

Referrals, migration and replication for NFSv4

User Guide

version 0.2

RFC 3530 defines mechanisms to manage migration and replication of filesystems. The filesystem locations (fs_location attribute) supplies a way for the client to obtain. The fs_location attribute contains a list of locations, composed by an hostname and a path name representing the root of the filesystem on the server. The fs_location meaning depends on the type of service being provided. In case of migration, the list provides the location where a data has moved. In case of replication, the list gives the places where the replicated filesystem(s) is (are).

This document aims at explaining shortly migration and replication, and to give information about configuration to use them.

1. Referrals:

Referral Situations

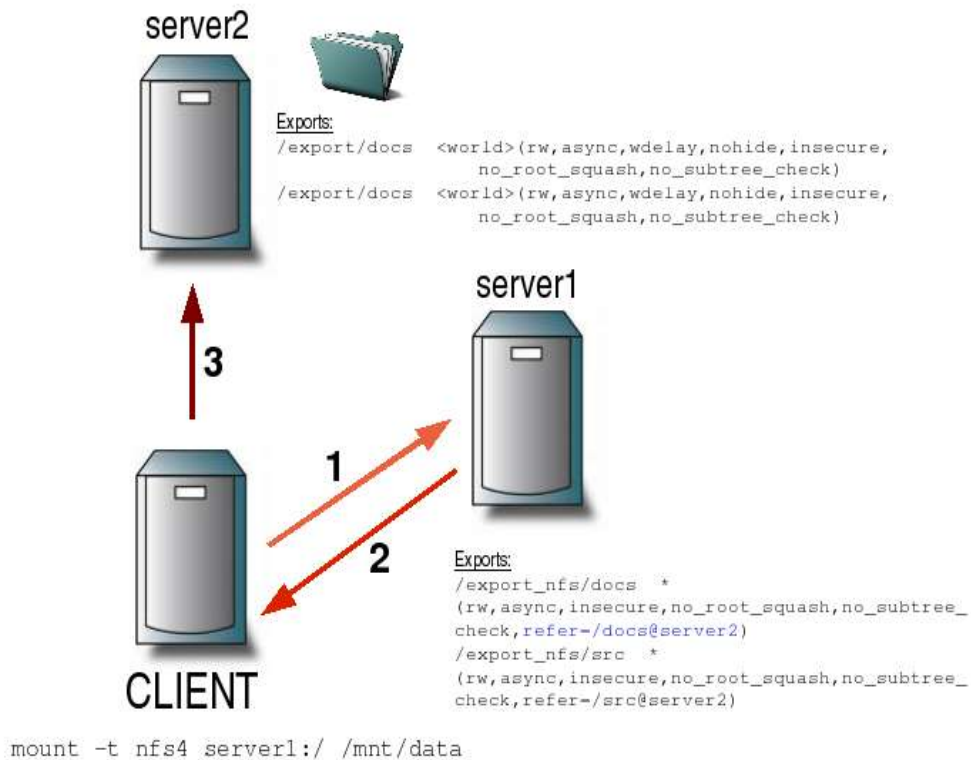
When a client is directed to a new location when first referencing a file system, the result is best described, from the client's point of view as a referral, rather than a migration event, since the client does not have any information on the file system before the migration occurred.

If, at the time of migration, none of the clients has referenced the filesystem yet, we have a pure referral situation: all that clients will ever see is the referral for an absent file system.

Given that clients can use such referrals to find the current location of file systems, servers can usefully provide such referrals even if the filesystem never resided on the server.

Referrals make possible a multi-server namespace to be build so that clients can address file systems in terms of the name of the filesystem on a server providing referrals. In that case server configuration changes which move file systems from server to server will not impact the client.

It is a good way to balance between several servers, rather than having one server hosting all exported filesystems. We can easily imagine an architecture where a server could only provide referrals to clients and dispatch request to servers that contain data.



Configuration:

The server must indicate the location of migrated filesystem(s) with the option:
refer=<directory>@<host>

in /etc/exports file.

Example:

```
/export/docs <world>
(rw, async, wdelay, insecure, no_subtree_check, refer=/docs@dataserver1)
/export/sources <world>
(rw, async, wdelay, insecure, no_subtree_check, refer=/sources@dataserver2)
```

The server indicates that the export/docs and export/sources have been migrated to 2 servers.

NB: Only the server needs to be configured. Neither the client nor the server where filesystem is migrated need special configuration. The client just needs to be able to manage fs_location. That is the case in 2.6.18-rc5 and above versions of the kernel. No patch is needed for the client userland part.

2. Migration:

Filesystem migration is used to move a filesystem from one server to another. Migration is typically used for a filesystem that is writable and has a single copy. It is a good way to move data exported via NFS, without disturbing client configuration, because the client is informed of each change in data locations.

To put it simply:

Clients have mounted a filesystem on server via NFS version 4. The server administrator can decide to move this exported filesystem to another server. The server must inform each client mounting filesystem on it, that the data has moved.

The method used to communicate the migration event between client and server is the following: once the servers participating in the migration have completed the move of the filesystem, the error NFS4ERR_MOVED will be returned for subsequent requests received by the original server. Upon receiving the NFS4ERR_MOVED error, the client obtains the value of the fs_location attribute. The client then uses the contents of the attribute to redirect its requests to the specified server.

Configuration:

On the server, the refer option has to be used, to indicate the location of the migrated filesystem. Moreover, on the server, you need to do something similar to:

```
mount --bind /dummy_mount /export_nfs
```

(This assumes /export is the nfs-root, and is not currently a mountpoint.)

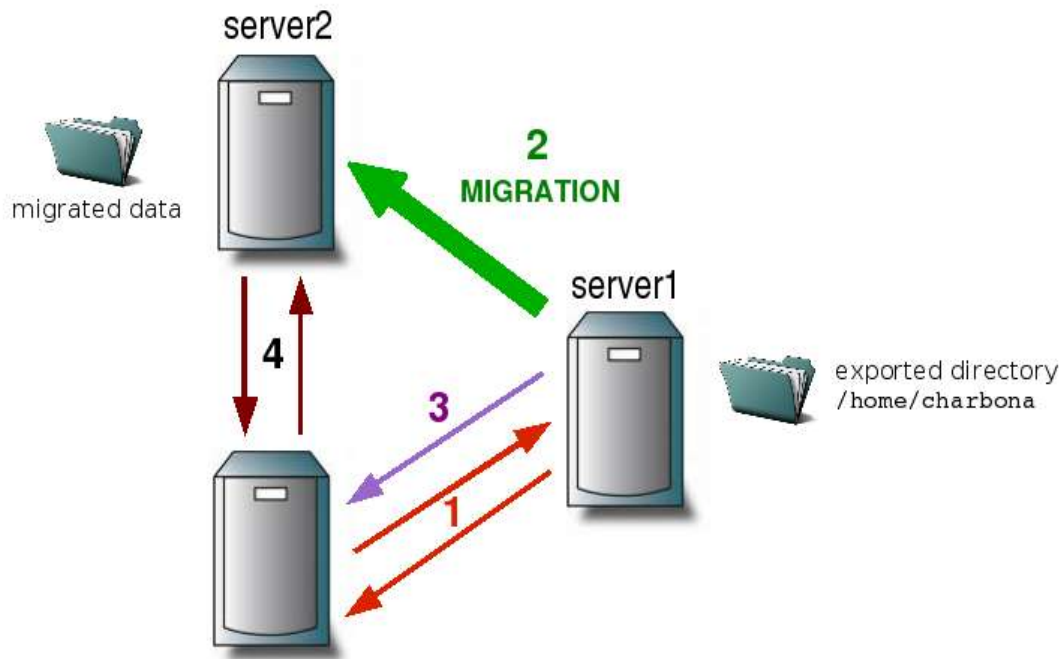
On the client, no configuration is needed, the operation must be transparent from a client point of view.

The migration server must be configured like a single server (single filesystem exports)

Example:

In file /etc/exports:

```
/home/charbona <world>  
(rw,async,wdelay,insecure,no_subtree_check,refer=/charbona@server2)
```



Typical sequence of operation:

- 1: the client mounts a filesystem exported by server1
- 2: filesystem is migrated from server1 to server2
- 3: when attempting to access to the mounted filesystem, client is sent a NFS4ERR_MOVED error associated to a fs_location attribute indicating the new location of the filesystem.
- 4: the client mounts the filesystem on server2.

NB: The way of migrating data is not defined by NFSv4.

Referrals and migration:

Referrals can be considered as a sub-case of migration.

If a server implements file system migration, individual clients, not being synchronized with the migration event, will encounter different situations (from a client point of view).

Some of them will have already received filehandles within the migrated filesystem, others may at that time have already accessed the filesystem (when it was present) and will need to be redirected.

Thus referrals are a limiting sub-case of migration. When a migration event occurs, some clients will see an ordinary migration in the case in which access (by that client) to the filesystem being moved has already occurred, while others will see a referral when they first attempt to access the absent filesystem.

3. Replication:

In case of replication, the `fs_location` attribute is used to supply to the client a list of places where the filesystem is replicated, and can be reached in case of problem (network problem, nfs daemon stopped etc.)

Filesystem replication is aimed to be used in case of read-only data. Typically, the filesystem will be replicated on two or more servers. The `fs_location` attribute will provide the list of these locations to the client. On first access of the filesystem, the client should obtain the value of the `fs_location` attribute. If, in the future, a client request times out, the client may attempt to use another of the servers specified in `fs_location` attribute.

In this case exported data are replicated on other server(s), called replicas.

NB: the way of replicating a filesystem is not managed by NFSv4.

We can imagine to cases of replication:

- the data can be duplicated manually (copy of the filesystem from a server to another). In this case, data should be available in read-only mode.
- the replicated data could be managed by an external tool, that manage replica update, data coherence between replicas etc.

Configuration:

Client configuration: nothing to be configured. Just a linux-2.6.18-rc5 kernel or above is necessary to support `fs_location` function.

Server configuration: the export list must contain the list of replica location of each replicated filesystem with the option `replica=<dir>@<host>`

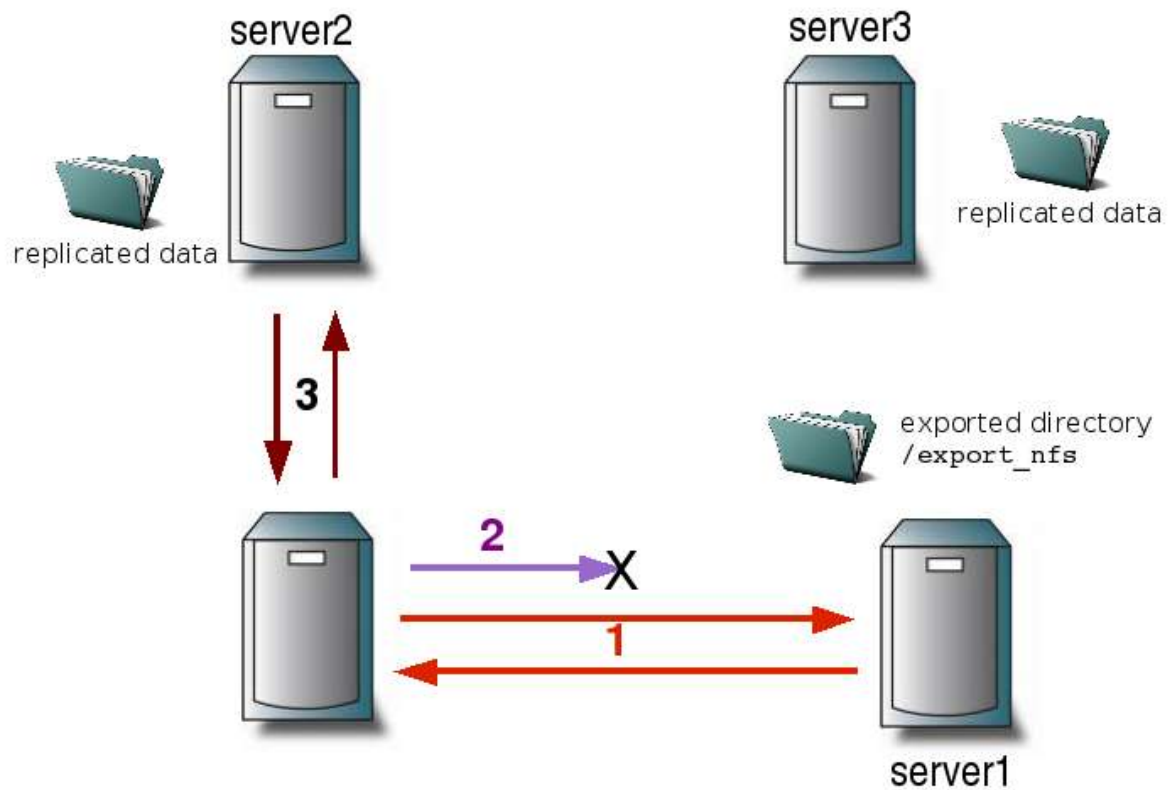
Replica configuration: the replica is a single NFS-server, that can be addressed like a replica or a single server.

Example:

In file `/etc/exports`:

```
/export_nfs <world>(rw,[...],  
insecure,no_root_squash,replicas=@replica1:/data@replica2)
```

Typical sequence of operation:



- 1- The client mount a filesystem on the client, and obtains the replica location(s)
- 2- The server is no more reachable by the client
- 3- The client mount a replicated filesystem.

4. Software needed (to be confirmed)

Kernel: linux-2.6.18-rc5

NFS userland package: nfs-utils-1.0.10