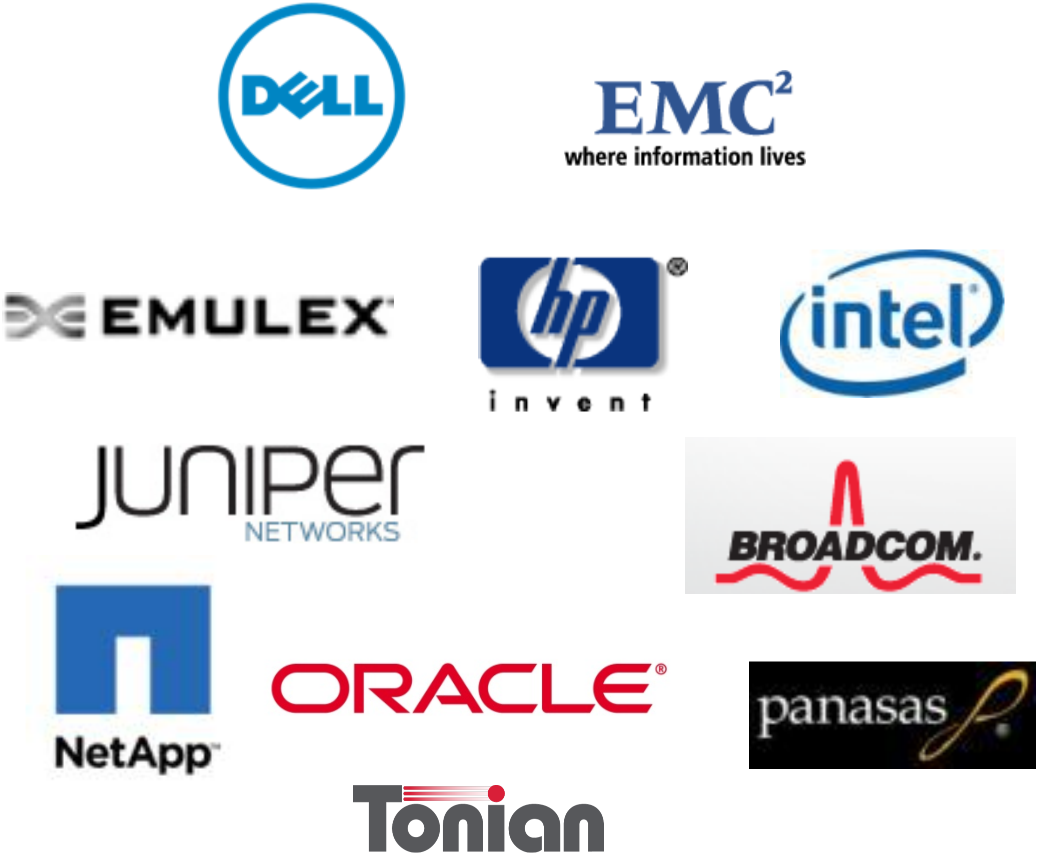
Advances in NFS; NFSv4.1, pNFS and NFSv4.2

SNIA™ WEBCAST

Presented by:
Alex McDonald
CTO Office, NetApp

HOSTED BY THE
ETHERNET STORAGE FORUM

Ethernet Storage Forum Members



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

SNIA Europe™

Webcast Presenter

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.

Alex is co-chair of the SNIA NFS Special Interest Group and co-chair of SNIA's Cloud Storage Initiative, and has a specific interest in promoting the NFS file protocol and CDMI (the Cloud Data Management Interface).

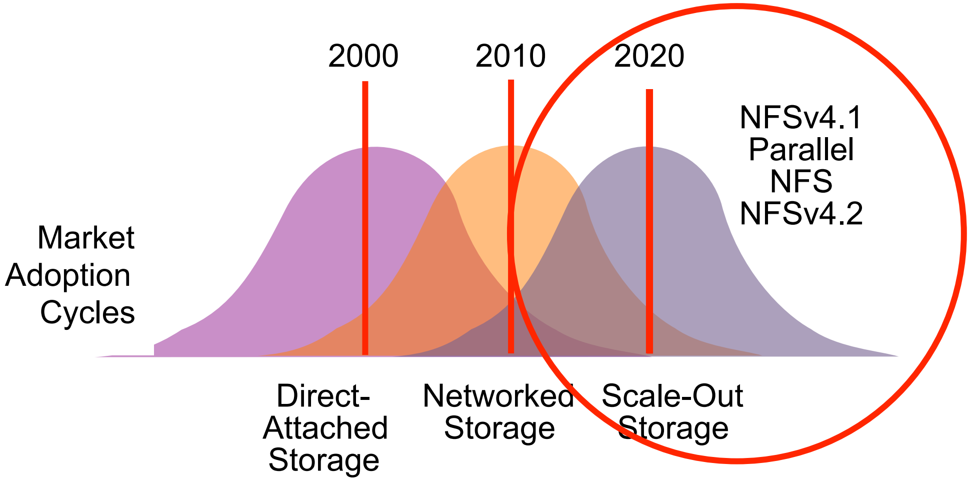
SNIA's NFS SIG

- SNIA's NFS Special Interest Group (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"

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

NFS; Ubiquitous & Everywhere

- NFS is ubiquitous and everywhere
- NFSv3 very successful
  - Protocol adoption is over time, and there have been no big incentives to change
- Industry – and hence NFS – doesn't stand still
  - NFSv2 in 1983
  - NFSv3 in 1995
  - NFSv4 in 2003
  - NFSv4.1 in 2010
  - NFSv4.2 to be agreed at IETF shortly
  - Faster pace for minor revisions
- But…



Evolving Requirements

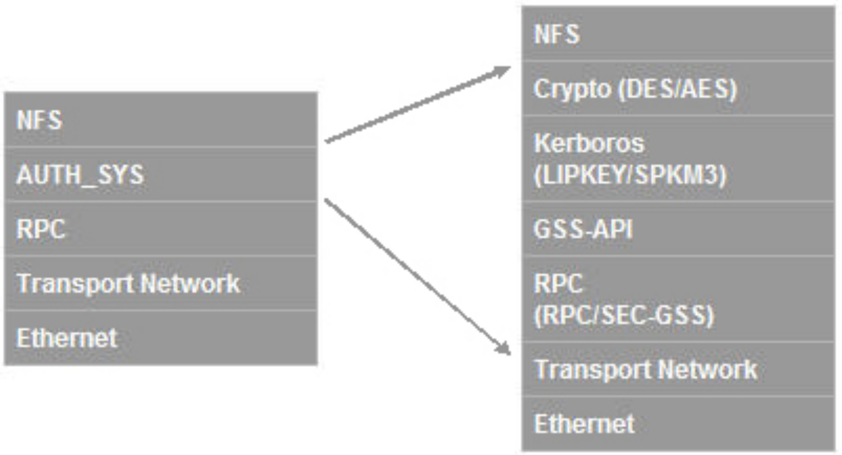
- Adoption has been slow; why?
  - Lack of clients was a problem with NFSv4
  - NFSv3 was just "good enough"
- Industry is changing, as are requirements
  - Economic Trends
    - Cheap and fast computing clusters
    - Cheap and fast network (1GbE to 10GbE, 40GbE and 100GbE in the datacenter)
    - Cost effective & performant storage based on Flash & 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
    - Beyond performance; NFSv4.1 brings increased scale & flexibility
    - Outside of the datacenter; requires good security, scalability

NFSv4 and beyond

- Areas address by NFSv4, NFSv4.1 and pNFS
  - Security
  - Uniform namespaces
  - Statefulness & Sessions
  - Compound operations
  - Caching; Directory & File Delegations
  - Parallelisation; Layouts & pNFS
- Future with FedFS and NFSv4.2
  - FedFS: Global namespace; IESG has approved Dec 2012
  - New features in NFSv4.2

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

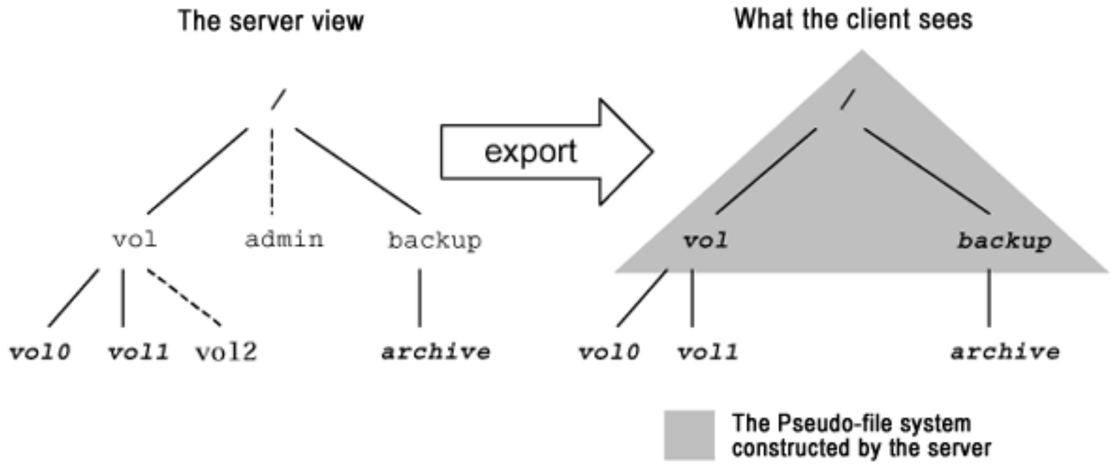


NFS
AUTH_SYS
RPC
Transport Network
Ethernet

NFS
Crypto (DES/AES)
Kerberos (LIPKEY/SPKM3)
GSS-API
RPC (RPC/SEC-GSS)
Transport Network
Ethernet

NFSv4 Major Features; Namespace

- Uniform and "infinite" namespace
  - Moving from user/home directories to datacenter & corporate use
  - Meets demands for "large scale" protocol
  - UTF-8 support for Unicode codepoints

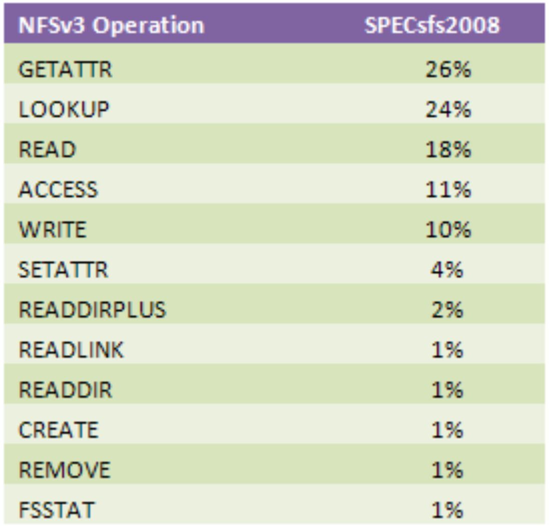


NFSv4 Major Features; Stateful Clients

- NFSv4 gives client independence
  - Previous model had "dumb" stateless client; server had the smarts
- Allows delegations & caching
- No automounter required, simplified locking
  - Mounting & locking incorporated into the protocol
  - Simplifies administration
- 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
  - Removes major source of NFSv3 irritation; stale locks

NFSv4 Major Features; Compound Operations

- NFSv3 protocol can be "chatty"; unsuitable for WANs with poor latency
- Typical NFSv3; open, read & close a file
  - LOOKUP, GETATTR, OPEN, READ, SETATTR, CLOSE
- NFSv4 compounds into a single operation
  - Reduce wire time
  - Simple error recovery



NFSv3 Operation	SPECsfs2008
GETATTR	26%
LOOKUP	24%
READ	18%
ACCESS	11%
WRITE	10%
SETATTR	4%
READDIRPLUS	2%
READLINK	1%
READDIR	1%
CREATE	1%
REMOVE	1%
FSSTAT	1%

Table 1; SPECsfs2008 %ages for NFSv3 operations

NFSv4.1 Major Features; Sessions

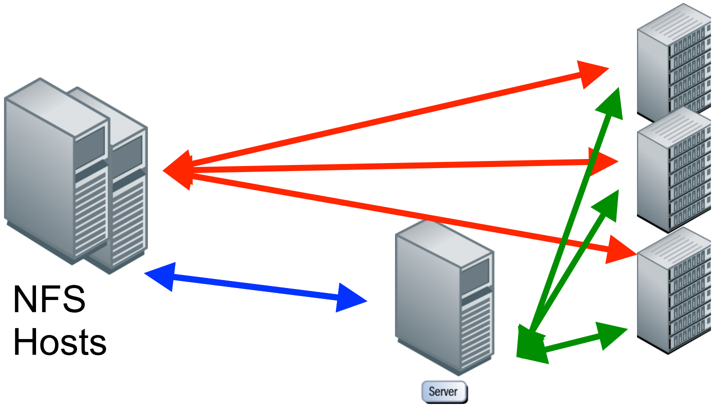
- NFSv3 server never knows if client got reply message
- NFSv4.1 introduces Sessions
  - Major protocol infrastructure change
  - Exactly Once Semantics (EOS)
  - Bounded size of reply cache
  - Unlimited parallelism
- A session maintains the server's state relative to the connections belonging to a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

NFSv4.1 Major Features; Delegations

- Server delegates certain responsibilities to the client
  - Directory & file
- At OPEN, the server can provide
  - READ delegation; server guarantees no writers
  - WRITE delegation; server guarantees exclusive access
- Allows client to locally service operations
  - E.g OPEN, CLOSE, LOCK, LOCKU, READ, WRITE

NFSv4.1 Major Features; Layouts

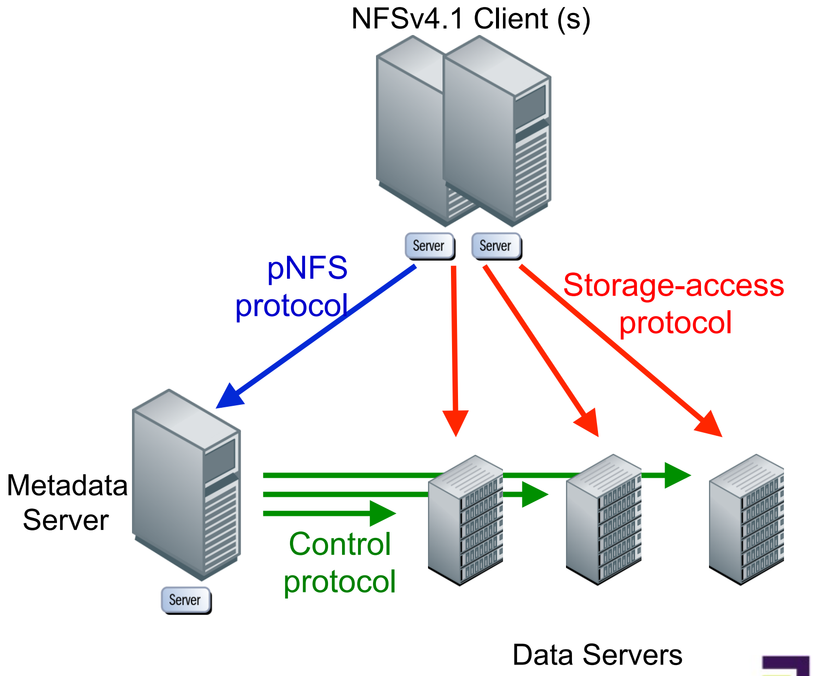
- Layouts
  - Files, objects and block layouts
  - Provides flexibility for storage that underpins it
  - Location transparent
    - Striping and clustering
- Examples
  - Blocks, Object and Files layouts all available from various vendors



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

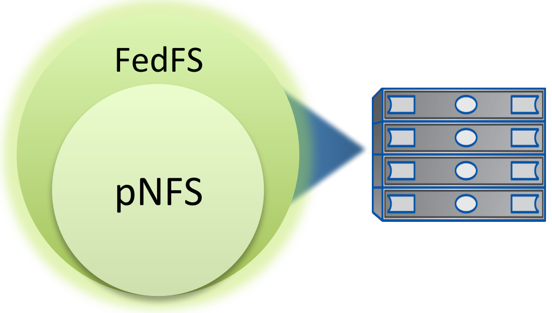


pNFS Filesystem Implications

- 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; no-one cares any more
  - IETF process defines how you get to storage, not what your storage looks like
  - NetApp pNFS implemented differently from Panasas or BlueArc or EMC or…

Federated File System: 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

What is FedFS?

- FedFS is a set of open protocols that permit the construction of a scalable, cross-platform federated file system namespace accessible to unmodified NFSv4[.1] clients.
- Key points:
  - Unmodified clients
  - Open: cross-platform, multi-vendor
  - Federated: participants retain control of their systems
  - Scalable: supports large namespaces with many clients and servers in different geographies

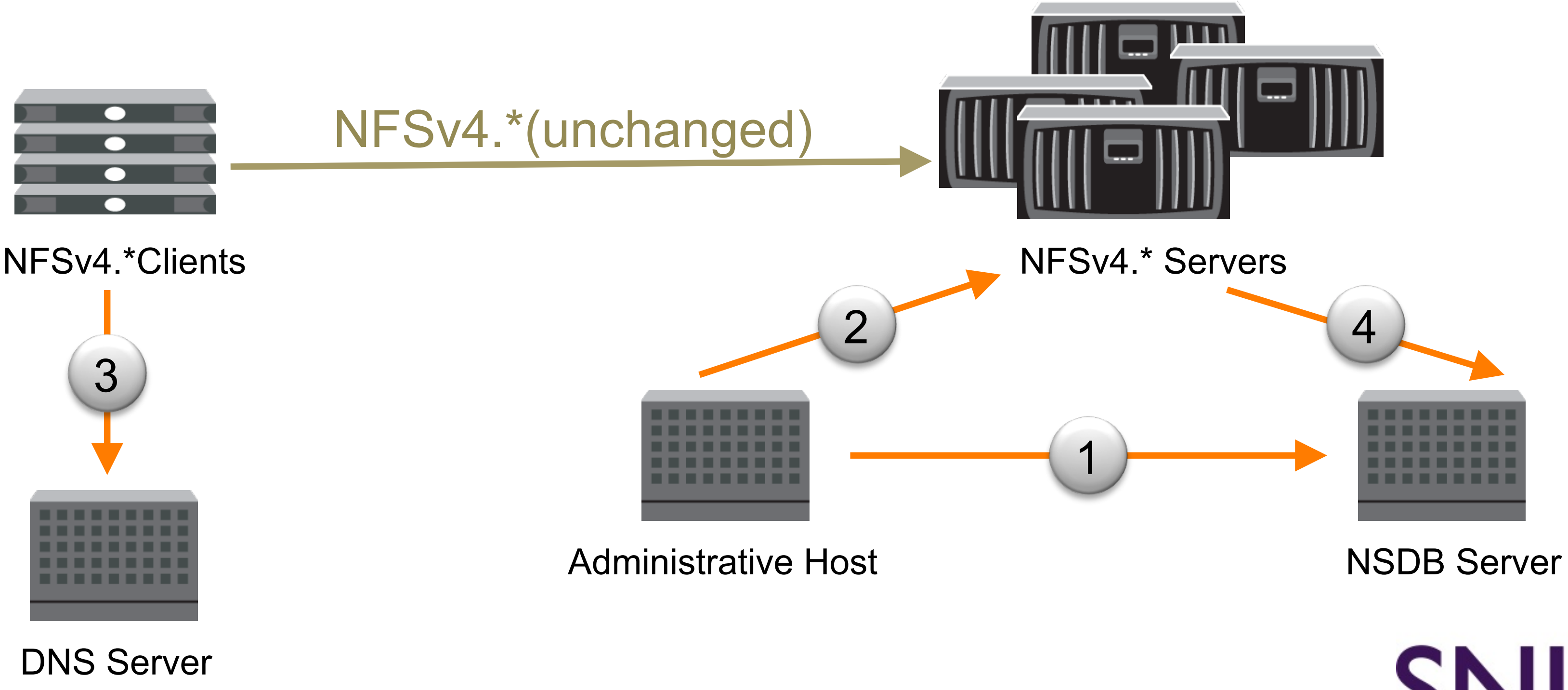
FedFS Protocols

Namespace Management

1. NSDB Management (LDAP)
2. Junction Management (ONC RPC)

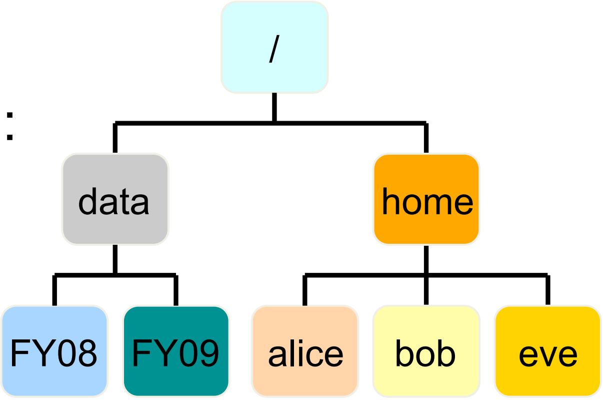
Namespace Navigation

3. Namespace discovery (DNS)
4. Junction resolution (LDAP)

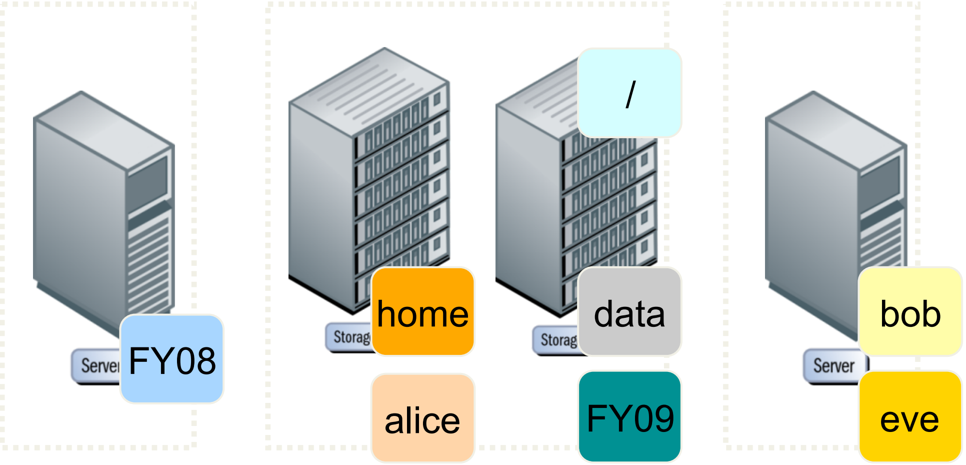


FedFS Example

The illusion:



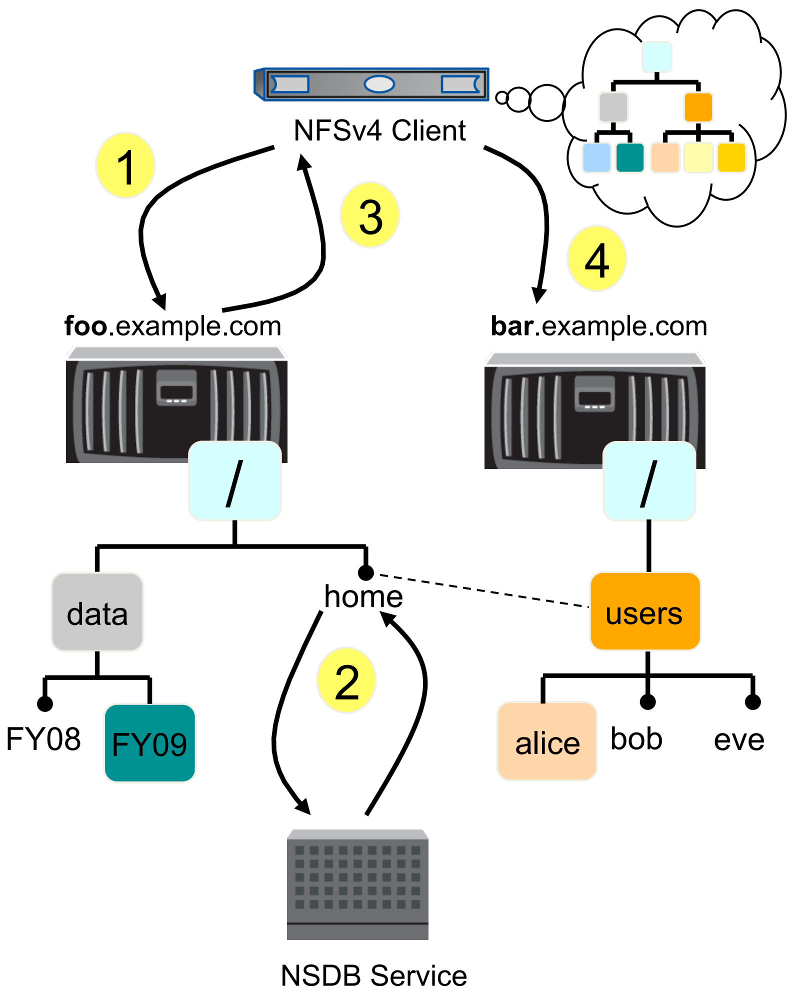
The reality:



London Frankfurt Paris

- The user and application software see a simple, hierarchical namespace
- Behind the scenes, simple management operations allow data mobility for high performance, high reliability, and high availability

FedFS Example



The user requests `/home/alice`:

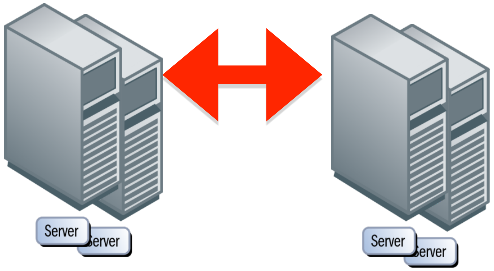
1. The client attempts to access `/home/alice` on server **foo**.
2. Server **foo** discovers that `home` is a namespace junction and determines its location using the FedFS NSDB service.
3. Server foo returns an NFSv4 referral to the client directing it to server **bar**'s `/users`.
4. The client accesses `/users/alice` on server **bar**.

Benefits of FedFS

- Simplified management
  - Eliminates complicated software such as the automounter
- Separates logical and physical data location
  - Allows data movement for cost/performance tiering, worker mobility, and application mobility
- Enhances:
  - Data Replication
    - Load balancing or high availability
  - Data Migration
    - Moving data closer to compute or decommissioning systems
  - Cloud Storage
    - Dynamic data center, enterprise clouds, or private internet clouds.

New Features in NFSv4.2

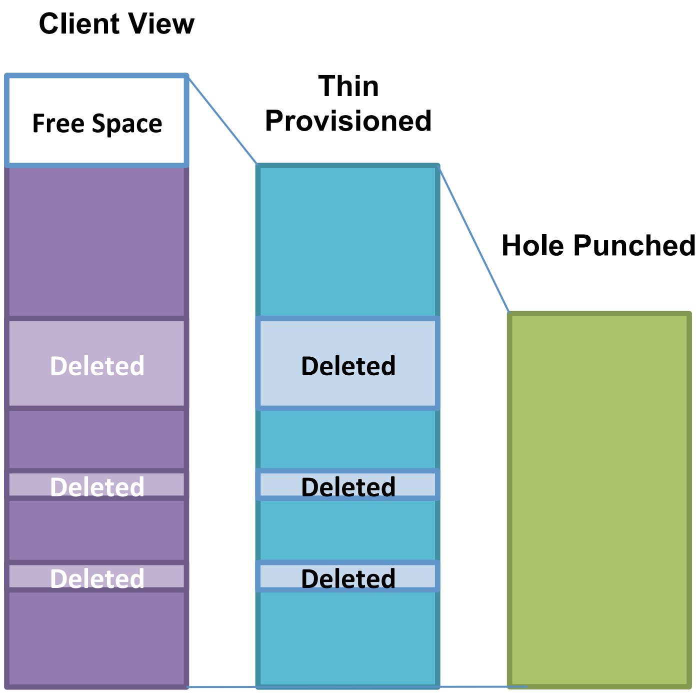
- Server-Side Copy (SSC)
  - Removes one leg of the copy
  - Destination reads directly from the source



- 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



The Four Reasons for NFSv4.1

	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

Summary/Call to Action

- pNFS is the first open standard for parallel I/O across the network
- NFSv4.1 & pNFS has industry support
  - Commercial implementations and open source
  - Ask vendors to include NFSv4.1 & pNFS support for clients & servers
- Start using NFSv4.1 today
  - NFSv4.2 nearing approval
  - FedFS brings true global namespace

Next Presentation

- **NFSv4.1: Plan For A Smooth Migration**
  - NFSv4.1 implementation steps and guidelines
  - Taking advantage of pNFS
  - Availability of NFSv4.1 and pNFS clients and servers
  - Application support for NFSv4.1 and pNFS
  - Next BrightTalk on
    - Feb 05 2013 16:00GMT, 17:00 CET

BrightTALK™

SNIA Europe™

To download this Webcast

after the presentation, go to

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

Question & Answer