
Education

# pNFS & NFSv4.2; a filesystem for grid, virtualization and database

David Dale, Director Industry Standards, Netapp

Author: Joshua Konkle, DCIG

# 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.

# Abstract

- pNFS & NFSv4.2; a filesystem for grid, virtualization and database
  - This session will appeal to Virtual Data Center Managers, Database Server administrators, and those that are seeking a fundamental understanding pNFS. This session will cover the four key reasons to start working with NFSv4 today, and explain the storage layouts for pNFS (parallel NFS), NFSv4.1 and the upcoming NFSv4.2 standard. The session includes use cases for database access, enterprise and desktop virtualization, HPC and datacenter use.

# Tutorial Agenda

- Introduction to NFS and NFS Special Interest Group
- NFS v4 – Security, High Availability, Internationalization and Performance
- pNFS and NFSv4.1
  - pNFS Use Cases – Virtualization, Database, etc
- OpenSource Client Status
- NFSv4.2; the next wave

# SNIA's NFS Special Interest Group

- 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 paper on migration from NFSv3 to NFSv4
  - "Migrating from NFSv3 to NFSv4"

SNIA Ethernet Storage Forum

*Learn more about us at: www.snia.org/forums/esf*

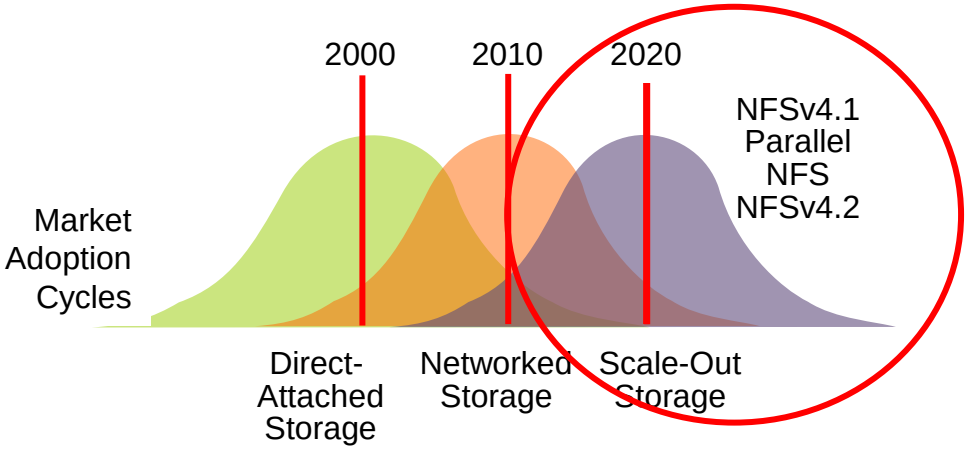
# Background Information

- Network File System
  - A protocol that provides transparent remote access to shared file systems across networks
  - NFS clients are included in all commonly used Operating Systems, e.g. Linux, Solaris, AIX, Windows etc…..
  - Application and OSI layers (remote procedure calls)

- NFS Server; Inspiration to NAS and appliances
  - Many Operating Systems have NFS servers
  - NAS Appliance – Control, Consistency and Cadence
  - Vendors offer commodity & custom hardware, with NFS & storage management software

# NFS; Ubiquitous & Everywhere

- 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



# Evolving 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 & SATA
- Performance
  - Exposes single threaded bottlenecks in applications
  - Increased demands of compute parallelism and consequent data parallelism
- Powerful compute systems
  - Analysis begets more data, at exponential rates
  - Competitive edge (ops/sec)
- Business requirement to reduce solution times
  - Beyond performance; NFS 4.1 brings increased scale & flexibility
  - Outside of the datacenter; requires good security

# Performance, Management and Reliability

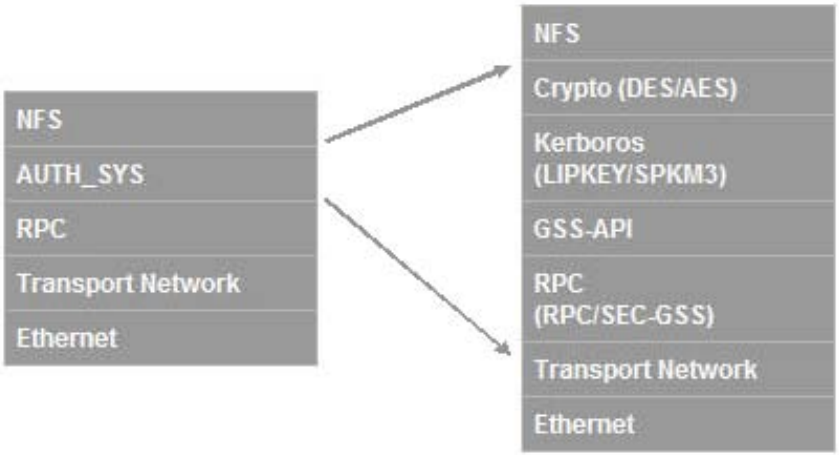
- Random I/O and Metadata intensive workloads
  - Memory and CPU are hot spots
  - Load balancing limited to pair of NFS heads; originally designed for HA
    - Not a limitation of the NFS 4.1 protocol
- Compute farms are growing larger in size
  - NFS head can handle a 1000+ NFS clients
  - NFS head hardware comparable to client CPU, I/O, Memory
  - NFS head requires more spindles to distribute the I/O
- Reliability and availability are challenging
  - Data striping limited to single head and disks
  - Non-disruptive upgrades affect dual-head configurations
  - Access and connectivity is typically limited to a pair of NFS server heads

# NFS Major Features

- NFSv4
  - Security, Namespace, FedFS, "Statefulness"
- NFSv4.1
  - Sessions, Layouts, pNFS
- The Linux Client
- Proposed for NFSv4.2
  - Server Side Copy, ADB, Sparse Files, Space Reservations…
- Implications & Applications

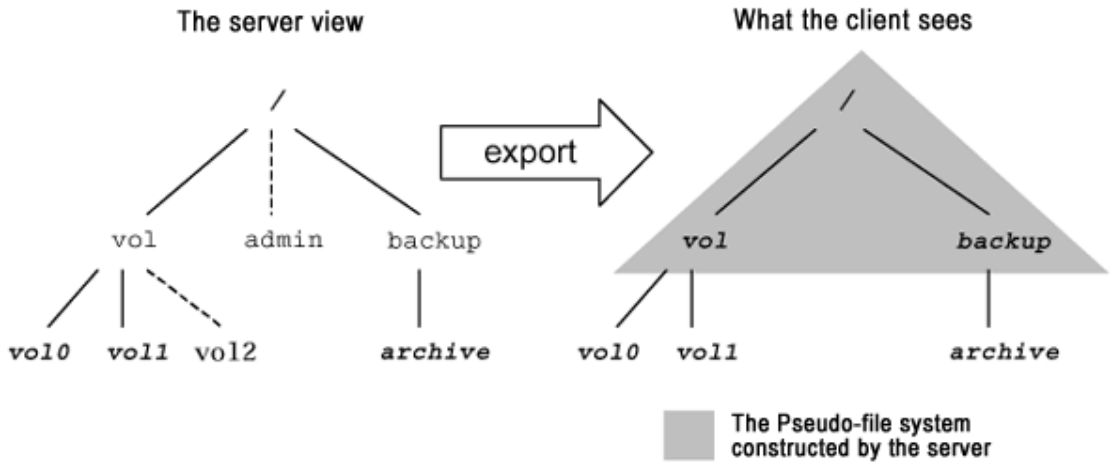
# NFSv4 Major Features; Security

- Strong security framework
- Access control lists (ACLs) for security and Windows® compatibility
- Mandatory security with Kerberos
  - Negotiated RPC security that depends on cryptography, RPCSEC_GSS

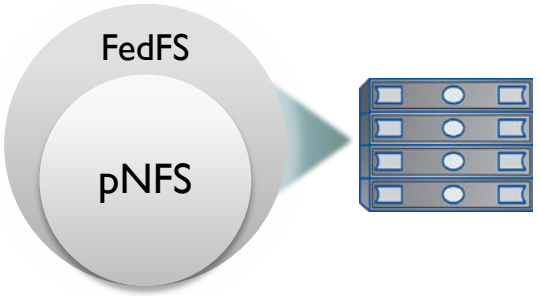


# NFSv4 Major Features; Namespace

- 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



# NFSv4 Major Features; FedFS




- Federated File System
  - Uniform namespace that has local and geographically global referral infrastructure
  - Accessible to unmodified NFSv4 clients
  - Addresses directories, referrals, nesting, and namespace relationships
- Client finds namespace via DNS lookup
  - Sees junctions (directories) and follows them as NFSv4 referrals

# NFSv4 Major Features; Stateful Clients

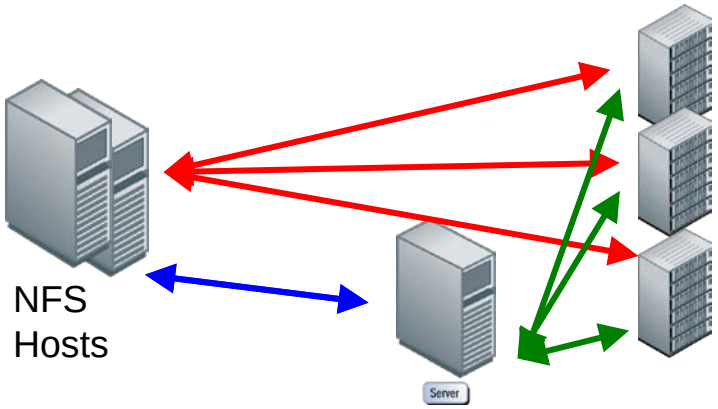
- NFSv4 gives client independence
  - Previous model had "dumb" stateless client
  - Server had the smarts
- Pushes work out to client through delegations & caching
- Why?
  - Compute nodes work best with local data
  - NFSv4 eliminates the need for local storage
  - Exposes more of the backend storage functionality
    - Client can help make server smarter by providing hints

# NFSv4.1 Major Features; Sessions

- NFSv3 server never knows if client got reply msg
- NFSv4.1 introduces Sessions
  - Major protocol infrastructure change
  - Exactly Once Semantics (EOS)
  - Bounded size of reply cache
  - Unlimited parallelism
- A session maintains the server's state relative to the connections belonging to a client.

# NFSv4.1 Major Features; Layouts

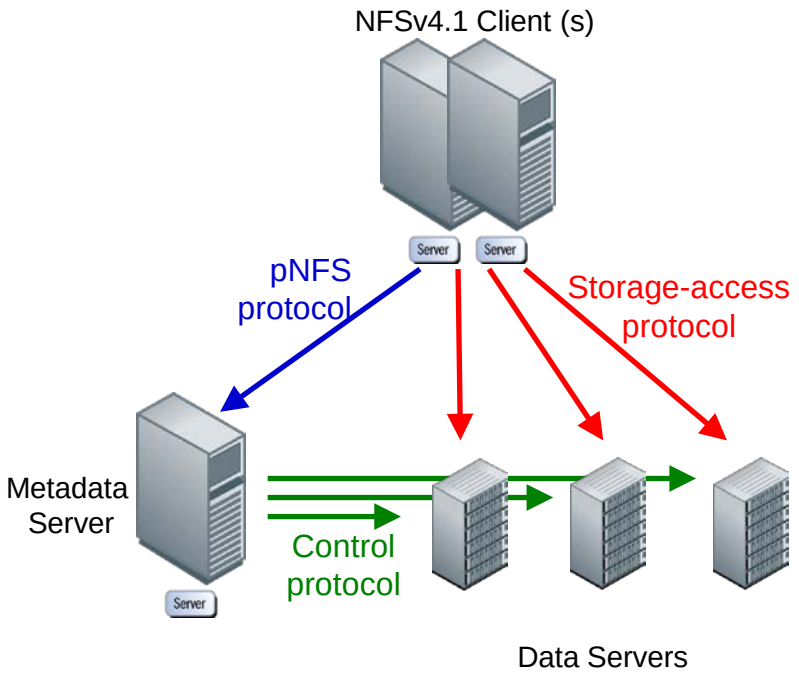
- Layouts
  - Files, objects and block layouts
  - Provides flexibility for storage that underpins it
  - Location transparent
    - Striping and clustering
- Examples
  - Blocks, Object and Files layouts all available from various vendors



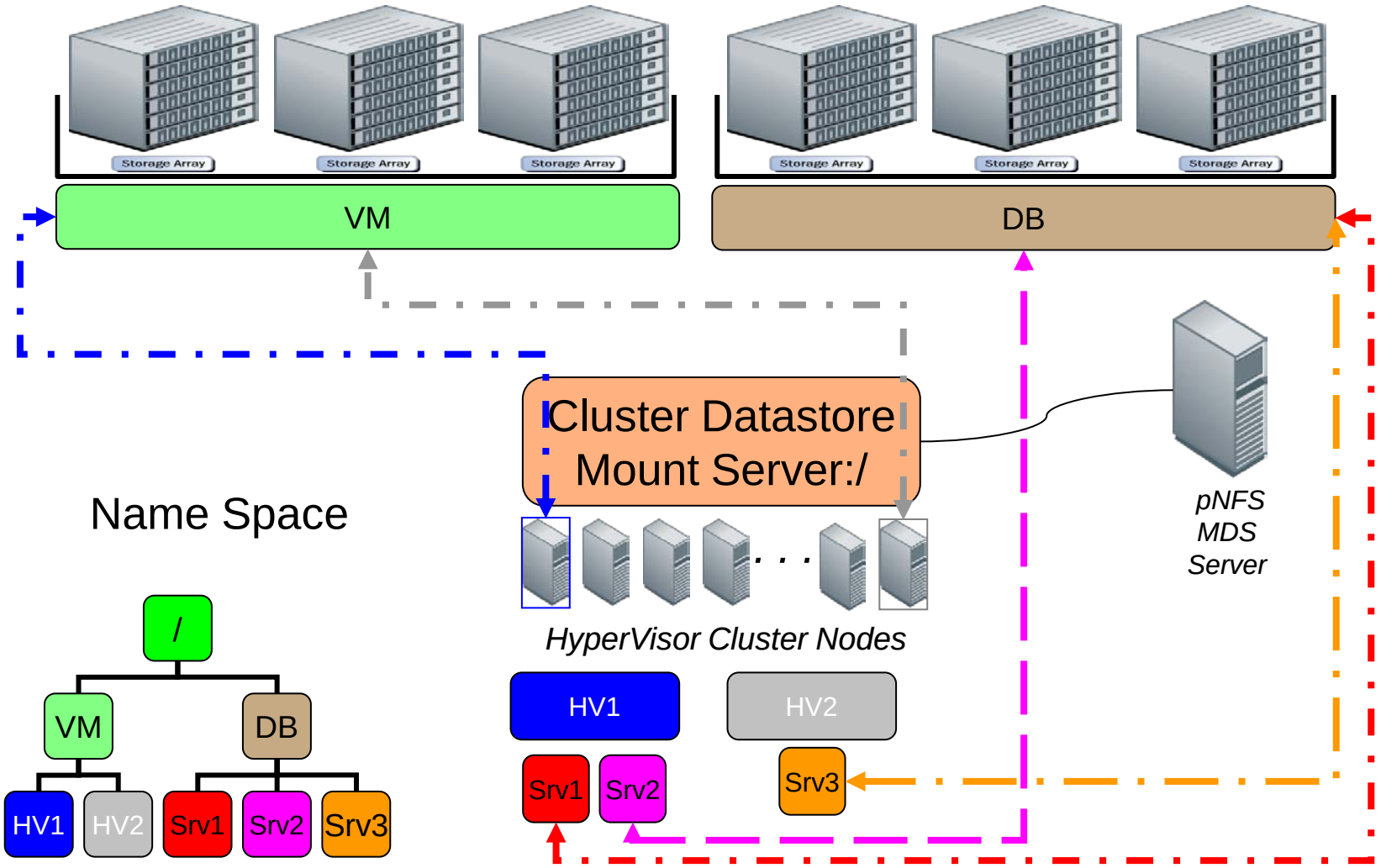
# NFSv4.1 Major Features; pNFS

- 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



# NFSv4.1: The Virtualized Datacenter



# 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,
  - O_DIRECT reads and writes are not yet supported.

# Linux Client and NFSv4.1

- 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

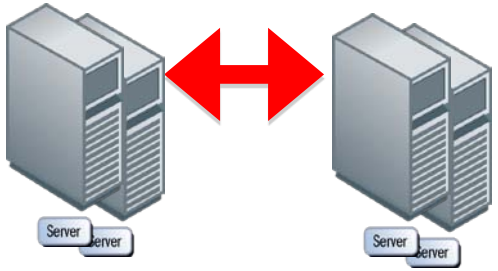
# New Features in NFSv4.2

- New direction for protocol changes

"Instead of server vendors putting in new features that might attract application developers and vendors, they're approaching server vendors requesting features that are available on local storage, but that you can't get to currently via NFS"

# New Features in NFSv4.2

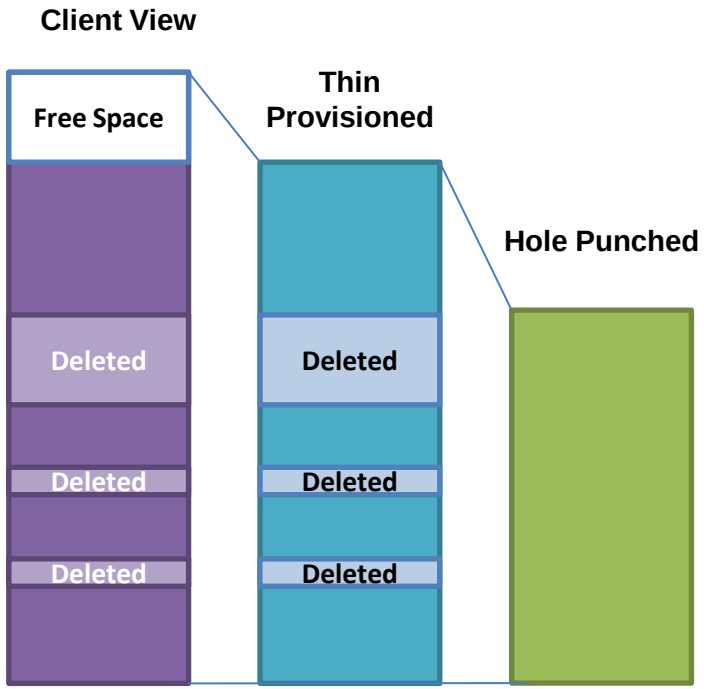
- Server-Side Copy (SSC)
  - Removes one leg of the copy
  - Destination reads directly from the source




- Application Data Blocks
  - 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.

# New Features in NFSv4.2

- Space reservation
  - Ensure a file will have storage available
- Sparse file support
  - "Hole punching" and the reading of sparse files
- Labeled NFS (LNFS)
  - MAC checks on files
- IO_ADVISE
  - Client or application can inform the server caching requirements of the file



# Filesystem Implications

- Files, blocks, objects can co-exist in the same storage network
  - Can access the same filesystem; even the same file
- NFS flexible enough to support unlimited number of storage layout types
  - Three IETF standards, files, blocks, objects
  - Others evaluated experimentally
- NAS vs SAN?
  - IETF process defines how you get to storage, not what your storage looks like
  - Each vendor can implement their pNFS system differently; the standard does not speak to the implementation

# Virtualization; The Game Changer

- Server virtualization a major area of use
  - VMware, Citrix Xen
- Demands of 1000s of images on 100s of servers
- Requirements from a storage system
  - Single system image, Resiliency, Load balanced, Transparent & non-disruptive upgrades…
- NFS a better fit to requirements than SANs
  - Use cases much wider & broader
  - Ubiquitous like Linux; available everywhere
  - Runs on widely available Ethernet & TCP/IP infrastructure

# Conclusions

- NFS has more relevance today for commercial, HPC and other use cases than it ever did
  - Features for a virtualized data centers
  - Performance, scalability, WAN security
- Developments driven by application requirements
- Adoption slow, but will continue to increase
  - NFSv4 support widely available
  - New NFSv4.1 with client & server support

# Summary/Call to Action

- pNFS is the first open standard for parallel I/O across the network
  - Ask application vendors to include NFSv4.1 support for client/servers
- pNFS has wide industry support
  - Commercial implementations and open source
- Start using NFSv4.0, NFSv4.1 today
  - NFSv4.2 nearing approval

# Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: tracktutorials@snia.org

**Many thanks to the following individuals for their contributions to this tutorial.**

**- SNIA Education Committee**

Joshua Konkle (author)
Mike Eisler, Co-Editor of NFSv4.1
J. Bruce Fields
Brian "Beepy" Pawloski, (Co-Chair, NFSv4.1)
Joe White,
Howard Goldstein,
Ken Gibson
Omer Asad
Sachin Chheda
Jason Blosil
Sorin Faibash
Rob Peglar
Dave Hitz
Dave Noveck
Peter Honeyman
Brent Welch
David Black
Piyush Shivam
Mark Carlson
Andy Adamson
Pranoop Ersani
Ricardo Labiaga
Tom Haynes
Alex McDonald
Simon Gordon