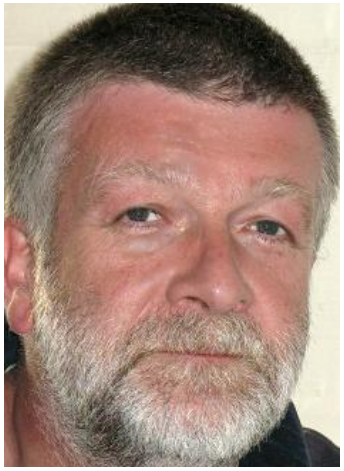
# What's New in NFS 4.2?

April 28, 2015

SNIA™
Ethernet Storage Forum

# Webcast Presenters

**J Metz, SNIA Board of Directors**
**Cisco**



**Alex McDonald, Vice Chair SNIA-ESF**
**NetApp**

# SNIA Legal Notice


- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

# Agenda

SNIA Ethernet Storage Forum

- Before we start: NAS Protocols
- NFSv4 Background
- New features in NFSv4.2
  - NFSv4.2 not yet standardized, but many features already available
- pNFS New Layouts; FlexFiles & SCSI
- Other Developments & Beyond NFSv4.2

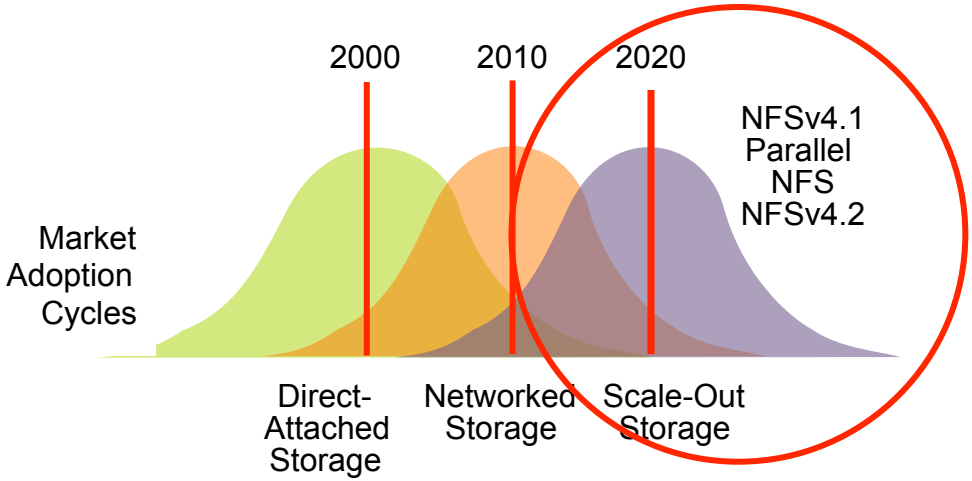
# Before we start: NAS Protocols

SNIA Ethernet Storage Forum

- Compare & contrast with SAN
  - This is a gross simplification!
- NAS (Network Attached Storage) vs SAN (Storage Area Network)
  - Filesystem based (directories, files) vs volume based (LUN)
  - Byte addressable, byte size chunks vs 4K aligned 4K blocks
  - Primarily Ethernet, RDMA vs Fibre Channel, Ethernet …
- Two major dominant NAS protocols
  - NFS (Network File System) protocol
  - SMB (Microsoft's Server Message Block) protocol
    - Formerly known as CIFS (Common Internet File System)

# NFS: Ubiquitous & Everywhere

- NFS is ubiquitous and everywhere
- Industry – and hence NFS – doesn't stand still
  - NFSv2 in 1983
  - NFSv3 in 1995
  - NFSv4 in 2003, updated 2015
  - NFSv4.1 and pNFS in 2010
  - NFSv4.2 to be agreed at IETF shortly
  - Faster pace for minor revisions
  - http://datatracker.ietf.org/wg/nfsv4



# Evolving Requirements

SNIA Ethernet Storage Forum

- Pace of NFSv4 adoption now increasing
- Beyond traditional home directories
  - VMware announces support for NFSv4.1 as a client for storing VMDKs
  - Amazon announces support for NFSv4.0 in AWS Elastic File System (EFS)
- Industry is changing, as are requirements
  - Economic Trends
    - Cheap and fast computing clusters
    - Cheap and fast network (1GbE to 10GbE, 40GbE and 100GbE in the datacenter)
    - Cost effective & performant storage based on flash, flash & SATA
  - Performance
    - Exposes NFSv3 single threaded bottlenecks in applications
    - Increased demands of compute parallelism and consequent data parallelism
    - Analysis begets more data, at exponential rates
    - Competitive edge (ops/sec)
  - Business requirement to reduce solution times
    - NFSv4.1 brings increased scale & flexibility
    - Outside of the datacenter; requires good security, scalability

# Agenda

SNIA Ethernet Storage Forum

- Before we start: NAS Protocols
- NFSv4 Background
- New features in NFSv4.2
  - NFSv4.2 not yet standardized, but many features already available
- pNFS New Layouts; FlexFiles & SCSI
- Other Developments & Beyond NFSv4.2

# NFS as a Standard

- How the IETF Works
  - https://www.ietf.org/about/standards-process.html
  - “a specification undergoes a period of development and several iterations of review by the Internet community and revision based upon experience, is adopted as a Standard by the appropriate body... and is published.”
  - Open process
  - Technical competence; “engineering quality”
  - Volunteer Core
  - Rough consensus and running code
  - Protocol ownership
- NFS Working Group
  - http://datatracker.ietf.org/wg/nfsv4/charter/

# NFSv4 background

SNIA
Ethernet Storage Forum

- Areas addressed by NFSv4, NFSv4.1 and pNFS
  - Security
  - Uniform namespaces
  - Statefulness & Sessions
  - Compound operations
  - Caching; Directory & File Delegations
  - Layouts & pNFS (parallel NFS)
  - Trunking (NFSv4.1 & pNFS)
- SNIA has entire set of white papers & tutorials
  - https://www.brighttalk.com/search?duration=0..&keywords[]=nfs&q=snia&rank=webcast_relevance
    http://www.snia.org/sites/default/files/SNIA_An_Overview_of_NFSv4-3_0.pdf
    http://www.snia.org/sites/default/files/Migrating_to_NFSv4_v04_-Final.pdf
- NB: NFSv4.2 not yet standardized
  - But some features already available

# Agenda

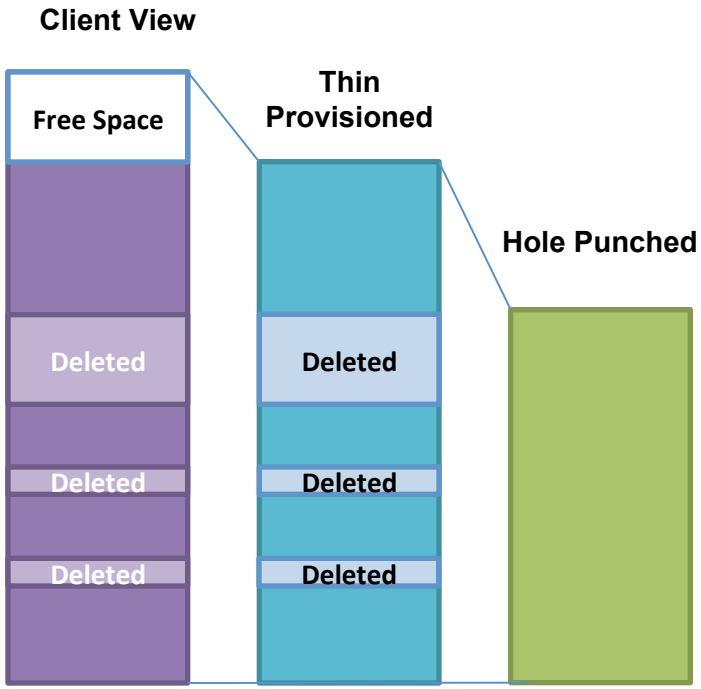
SNIA Ethernet Storage Forum

- Before we start: NAS Protocols
- NFSv4 Background
- New features in NFSv4.2
  - NFSv4.2 not yet standardized, but many features already available
  - Sparse File Support
  - Space Reservation
  - Labeled NFS
  - IO_ADVISE
  - Server Side Copy
  - Application Data Holes
- pNFS New Layouts; FlexFiles & SCSI
- Other Developments & Beyond NFSv4.2

# New Features in NFSv4.2

SNIA Ethernet Storage Forum

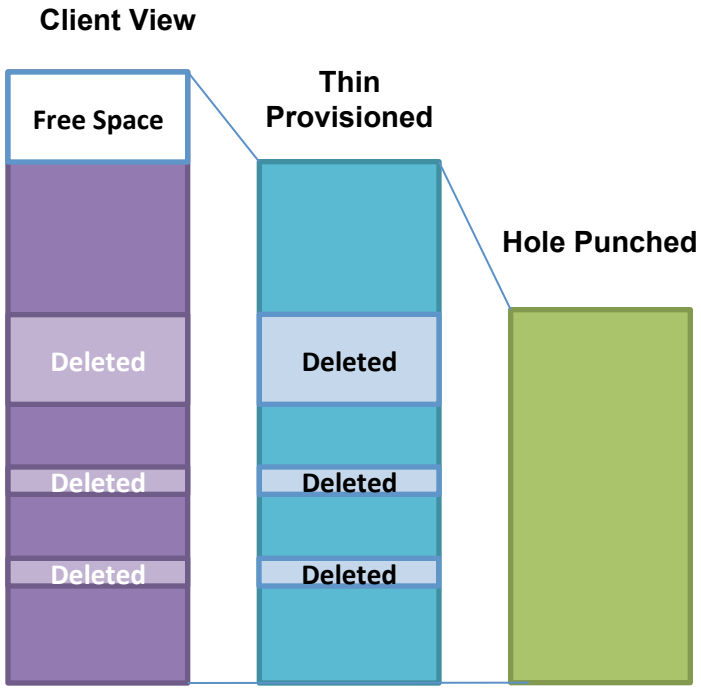
- Sparse file support
  - "Hole punching" and the reading of sparse files
  - Key to effective management of expensive storage devices (like SSDs)
  - VM datastores benefit
  - In 3.18 kernel (October 2014)



# New Features in NFSv4.2

SNIA Ethernet Storage Forum

- Space reservation
  - Ensure a file will have storage available
  - Make sure client view of the storage is reflected by the server's space allocation policies
  - Still able to hole punch & thin provision; it's a commitment, not a physical requirement
  - In 3.19 kernel (December 2014)



# New Features in NFSv4.2

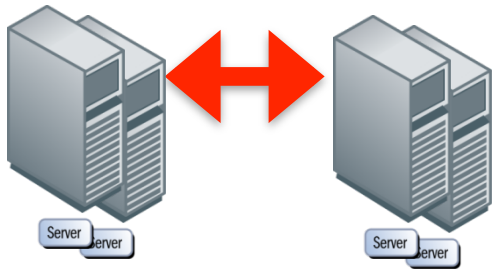
SNIA Ethernet Storage Forum

- Labeled NFS (LNFS)
  - Allows (partial) SELinux support
  - In 3.11 (September 2013), in RHEL7 (June 2014)
- IO_ADVISE
  - Client or application can inform the server IO patterns and hence possible caching requirements of the file (including hints for pNFS)
  - Predicting IO patterns is hard without hints
  - Sequential, random, read, write…

# New Features in NFSv4.2

SNIA Ethernet Storage Forum

- Server-Side Copy (SSC)
  - Removes one leg of the copy
  - Destination reads directly from the source
  - Two proposed types
    - CLONE: local to the server
    - COPY: server to server, potentially different physical systems
  - Security a big issue, requires updated security model (RPCSEC_GSS Version 3)



# New Features in NFSv4.2

SNIA Ethernet Storage Forum

- Application Data Holes
  - (previously Application Data Blocks or ADB)
  - Allows definition of the format of file
  - Examples: database or a VM image.
  - INITIALIZE blocks with a single compound operation
    - Initializing a 30G database takes a single over the wire operation instead of 30G of traffic.

# Agenda

SNIA Ethernet Storage Forum

- Before we start: NAS Protocols
- NFSv4 Background
- New features in NFSv4.2
  - NFSv4.2 not yet standardized, but many features already available
- pNFS New Layouts; FlexFiles & SCSI
- Other Developments & Beyond NFSv4.2

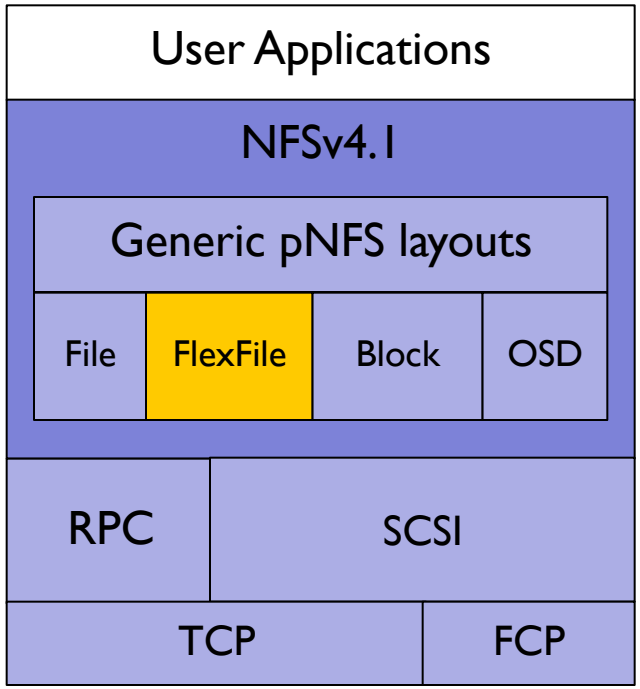
# Flex Files: A New pNFS Layout

SNIA Ethernet Storage Forum

- Advance in Clustering
  - Aggregation of standalone NFS servers
    - Customers heavily invested in NFSv3
    - Allows reuse of legacy filers as data servers in a clustered configuration
  - Exporting of existing clustered file system
    - For example: Ceph, Gluster
    - No standard storage access protocol; pNFS could be used instead
  - Flexible, per-file striping patterns
    - Application SLAs and management policies as well as dynamic load balancing and tiering decisions require per-file control over striping
    - Existing clustered file systems do not map to the files layout striping patterns

# Flex-files pNFS layout

- pNFS is dependent on session support, which is only available in NFSv4.1
- Flex-files pNFS layout
  - Flexible, per-file striping patterns and simple device information suitable for aggregating standalone NFS servers into a centrally managed pNFS cluster
- SCSI pNFS Layout
  - Extends pNFS Block/Volume Layout
  - Provides closer integration into the SCSI Architecture
- These are proposed, but remember not yet ratified or available!
- Brief pNFS backgrounder



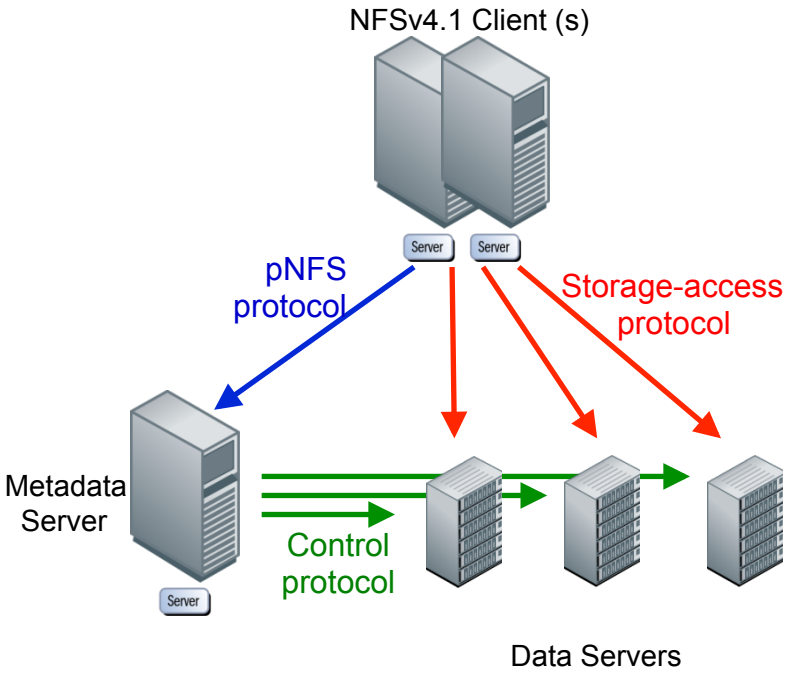

OSD: Object based Storage Device

# Why pNFS

## NFSv4.1 (pNFS) can aggregate bandwidth

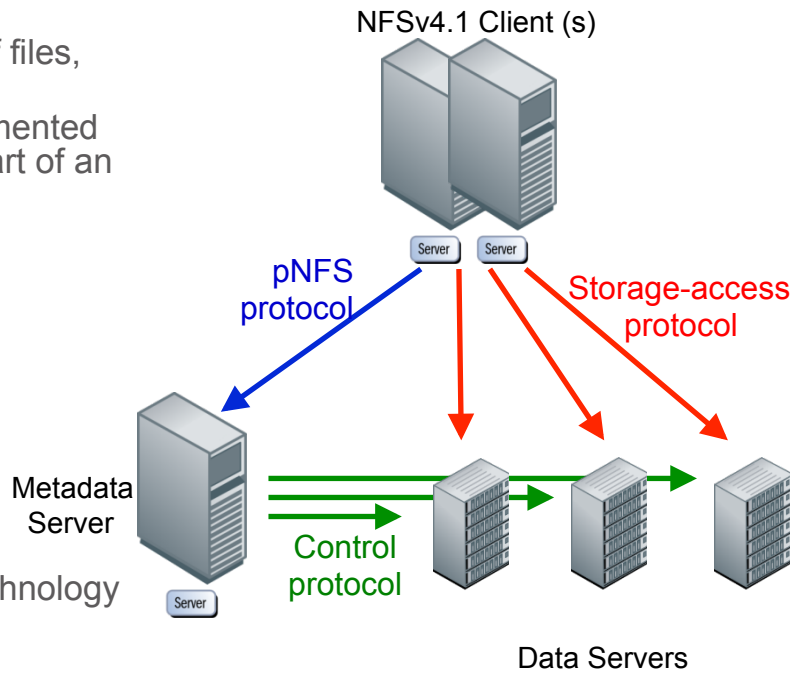
- Modern approach; relieves issues associated with point-to-point connections

- 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



# pNFS Terminology

- Metadata Server; the MDS
  - Maintains information about location and layout of files, objects or block data on data servers
  - Shown as a separate entity, but commonly implemented on one or across more than one data server as part of an array
- pNFS protocol
  - Extended protocol over NFSv4.1
  - Client to MDS communication
- Storage access protocol
  - Files; NFS operations
  - Objects: OSD SCSI objects protocol (OSD2)
  - Blocks; SCSI blocks (iSCSI, FCP)
- Control protocol
  - Not standardised; each vendor uses their own technology to do this
- Layout
  - Description of devices and sector maps for the data stored on the data servers
  - 3 types; files, block and object
- Callback
  - Asynchronous RPC calls used to control the behavior of the client during pNFS operations



# pNFS Operations

- Client requests layout from MDS
- Layout maps the file/object/ block to data server addresses and locations
- Client uses layout to perform direct I/O to the storage layer
- MDS or data server can recall the layout at any time using callbacks
- Client commits changes and releases the layout when complete
- pNFS is optional
  - Client can fall back to NFSv4

- pNFS operations
  - LAYOUTCOMMIT Servers commit the layout and update the meta-data maps
  - LAYOUTRETURN Returns the layout or the new layout, if the data is modified
  - GETDEVICEINFO Client gets updated information on a data server in the storage cluster
  - GETDEVICELIST Clients requests the list of all data servers participating in the storage cluster
  - CB_LAYOUT Server recalls the data layout from a client if conflicts are detected

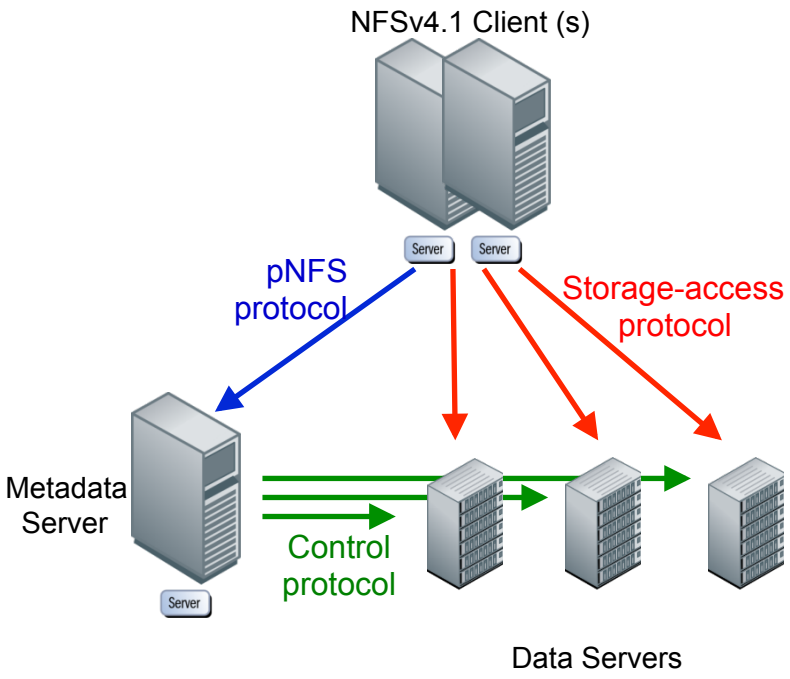
# Flex Files: A New pNFS Layout

SNIA Ethernet Storage Forum

- Flex-files pNFS layout
  - Flexible, per-file striping patterns and simple device information suitable for aggregating standalone NFS servers into a centrally managed pNFS cluster
  - Assumption was that data servers would be NFSv4.1 or better
  - Flex-files layout allows various data servers
- Encourage best-of-breed solutions
  - NFS as the basic back-end control protocol allows one to mix and match a metadata server and data servers from different vendors

# Flex Files: A New pNFS Layout

- Permit layout to extend over non-pNFS data servers
- Example with NFSv3
  - File gets private UID and GID that client uses to access the file
  - To fence the file: the MDS changes the UID or GID
    - Requires exclusive root access to the data server
  - Fences access from clients, and forces clients to:
    - Return the file's layout
    - Request a new layout for the file
  - MDS grants access via new UID/ GID to clients it does NOT want to fence
  - Only AUTH_SYS is supported to the data servers, not full Kerberos



# Agenda

SNIA Ethernet Storage Forum

- Before we start: NAS Protocols
- NFSv4 Background
- New features in NFSv4.2
  - NFSv4.2 not yet standardized, but many features already available
- pNFS New Layouts; FlexFiles & SCSI
- Other Developments & Beyond NFSv4.2

# Other Developments & Beyond NFSv4.2



- Other work in Progress
  - Formalization of NFS/RDMA
  - RPCSEC_GSSv3 (security)
- Beyond NFSv4.2
  - NFS xattrs?
  - pNFS for directories (metadata striping)?
  - Byte-range delegations?

# Summary/Call to Action

SNIA
Ethernet Storage Forum

- NFS has more relevance today for commercial, HPC and other use cases than it ever did
  - Features for a virtualized data centers
- Developments driven by application & business requirements
- Adoption slow, but will continue to increase
  - NFSv4 support widely available
  - New NFSv4.1 with client & server support
  - NFS defines how you get to storage, not what your storage looks like
- Start using NFSv4.1 today
  - It works & it's available
  - pNFS offers performance support for modern NAS devices
  - Ask vendors to include NFSv4.1 and pNFS support for client/servers
  - pNFS has wide industry support
  - Commercial implementations and open source
- NFSv4.2 & future pNFS
  - Indicates industry commitment & development to NFS

- Q&A
- Supporting white papers and information can be found at

https://www.brighttalk.com/search?duration=0..&keywords[]=nfs&q=snia&rank=webcast_relevance
http://www.snia.org/sites/default/files/SNIA_An_Overview_of_NFSv4-3_0.pdf
http://www.snia.org/sites/default/files/Migrating_to_NFSv4_v04_-Final.pdf
http://linux-nfs.org
http://datatracker.ietf.org/wg/nfsv4
https://tools.ietf.org/wg/nfsv4/draft-ietf-nfsv4-minorversion2/

# After This Webcast

SNIA Ethernet Storage Forum

- This webcast and a PDF of the slides will be posted to the SNIA Ethernet Storage Forum (ESF) website and available on-demand
  - http://www.snia.org/forums/esf/knowledge/webcasts
- A full Q&A from this webcast, including answers to questions we couldn't get to today, will be posted to the SNIA-ESF blog
  - http://sniaesfblog.org/
- Follow us on Twitter @ SNIAESF

# Thank You

SNIA™
Ethernet Storage Forum