

DFS Namespaces

Dan Lovinger
DFSN Development Lead
Microsoft

Virtualization for Remote File Systems

Agenda

- ❑ Overview of DFS Namespaces
- ❑ Terminology
- ❑ General Architecture
- ❑ Geo-Awareness
- ❑ Operation and Protocol
- ❑ Domain Metadata Store
- ❑ Examples

- ❑ DFS/N is part of Windows' native SMB/CIFS protocol
- ❑ Provides a service which lets clients access a file space on local and geographically distributed file servers
 - ❑ through a single, consistent, locally available folder structure
 - ❑ without user knowledge of the remote servers
 - ❑ with availability against namespace and storage failure
 - ❑ while being directed to optimal targets
- ❑ Value Proposition: enables deployment of large numbers of servers (or clusters) while providing availability and the benefits of a single namespace

- ❑ DFS/N has four component buckets
 - ❑ a client extension to the SMB redirector
 - ❑ a server-side service attached to the SMB server
 - ❑ a domain controller role for locating DFS/N service
 - ❑ API/CLI/GUI management and diagnostic tools
- ❑ As of Windows 2008, the DFS/N service is an optional role service for the File Server role

DFSN Is Not ...

- ❑ A few common DFSN misperceptions
- ❑ ... replication – in Windows, that comes from the DFSR technology
 - ❑ any replicator can be plugged under DFSN
- ❑ ... a “union” system that merges two shares into one view
\\server1\share ∪ \\server2\share ⇒ ...
- ❑ ... a clustered file system with integration between the replicas

Common Scenarios

- ❑ Simplified and persistent presentation of storage: user types [\\contoso\public](#) on their local machine and gets to the nearest server for all content underneath. The 100's of servers may change or move but user's and tool's paths can remain the same.
- ❑ Publication: users publish corporate data on [\\contoso\marketing\announce](#). The announce folder is hosted on multiple servers around the globe. Data and folder information is replicated and made available globally.
- ❑ Transparent consolidation: [\\oldserver](#) is decommissioned and its file service relocated to a new server infrastructure using DFSN consolidation namespaces. Clients continue to access data using the old pathnames.

Key Elements of the DFSN Service

- ❑ DFSN's service is a database of namespaces and their links and targets
- ❑ Functions at both the server and domain level
 - ❑ on the server, responsible for local namespaces
 - ❑ at the domain, responsible for locating domain namespaces, operating as a proxy for the DS store
 - ❑ ... also responsible for providing clients with a local domain list for domain identification.
- ❑ DFSN's databases are accessed by a simple query protocol – the “referral”. Each comes with a refresh TTL.
- ❑ On the namespace server, intercepts client access
 - ❑ DFSN attaches to the SMB server for each namespace share
 - ❑ DFSN handles opens which cross links, notifying the client via a distinguished error. This initiates the referral process.

Breaking down the Path


□ Domain Namespaces

\\domain \root\...\link \...\finance.xlsx



□ Standalone/Consolidated Namespaces

\\server \root\...\link \...\finance.xlsx



- Note: UNC paths do not differentiate domain and normal server paths
- DFSN identifies domains based on the client's home domain's relationships to its peer domains

- ❑ DFSN provides a domain identification and location service, reusing access algorithms in the client
- ❑ Identification
 - ❑ all local forest domains
 - ❑ domains in peer forests with a transitive trust relationship
- ❑ Location
 - ❑ enumeration of registered domain controllers of each identified domain
- ❑ Protocol is more general than implementation

- ❑ Summarizing the terminology ...
- ❑ **ROOT** – SMB share that hosts a set of **LINKS**
- ❑ Two base types, defined by how it is located
 - ❑ **DOMAIN** – located through Directory Services
 - ❑ \\domain\root\...
 - ❑ \\contoso.com\release, \\snia.org\files
 - ❑ **STANDALONE** - located on a single server or cluster
 - ❑ \\server\root\...
 - ❑ \\groupserver\users, \\teamcluster\builds

DFSN Terminology (cont.)

- ❑ **LINK** – path on a root pointing to **TARGETS**
- ❑ **TARGET** – any UNC path
 - ❑ SMB, NFS or DAV shares, other roots
- ❑ Roots also have targets
 - ❑ Domain – servers which host the root shares
 - ❑ Standalone – itself, as a means of self identification
- ❑ **INTERLINK** – link whose target is a **ROOT**
- ❑ **NAMESPACE** – one or more roots linked together to form a virtualized share spanning them and their link targets

Referral Terminology

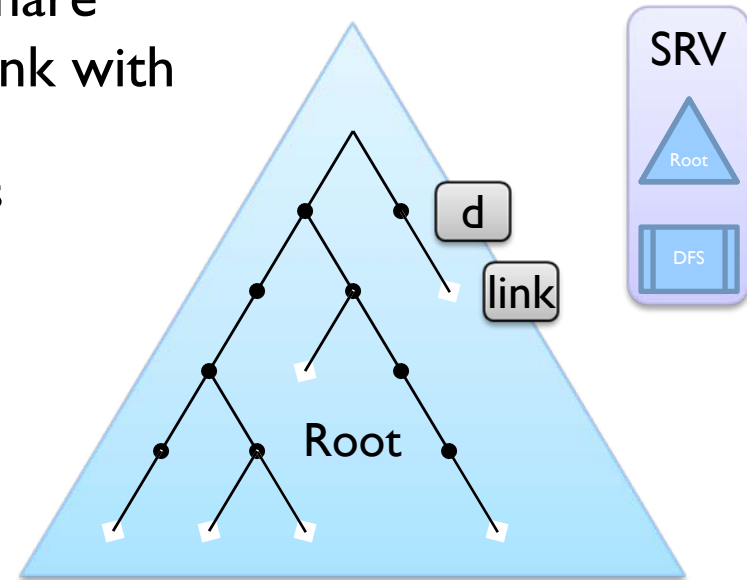
- ❑ DFSN database query
- ❑ **REFERRAL** – a **REQUEST** (query) to DFS resulting in a **RESPONSE** with its targets
 - ❑ request: \\contoso\release
 - ❑ response: \\contoso\release → $\left[\begin{array}{l} \underline{\\cosrv1\release} \\ \underline{\\cosrv2\release} \end{array} \right]$
- ❑ Four basic types
 - ❑ Link: link path → link targets
 - ❑ Root: root path → root targets
 - ❑ DC: domain → DCs of a domain
 - ❑ Domain: <null> → known domains
- ❑ Domain/DC form the domain identification and location service for DFSN clients

- ❑ DFSN clients process referrals as ordered lists
 - ❑ try the first, second ... nth target ... until success
 - ❑ on success, use the given target for subsequent IO
 - ❑ failover criteria – is error condition one expected **not** to be replicated to all other targets?
 - ❑ File not found v. Connection failed
- ❑ Ordered referral access allows the server to ...
 - ❑ direct clients through targets in geo-aware order
 - ❑ ... and fail *back* to local targets later
 - ❑ designate site or global targets of last resort
 - ❑ provide a low cost load leveling for targets

- ❑ Active Directory defines sites as a set of IP subnets, along with a graph of site to site costs. Sites are usually geographic.
- ❑ DFSN uses the costs to direct WAN aware access
 - ❑ client IP address → site
 - ❑ target IP addresses → sites
 - ❑ compute site to site costs for each client/target pair, sort
- ❑ Also provides optional policy to restrict clients to targets within their local site (insite)
- ❑ DFSN Protocol support
 - ❑ site cost boundary information in referrals (V4/2003 SPI)
 - ❑ (optional) periodic target failback checks
 - ❑ key: ordered referral access

Elements of a Root

- ❑ Recall that a root is hosted on an SMB share
- ❑ DFSN creates directory trees for each link with NTFS reparse points at the tips
 - ❑ if \\srv\root\d\link is a link, DFSN creates the directory “d” and a reparse point inside of it, “link”
- ❑ DFSN uses the reparse point for low-cost notification when an access crosses a link, triggering the referral sequence
- ❑ Of note:
 - ❑ There is nothing special about having a link somewhere below the root directory of the share
 - ❑ A root can link to another root; however, there is no such thing as a link inside of a link
 - ❑ Content can be placed anywhere in the root.

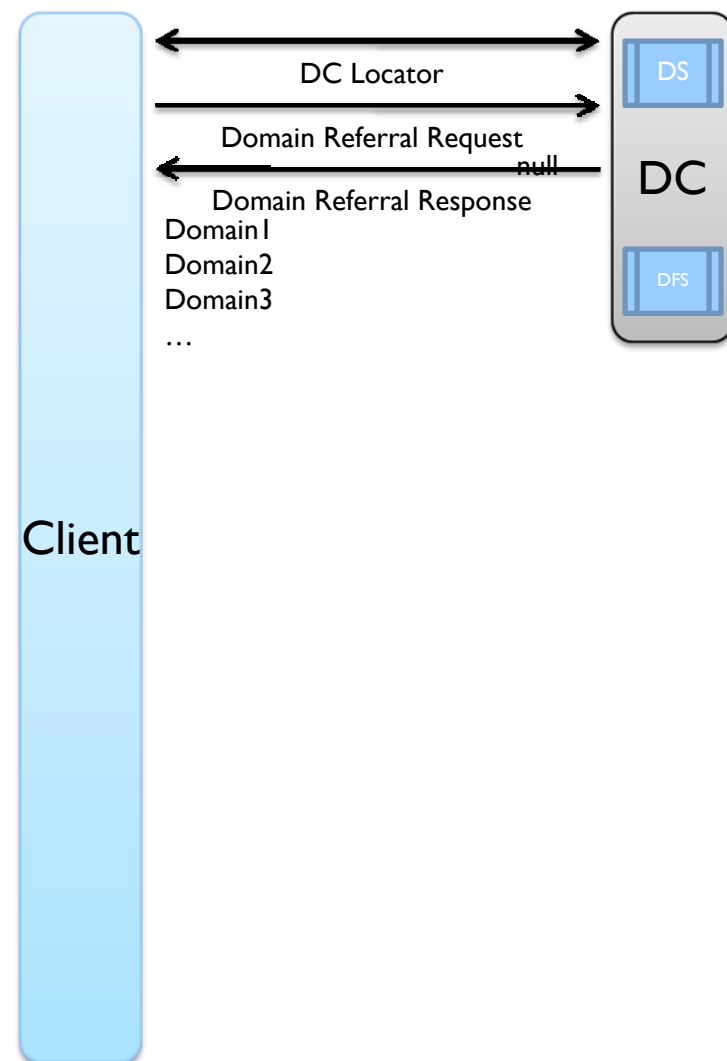


Key Elements of DFSN Access

- ❑ A DFSN client uses the DFSN service's database to crack opens
- ❑ Given [\\a\b\c\...](#), DFSN identifies if “a” is a domain name – **domain referral**
- ❑ Resolve the Namespace
 - ❑ First component is domain: e.g. [\\domain\b\c\...](#)
 - ❑ Locate domain services for “domain” – **DC referral**
 - ❑ Query its DFSN domain services for the location of “b” – **root referral**
 - ❑ ... if not ⇒ fail
 - ❑ First component is non-domain: e.g. [\\server\b\c\...](#)
 - ❑ Query DFSN on the server as to whether “b” is a namespace – **root referral**
 - ❑ ... if not ⇒ pass it over for normal resolution (plain SMB, DAV, NFS, etc.)
- ❑ Resolve the Rest
 - ❑ Tries to connect
 - ❑ On success, done!
 - ❑ On reaching a link, queries for the next servers – **link referral**
 - ❑ ... repeat
- ❑ Let's visualize the protocol ...

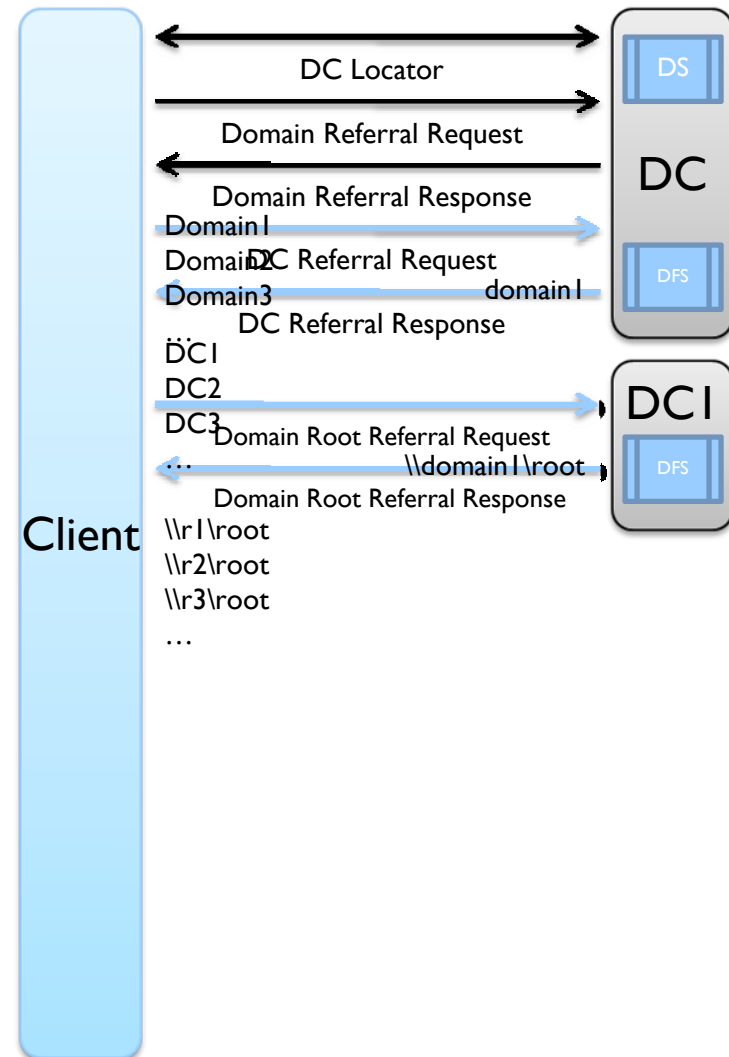
Domain Bootstrap

- At bootstrap, a DFSN client initializes against the DFSN domain services
 - discover a home domain DC
 - query the list of known domains – this forms the list of \\domain path prefixes it will recognize



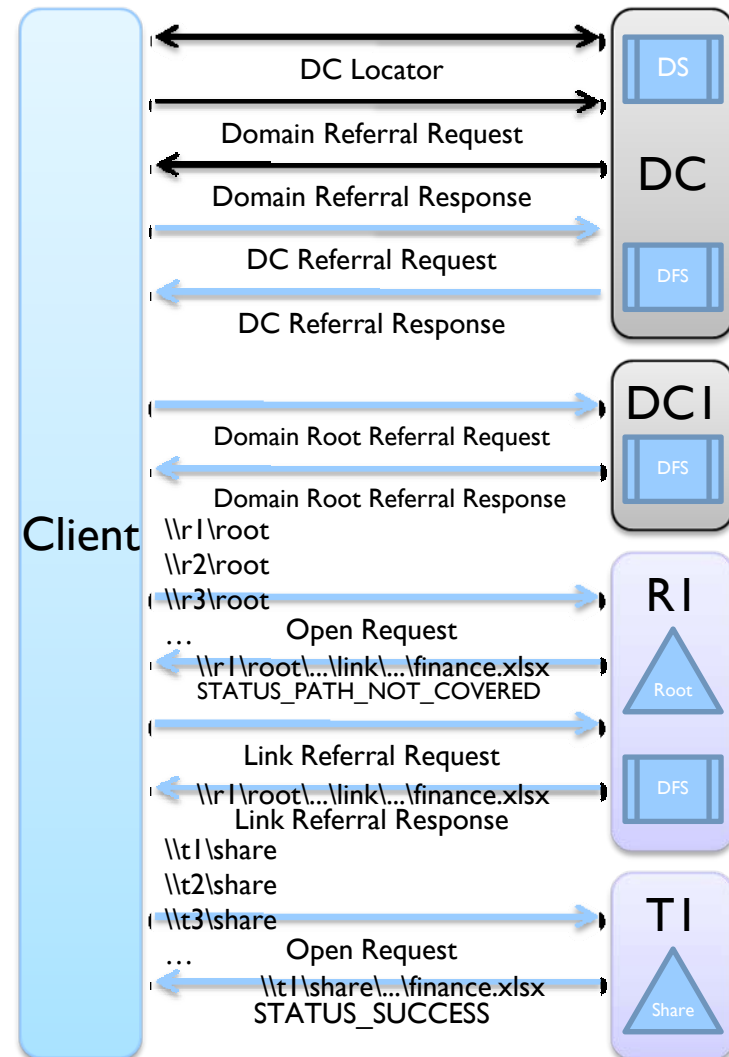
Discovering Domain Roots

- Now find \\domain I\root
 - DC referral – get DCs for Domain I
 - Choose a DC – DCI
 - Domain root referral – get list of root targets for \\domain I\root



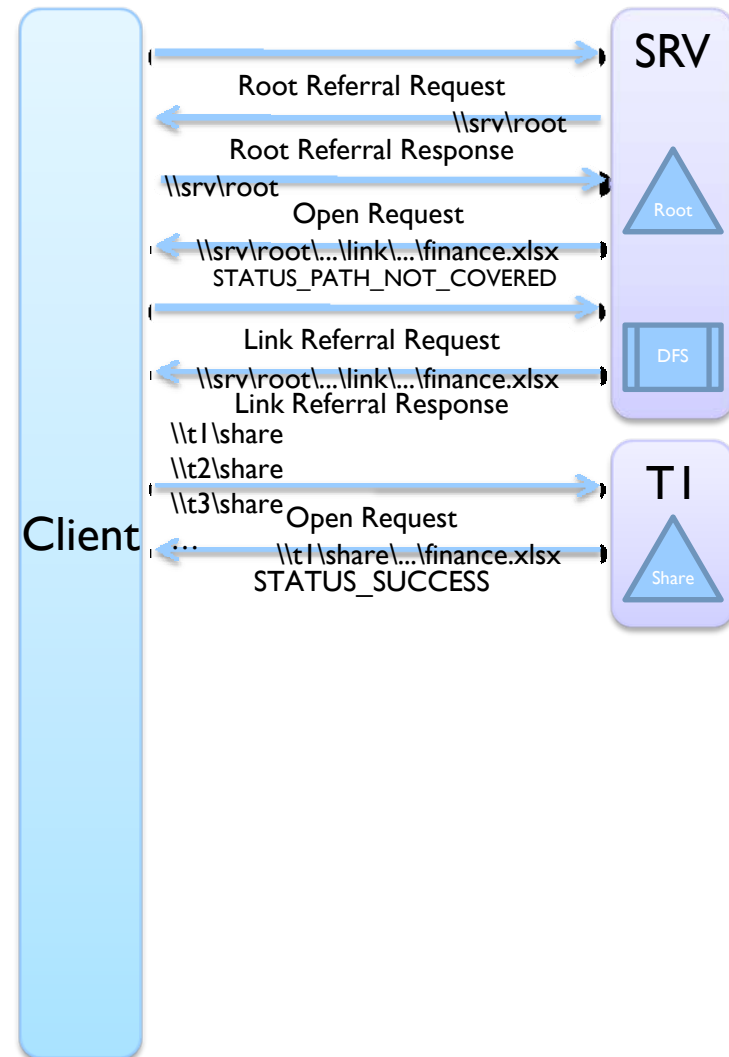
Domain Root (cont.)

- ❑ Now complete opening
`\\domain1\root\...\link\...\finance.xlsx`
- ❑ Choose a root target –RI
 - ❑ Issue open request
 - ❑ Issue link referral request
- ❑ Choose a link target –TI
 - ❑ Issue open request
- ❑ Success



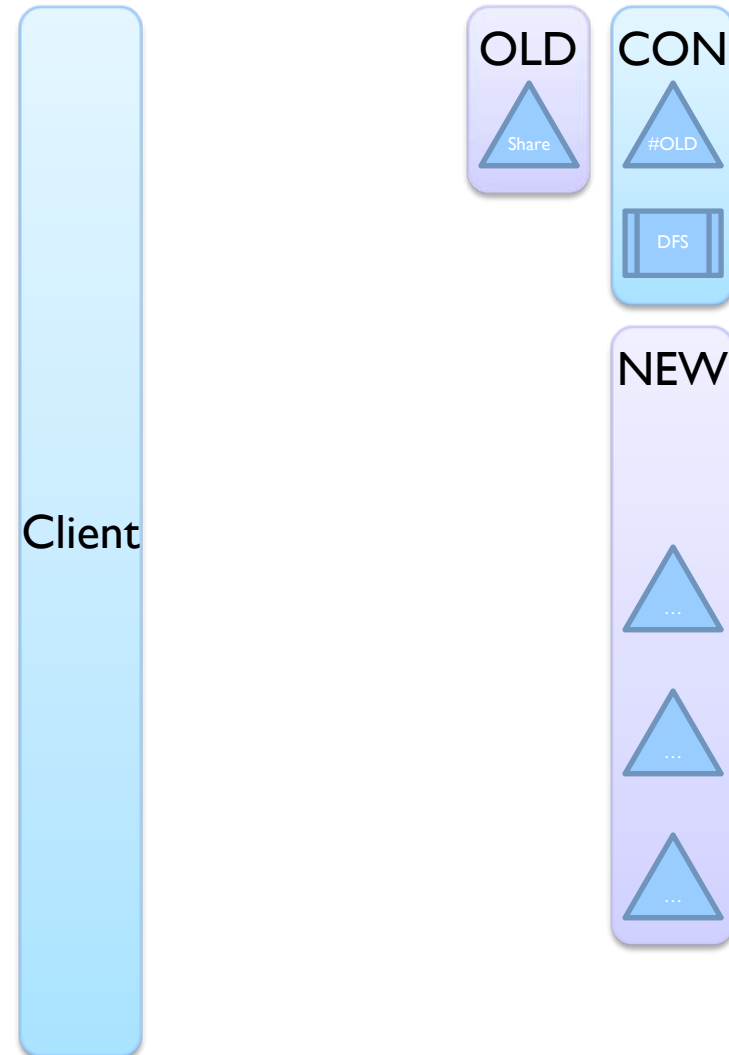
Discovering Standalone Roots

- ❑ For non-domain opens, the client first asks if it is DFS
 - ❑ Get a root referral
- ❑ Confirmed - a DFS root
 - ❑ Issue open request
 - ❑ Issue link referral request
- ❑ Choose a link target – TI
 - ❑ Issue open request
- ❑ Success



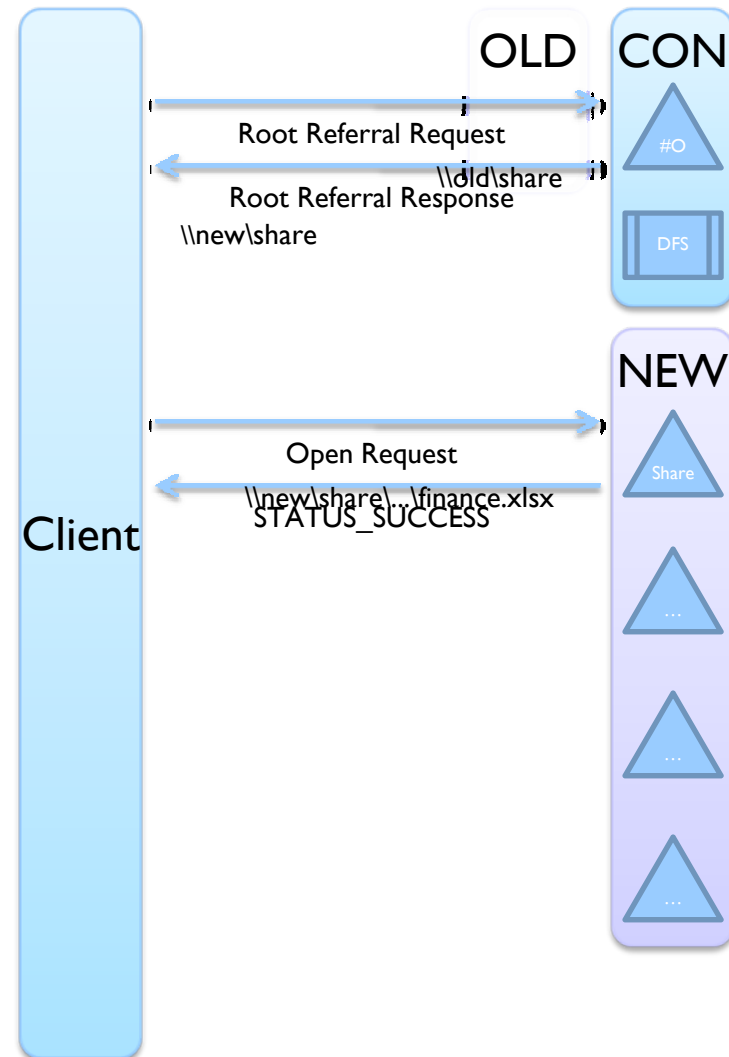
Discovering Consolidation Roots

- ❑ The fact DFSN clients query for standalone roots – and use the response – is the basis of consolidation
- ❑ A consolidation root is a special namespace that DFSN uses to simulate a retired server, here “OLD”
- ❑ OLD will be consolidated in 4 steps
 - ❑ share is transferred to NEW
 - ❑ DNS record re-pointed to CON, the DFSN consolidation server
 - ❑ root created on CON - #OLD
 - ❑ links are created in [\\con\#old](#) that map each share which used to exist on OLD to their new locations, like this:
 - ❑ [\\con\#old\share](#) → [\\new\share](#)



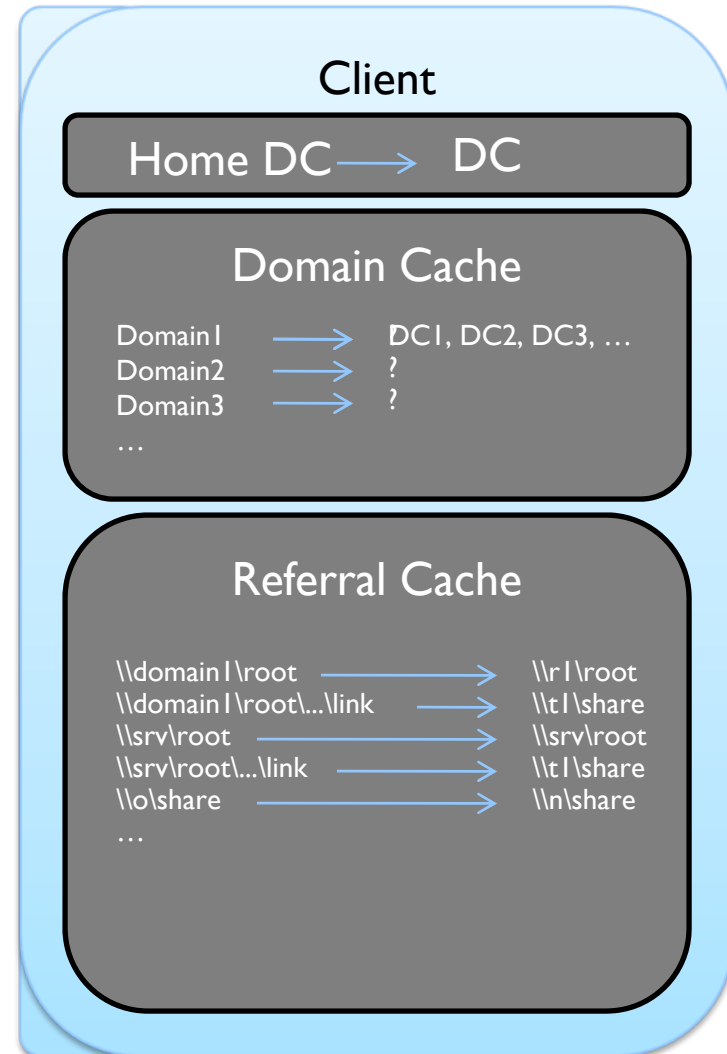
Consolidation Root (cont.)

- ❑ Now complete opening `\\old\share\...\finance.xlsx`
 - ❑ Get a root referral
 - ❑ DFS recognizes OLD is consolidated, and uses `\\con\#old` to determine where `\\old\share` should be redirected.
 - ❑ Note that `\\new\share` is returned
 - ❑ Issue open request
- ❑ Success
- ❑ NEW can host the shares of many retired servers
- ❑ Shares of a retired server can be spread among any number of servers



DFS Client Caches

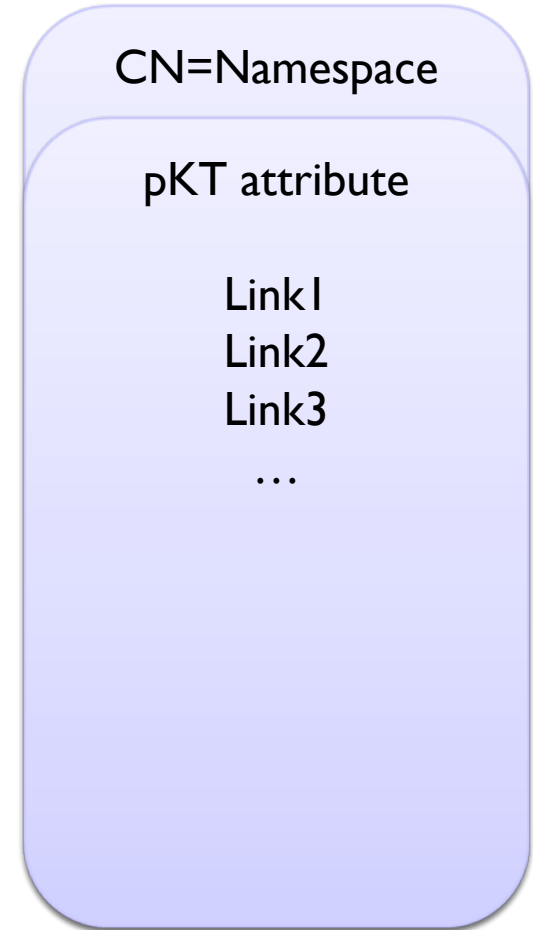
- ❑ Simplest way to visualize how your client is traversing namespaces too
- ❑ Three caches
 - ❑ Home DC
 - ❑ Domain Cache
 - ❑ Referral Cache
- ❑ Home DC is refreshed on a 15 minute timer from the WorkStation Service
 - ❑ triggers domain list refresh
- ❑ Caches can be seen via DFSUTIL
 - ❑ *dfsutil cache referral*
 - ❑ *dfsutil cache domain*
- ❑ .. and the explorer property page (DFS tab)



- ❑ DFSN has had three database stores
 - ❑ Standalone – key/value tree in the registry
 - ❑ Domain 2000 mode – database marshaled as a binary blob attribute on an object in AD via LDAP
 - ❑ Domain 2008 mode – object tree in AD via LDAP
- ❑ Standalone is the original store
 - ❑ Scales to 50,000 links per namespace
 - ❑ HKLM\Software\Microsoft\Dfs\Roots\Standalone
- ❑ Domain Container
 - ❑ CN=Dfs-Configuration,CN=System,DC=...

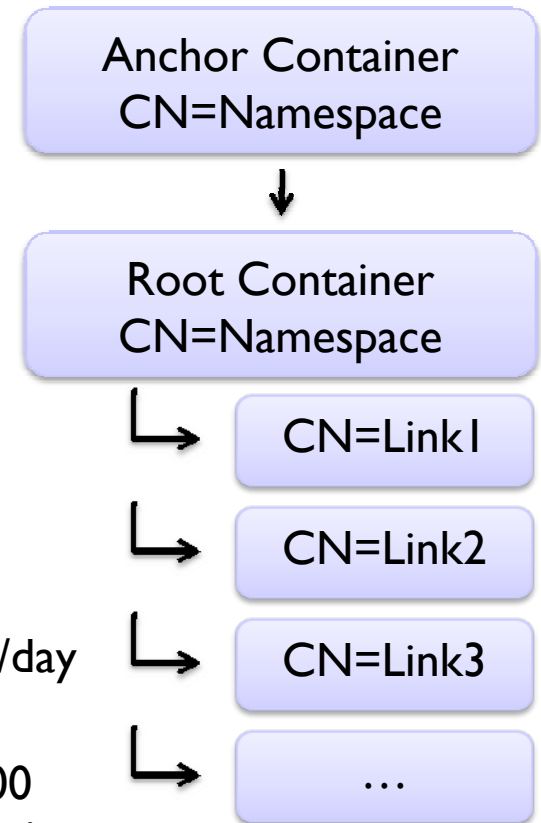
Domain 2000 Mode

- ❑ Simple design
 - ❑ Marshaled an image of what a standalone namespace would place in the registry
- ❑ Limited in scale
 - ❑ AD caps blob attributes at 10MB
 - ❑ ... results in about a 5MB/5000 link limit for these roots, for safety
 - ❑ Updates rewrite the blob, leading to $O(n^2)$ traffic on bulk operations
 - ❑ Compounded with multiple root targets
- ❑ Led to a tension in the system design
 - ❑ Standalone namespaces scale
 - ❑ Domain namespaces are distributable
 - ❑ Couldn't do both at the same time
- ❑ Deprecated with Windows 2008



Domain 2008 mode

- ❑ Namespace root container is split into
 - ❑ a generic anchor container
 - ❑ a version-specific container holding root targets and namespace properties
- ❑ Links broken down into objects
 - ❑ operations take constant time
 - ❑ updates are incremental and search-based – based on AD object USNs
 - ❑ metadata is visualized in the schema
- ❑ Windows' release infrastructure was updating 2,000 links/day with ~60,000 live around Windows 2008 RTM
 - ❑ at full localization, the Windows build produces ~1000 chunks, each of which translates to an individual share/link
 - ❑ DFSN is used to re-integrate the build into a single view
- ❑ 2008 functional level domain required



Internal Microsoft Examples

- ❑ [\\products\public](#)
 - ❑ clustered standalone root
 - ❑ hosts application installation images for the worldwide corporate network
 - ❑ \\products\public\products\applications redirects to 34 replicas
- ❑ [\\ntdev.corp.microsoft.com\release](#)
 - ❑ domain root
 - ❑ distribution point for all Windows builds
 - ❑ ~100 servers and low 100,000's of physical shares in interlinked namespaces
- ❑ Many previous-generation Microsoft IT hosted file servers have been consolidated using DFSN's consolidation namespaces

- ❑ <http://www.microsoft.com/dfs>
- ❑ <http://www.microsoft.com/protocols>
 - ❑ [MS-DFSC] – client access protocol
 - ❑ [MS-DFSNM] – server API and domain store
- ❑ Windows 2008: DFSN installs through the Server Manager, along with the admin console
- ❑ Vista SP1: Remote Server Administration Tools (RSAT)
- ❑ [mailto: DanLo \(at\) microsoft.com](mailto:DanLo@microsoft.com)
- ❑ Q & A