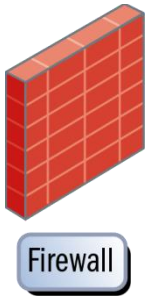
Planning; Stateful Clients & Sessions

- Statefulness
  - NFSv4 gives client independence
  - Previous model had “dumb” stateless client
  - Server had the “smarts”
- Pushes work out to client through delegations & caching
  - Compute nodes work best with local data
  - NFSv4 eliminates the need for local storage
  - Exposes more of the backend storage functionality
    - Client can help make server smarter by providing hints
- Sessions
  - NFSv3 server never knows if client got reply message
  - NFSv4.1 introduces Sessions
  - A session maintains the server's state relative to the connections belonging to a client
- Action
  - None; use delegation & caching transparently; client & server provide transparency
  - NFSv4 advantages include session lock clean up automatically

Planning; Firewalls

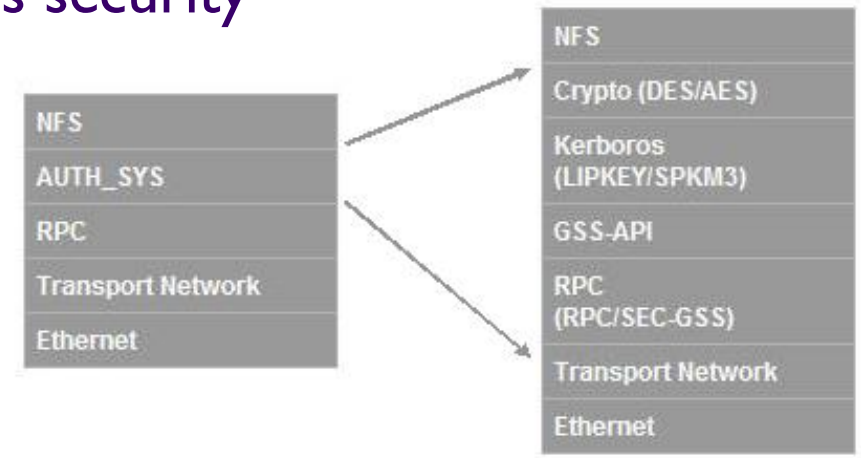
SNIA WEBCAST

- Firewalls
  - NFSv3 promiscuously uses ports; including111, 1039, 1047, 1048, and 2049 (and possibly more)
  - NFSv4 has no "auxiliary" protocols like `portmapper`, `statd`, `lockd` or `mountd`; uses port 2049 with TCP only
  - No floating ports required & easily supported by NAT
- Action
  - Open port 2049 for TCP on firewalls



Planning; Security

- Strong security framework
- Access control lists (ACLs) for security and Windows® compatibility
- Security with Kerberos
  - Negotiated RPC security that depends on cryptography, RPCSEC_GSS
- NFSv4 can be implemented without implementing Kerberos security
  - Not advised; but it is possible



Planning; Security

- Implementing without Kerberos
- NFSv3 represents users and groups via 32 bit integers
  - UIDs and GIDs with GETATTR and SETATTR
- NFSv4 represents users and groups as strings
  - user@domain or group@domain
- Requires NFSv3 UID and GUID 32 bit integers be converted to all numeric strings
  - Client side;
    - Run **`idmapd6`**
    - **`/etc/idmapd.conf`** points to a default domain and specifies translation service **`nsswitch`**.
  - Incorrect or incomplete configuration, UID and GUID will display **`nobody`**.
  - Using integers to represent users and groups requires that **every client and server that might connect to each other** agree on user and group assignments.
- Last resort!

Planning; Security

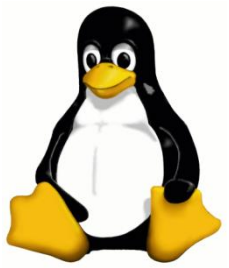
- Implementing with Kerberos
- Find a security expert
  - Requires to be correctly implemented
  - Do not use NFSv4 as a testbed to shake out Kerberos issues!
- User communities divided into realms
  - Realm has an administrator responsible for maintaining a database of users
  - Correct **user@domain** or **group@domain** string is required
  - NFSv3 32 bit integer UIDs and GUIDs are explicitly denied access
- NFSv3 and NFSv4 security models are not compatible with each other
  - Although storage systems may support both NFSv3 and NFSv4 clients, be aware that there may be compatibility issues with ACLs. For example, they may be enforced **but not visible** to the NFSv3 client.

Planning; Security

- **Action**
  - Review security requirements on NFSv4 filesystems
  - Use Kerberos for robust security, especially across WANs
  - If using Kerberos, ensure it is installed and operating correctly
    - Don't use NFSv4 as a testbed to shake out Kerberos issues
- Last resort
  - If using NFSv3 security, ensure UID and GUID mapping and translation is uniformly implemented across the enterprise

Linux Client and NFSv4.1

- Upstream (Linus) Linux NFSv4.1 client support
  - Basic client in Kernel 2.6.32
  - pNFS support (files layout type) in Kernel 2.6.39
  - Support for the 'objects' and 'blocks' layouts was merged in Kernel 3.0 and 3.1 respectively
- Full read and write support for all three layout types in the upstream kernel
  - Blocks, files and objects
  - O_DIRECT reads and writes are not yet supported



Linux Client and NFSv4.1

SNIA WEBCAST

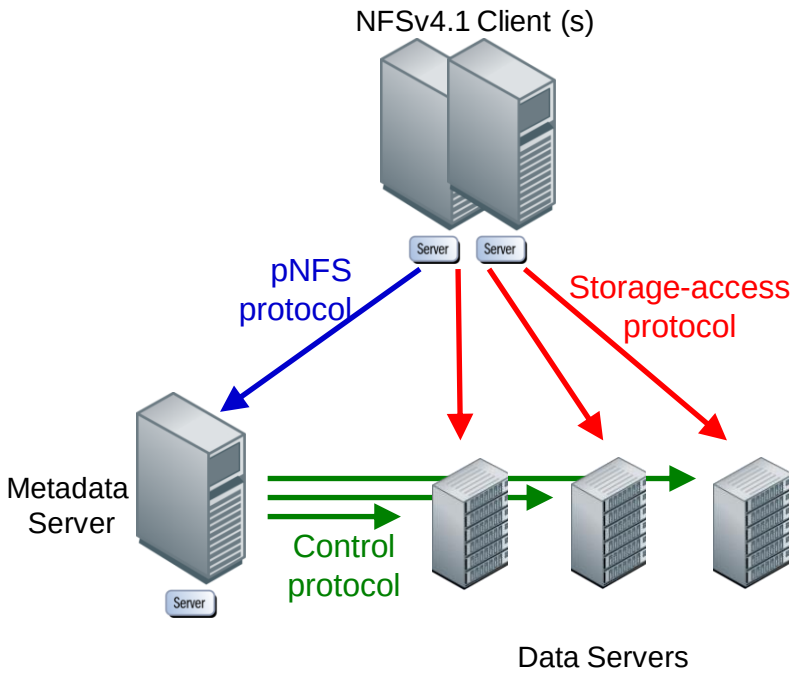
- pNFS client support in distributions
  - Fedora 15 was first for pNFS files
  - Kernel 2.6.40 (released August 2011)
- Red Hat Enterprise Linux version 6.2
  - “Technical preview" support for NFSv4.1 and for the pNFS files layout type
- Other Open Source
  - Microsoft NFSv4.1 Windows client from CITI

It's Up & Running; now for pNFS

NFSv4.1 (pNFS) can aggregate bandwidth

- Modern approach; relieves issues associated with point-to-point connections

- pNFS Client
  - Client read/write a file
  - Server grants permission
  - File layout (stripe map) is given to the client
  - Client parallel R/W directly to data servers
- Removes IO Bottlenecks
  - No single storage node is a bottleneck
  - Improves large file performance
- Improves Management
  - Data and clients are load balanced
  - Single Namespace



Summary/Call to Action

SNIA WEBCAST

- Start using NFSv4.0, NFSv4.1 today
  - NFSv4.2 nearing approval
- Planning is key
  - Application, issues & actions to ensure smooth implementations
- Next up; pNFS
  - First open standard for parallel I/O across the network
  - Ask vendors to include NFSv4.1 support for client/servers
  - pNFS has wide industry support
  - Commercial implementations and open source

Question & Answer

To download this Webcast

after the presentation, go to

http://www.snia.org/about/socialmedia/