SDC
STORAGE DEVELOPER CONFERENCE
SNIA ■ SANTA CLARA, 2014

Introducing FedFS On Linux

Chuck Lever
Oracle Corporation

In Brief

- FedFS problem statement
- Overview of protocols
  - DNS, NFS/SMB, NSDB, ADMIN
- The fedfs-utils command set
- Next steps

Open Source *Bona Fides*

- FedFS is a proposed Internet Standard
- fedfs-utils is a GPL implementation of FedFS for Linux
- No Oracle products were harmed in the making of FedFS or fedfs-utils

FedFS Problem Statement

- Name file resources globally
  - File names not based on storage location
- Enable client self-configuration
  - Scale-up number of file-access clients
- Enable advanced storage administration tools
  - Enable fileset migration, replication
- Distribute administration authority
  - Anyone can deploy a namespace securely
- No changes to existing file access protocols

FedFS Encapsulated

"A federated filesystem enables file access and namespace traversal in a uniform, secure and consistent manner across multiple independent fileservers within an enterprise or across multiple enterprises"

Familiar Terminology

- File-access protocol / server / client
  - Standard network file protocol
- Share
  - Basic unit of data administration on a fileserver
- Namespace
  - How applications locate files (*e.g.*, POSIX pathname)

File System Referrals

- A *referral* is a file-access protocol action where a fileserver **redirects** file-access clients to another fileserver and share

- Referrals enable power features
  - Hide physical location of shared data
  - Migration: moving filesets
  - Replication: load balancing, fault tolerance
  - Multi-server namespace

FedFS Terms

- FedFS *Fileset*
  - A share that has unique Name and Location attributes
- FedFS *Junction*
  - A non-protocol specific link between filesets
  - Contains a FedFS Fileset Name
- FedFS *Domain*
  - An independently administered file namespace and fileset namespace
- Globally Useful Name
  - File pathname that is the same on all file-access clients

FedFS Client-side

- *Globally Useful Name* is a file pathname that always refers to the same file
- A Globally Useful Name hides details of underlying storage
  - Fileserver hostname and export path are not exposed
  - Fileset replicas are transparent

/nfs4/example.net/users/cel/.vimrc

Protocol **Domain** **Directories and files**

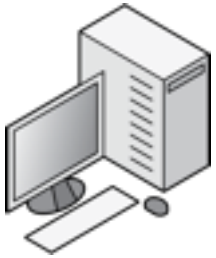
FedFS Server-side

- Junction Resolution
  - File-access client comes across a FedFS junction on a fileserver
  - Fileserver reads contents of junction (FSN, NSDB)
  - Fileserver converts these to a native referral by contacting the NSDB contained in the junction
- Namespace Database (NSDB)
  - Database mapping fileset Names to fileset Locations

FedFS Protocol Overview

- DNS protocol
  - A DNS SRV record advertises the root directory of each FedFS domain
- File-access protocol
  - File-access clients use this protocol to access files on fileservers
- NSDB protocol
  - Fileservers use this LDAP protocol to resolve FedFS junctions
- ADMIN protocol
  - Administrators use this ONCRPC protocol to manage FedFS junctions and provision NSDB connection parameters on fileservers

FedFS Operation: Mount

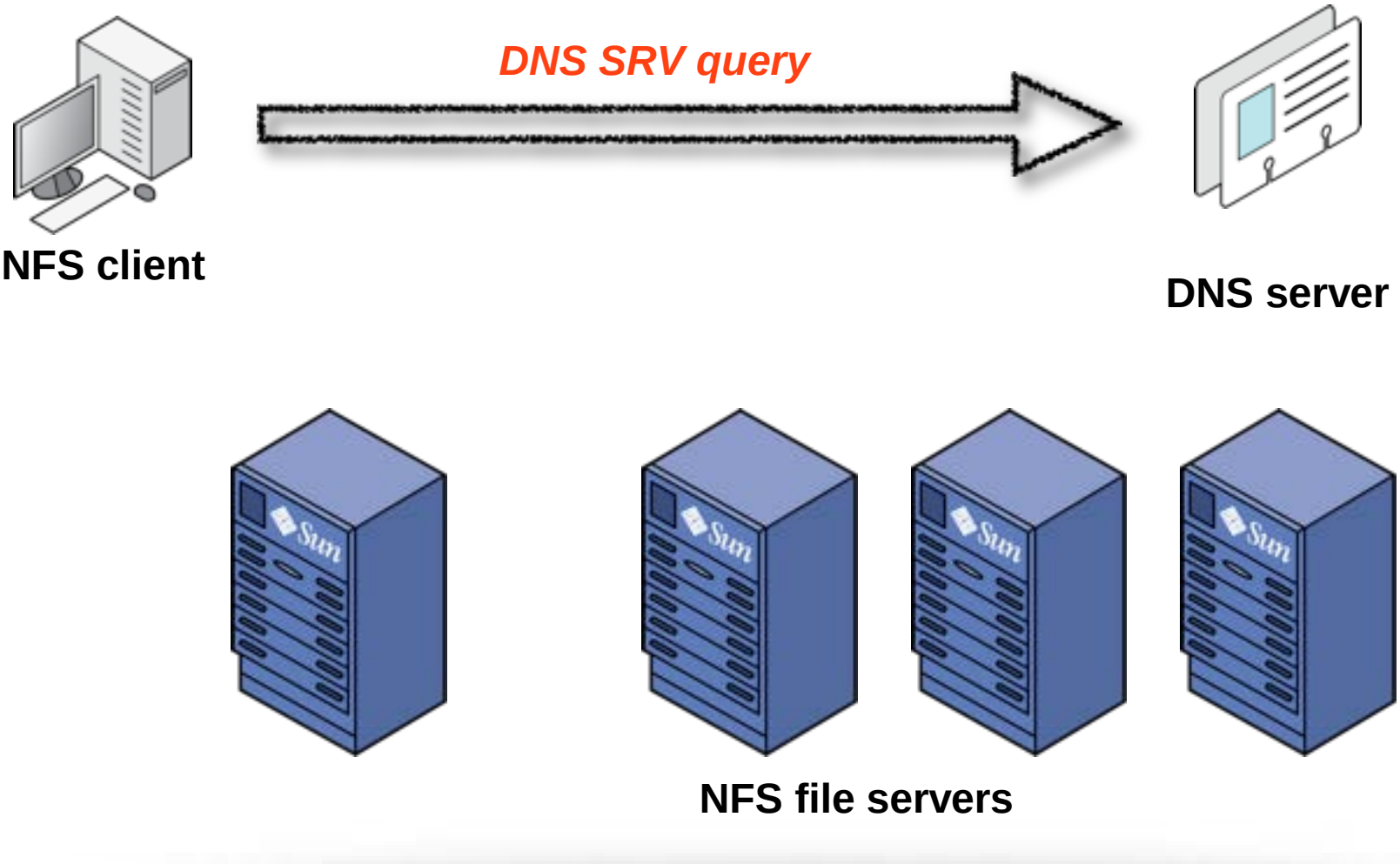


$ cd /nfs4/example.net

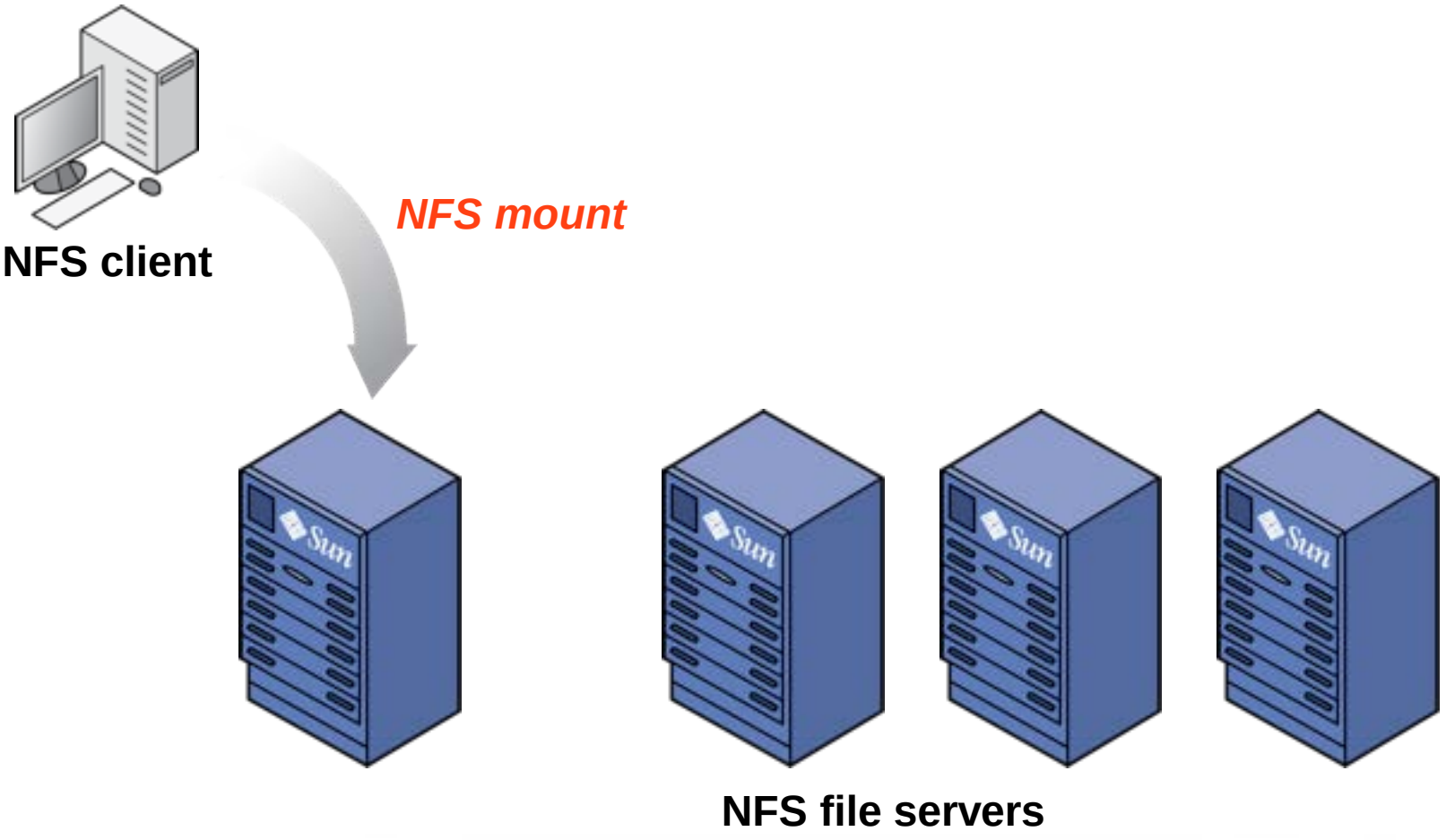
FedFS-enabled NFS client

NFS file servers

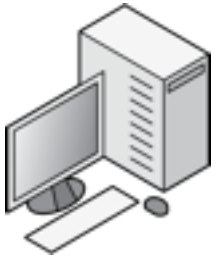
FedFS Operation: Mount



FedFS Operation: Mount



FedFS Operation: Mount

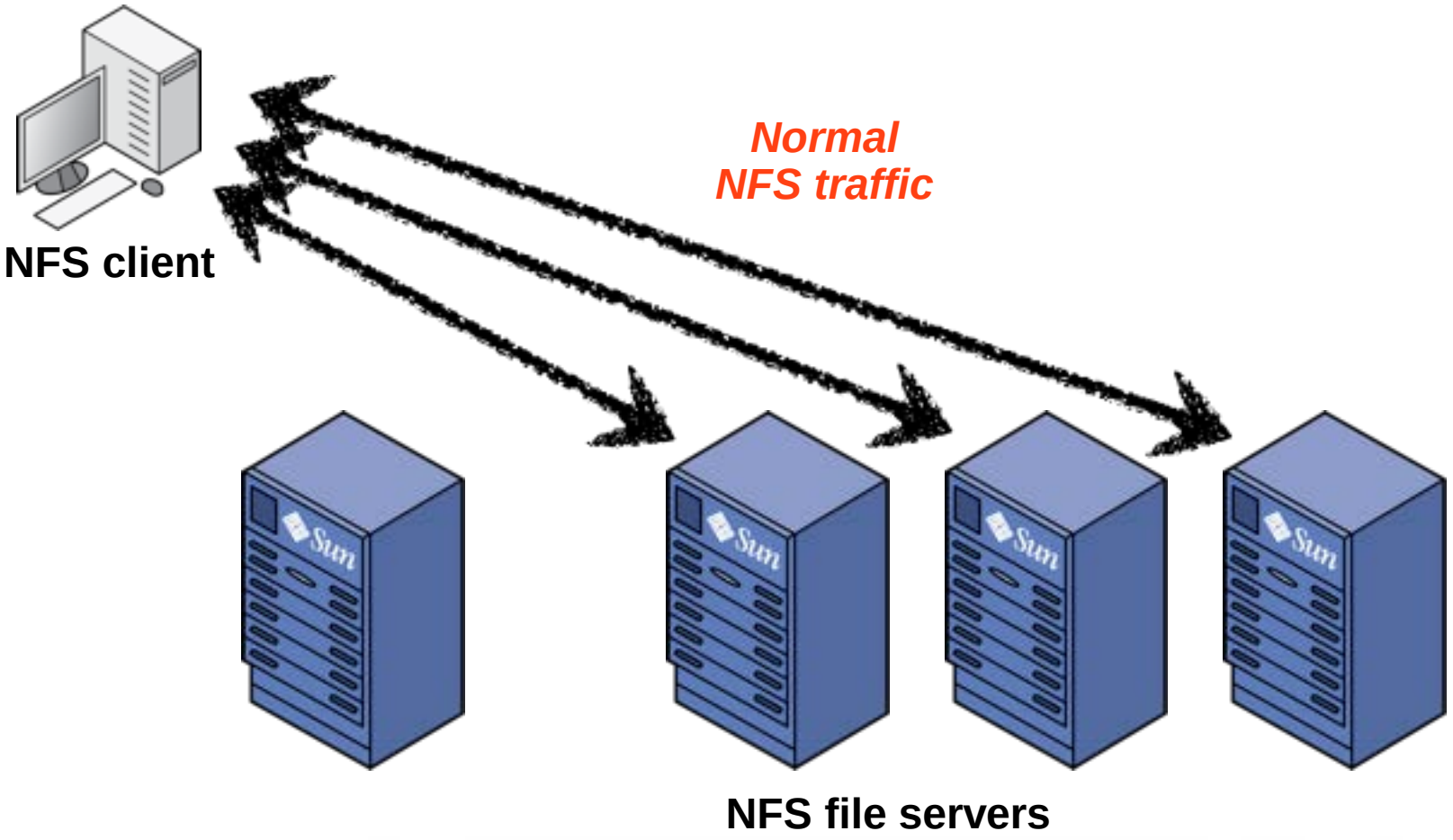


NFS client

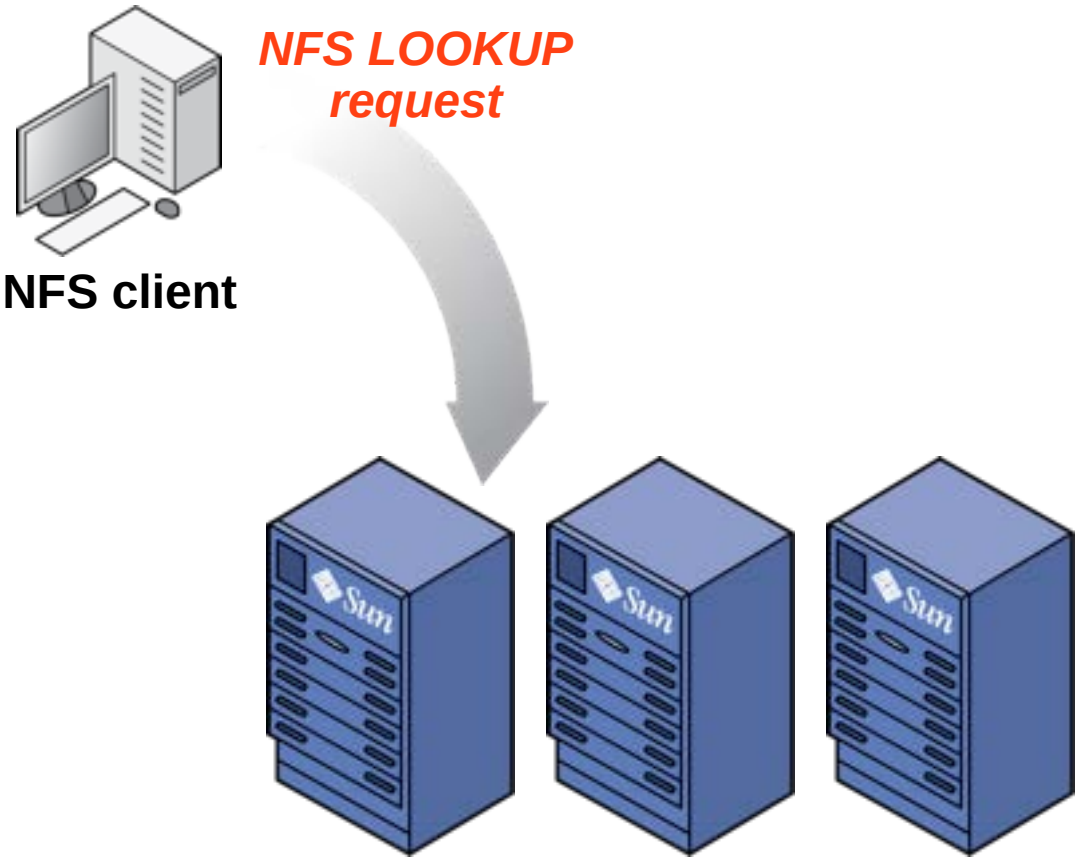
NFS referrals

NFS file servers

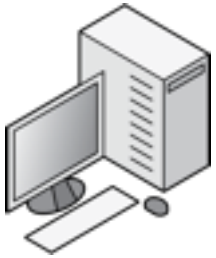
FedFS Operation: Mount



FedFS Operation: Junction Resolution



FedFS Operation: Junction Resolution

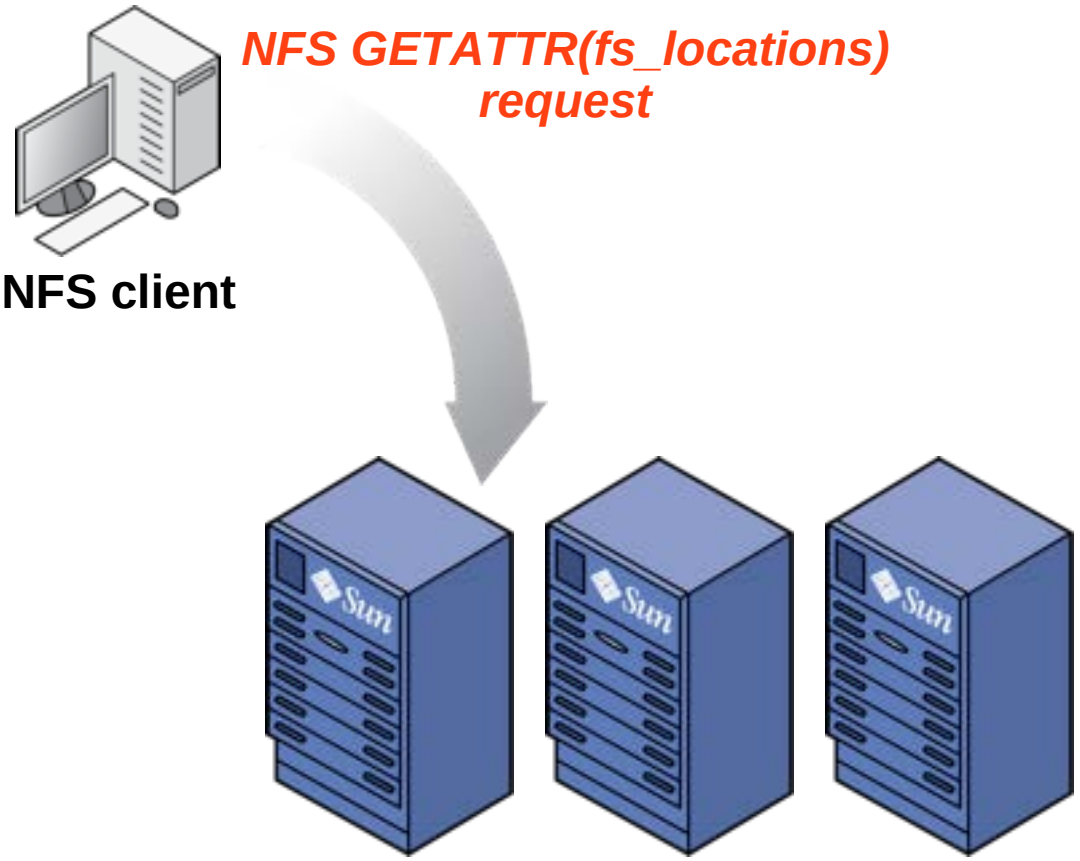


NFS client

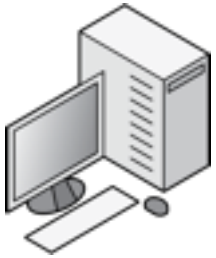
NFS4ERR_MOVED

FedFS-enabled fileservers

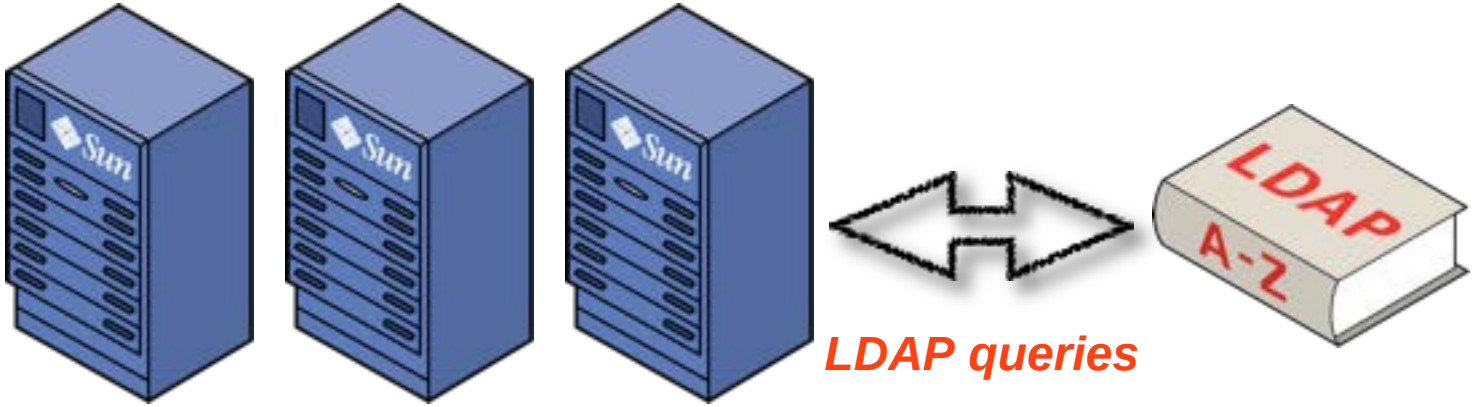
FedFS Operation: Junction Resolution



FedFS Operation: Junction Resolution



NFS client

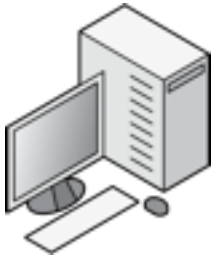


LDAP queries

FedFS-enabled fileservers

NSDB

FedFS Operation: Junction Resolution



NFS client

Location list

FedFS-enabled fileservers

Overview Of Linux fedfs-utils

- Not an in-kernel filesystem
  - Set of user space tools, plus an LDAP schema
  - Relies on automounter on file-access clients
- Does not store files, but rather locates them
  - Relies on standard network filesystem for data storage
  - Non-FedFS clients can mount and access shares normally
  - Constructs enterprise-wide file namespaces
  - Scalable referral manager

The fedfs-utils Client Command Set

- fedfs-map-nfs4
  - automounter program map
- mount.fedfs
  - For static mounts (rarely used)

The fedfs-utils Fileserver Command Set

- nfsref
  - Command line management of basic NFS referrals and FedFS junctions
- rpc.fedfsd
  - Remote management of FedFS junctions
  - Remote provisioning of NSDB connection parameters
  - Secured via RPCSEC GSSAPI
  - Entirely optional
- fedfs-domainroot
  - Command line provisioning of FedFS domain root directories

Sidebar: Basic NFS Junctions

- FedFS junction
  - Referral target locations stored on NSDB
  - Location metadata managed remotely via FedFS
- Basic NFS junction
  - Referral target locations stored directly on fileserver
  - No NSDB or other FedFS infrastructure required
- nfsref command handles both types

The fedfs-utils Admin Command Set

- nsdbparams
  - Command line management of local NSDB connection parameters
- nsdb-*
  - Remote management of NSDB content
  - Can use TLS for security
- fedfs-*
  - Remote management of FedFS junctions and NSDB connection parms
  - Can use RPCSEC GSSAPI for security
- nsdb-jumpstart
  - Command line provisioning of local NSDB based on OpenLDAP
  - Can create self-signed certificate to enable TLS-only access

Next Steps

- Ratification of pending FedFS standards
- Multi-realm file-access security
- SASL GSSAPI for NSDB access
- GUI administrative tools
- Enable NFSv4.1 features
  - Client and server support for GETATTR(fs_locations_info)

Where To Get fedfs-utils

- Source code
  - http://git.linux-nfs.org/?p=cel/fedfs-utils.git;a=summary
  - http://git.linux-nfs.org/?p=cel/fedfs-releases.git;a=summary
- Linux distributions
  - Fedora and Enterprise Linux 7: yum list fedfs-utils*
  - Enterprise Linux 6: EPEL 6
  - SuSE: coming soon

Thanks for your interest!

Appendix

Linux On-Disk Junction Format

- Directory with special mode bits
  - On server, mode looks like d - - - - - - - - - t
- Trusted xattr contains junction data
  - NFS basic: XML containing location list
  - FedFS junction: XML containing FSN UUID and NSDB hostname

Linux On-Disk Junction Format

- NFS basic junctions
  - Replaces refer=location_list in /etc/exports
  - Location list stored in junction on fileserver
  - Only NFS locations can be specified

- FedFS junctions
  - Location list stored in an NSDB, accessed via LDAP
  - Locations can be on any supported network filesystem

Client-side Junction Behavior

- On referral-enable file-access clients, junction object is automatically replaced by root directory of referred-to fileset
- On incapable file-access clients, junction object looks like an inaccessible directory

NSDB Schema

- NSDB Container Entry (NCE) record
  - Root of DIT containing FedFS information
  - DN: ou=fedfs,dc=example,dc=net
- Fileset Name (FSN) record
  - One per FedFS fileset
  - DN: fedfsFsnUuid=${FSNUUID},${NCE}
- Fileset Location (FSL) record
  - One for each physical instance of a FedFS fileset
  - Stores fileserver hostname and export path attributes
  - DN: fedfsFslUuid=${FSLUUID},fedfsFsnUuid=${FSNUUID},${NCE}