NFS v3/v4 Active Passive Deployment Overview

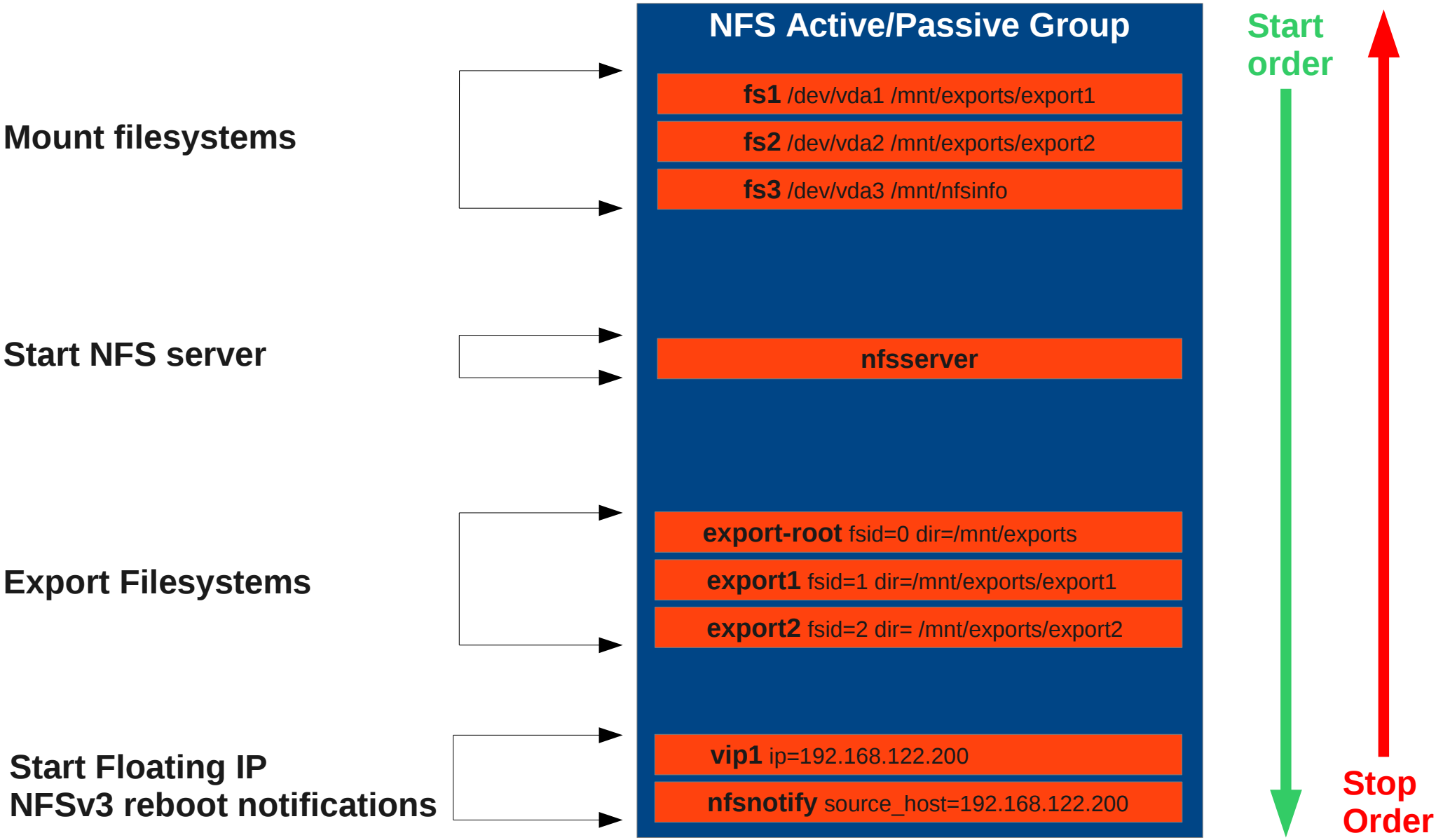
Author: David Vossel <dvossel@redhat.com>
Version: 2

An automated deployment script outlining the specifics of how to deploy HA NFS active-passive with Pacemaker can be found at the link below.

https://github.com/davidvossel/phd/blob/master/scenarios/nfs-active-passive.scenario

NFS v3/v4 Active Passive

Pacemaker Resource Group Example



Mount Filesystems

The nfs resource stack consists of shared filesystems mounted with the *Filesystem* agent.

These filesystems are used by both the nfsserver and the exports later on in the stack.

NOTE! All the filesystems *MUST* be ordered to start before the NFS daemons and stop after the NFS daemons. If exported filesystems are mounted after the NFS daemons, the filesystems will block during the umount if NFSv4 file leases are active. Ordering here is very important.

Start NFS Server Daemons

Next the NFS daemons (managed by the *nfsserver* agent) are started.

Note! One of the filesystems is used by the *nfsserver* agent to bind to the /var/lib/nfs directory. This allows the NFS server to maintain client data on shared storage for recovery failover.

Export Filesystems

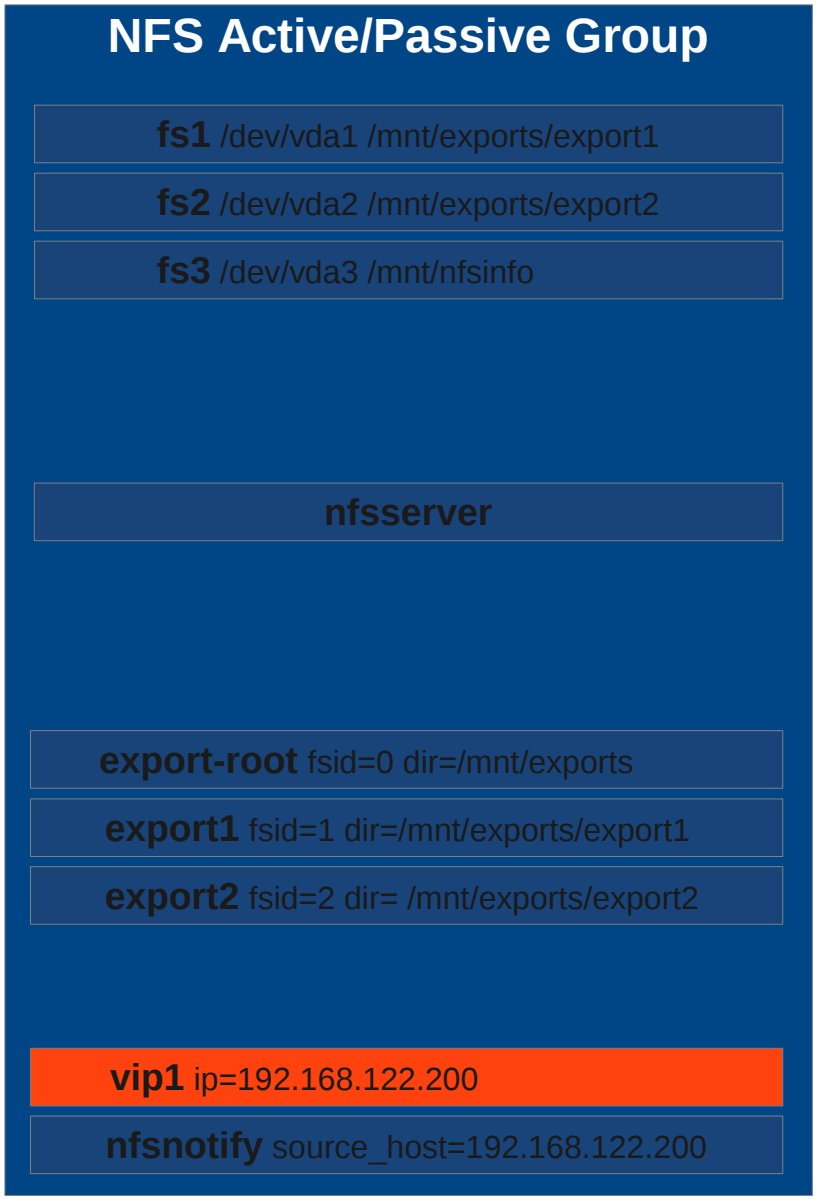
Next the shared filesystems are exported using the *exportfs* agent.

Note the ordering here. All the exports *MUST* start after the NFS daemons start.

Floating IP

The server's floating IP must be started after all the Filesystem exports.

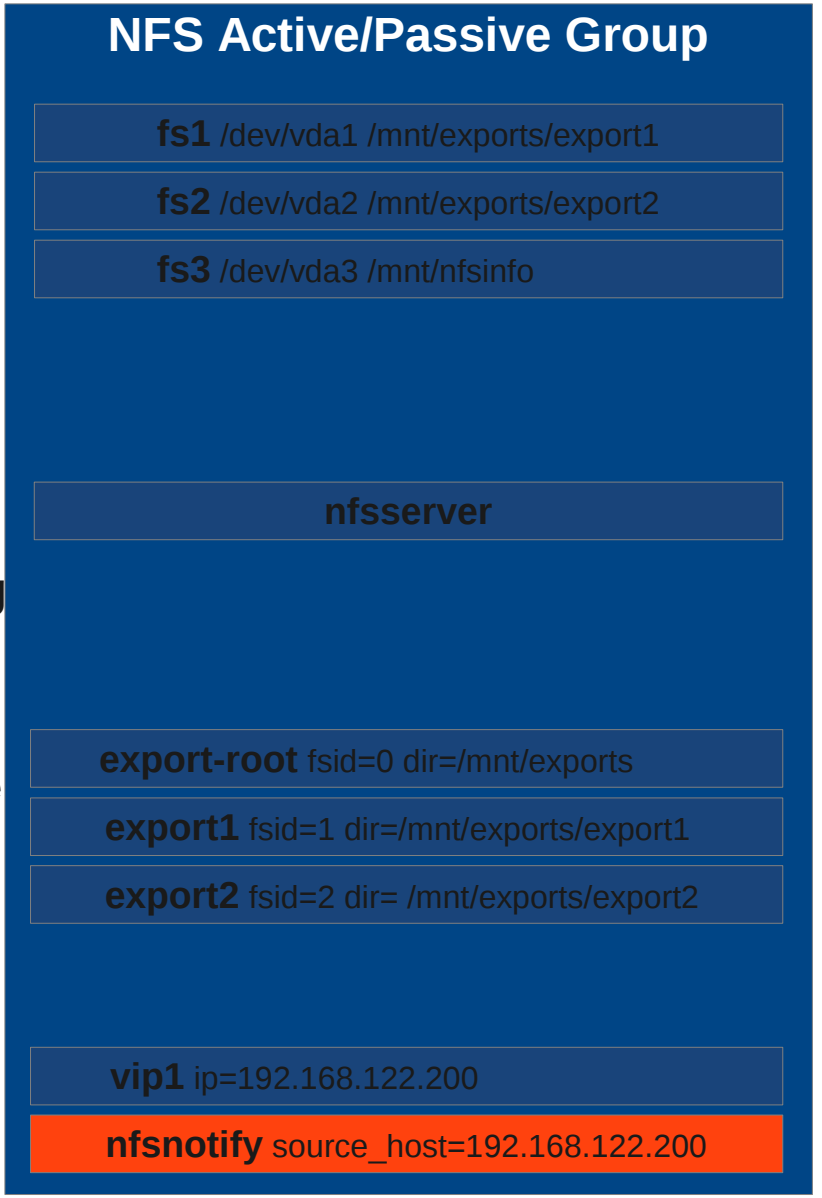
We do not want it to be possible for the clients to contact the server after a failover until all the exports are up again.



NFSv3 Lock recovery

Once the floating IP is initialized the NFSv3 reboot notifications are sent to inform previous clients to reclaim their locks during the server's grace period.

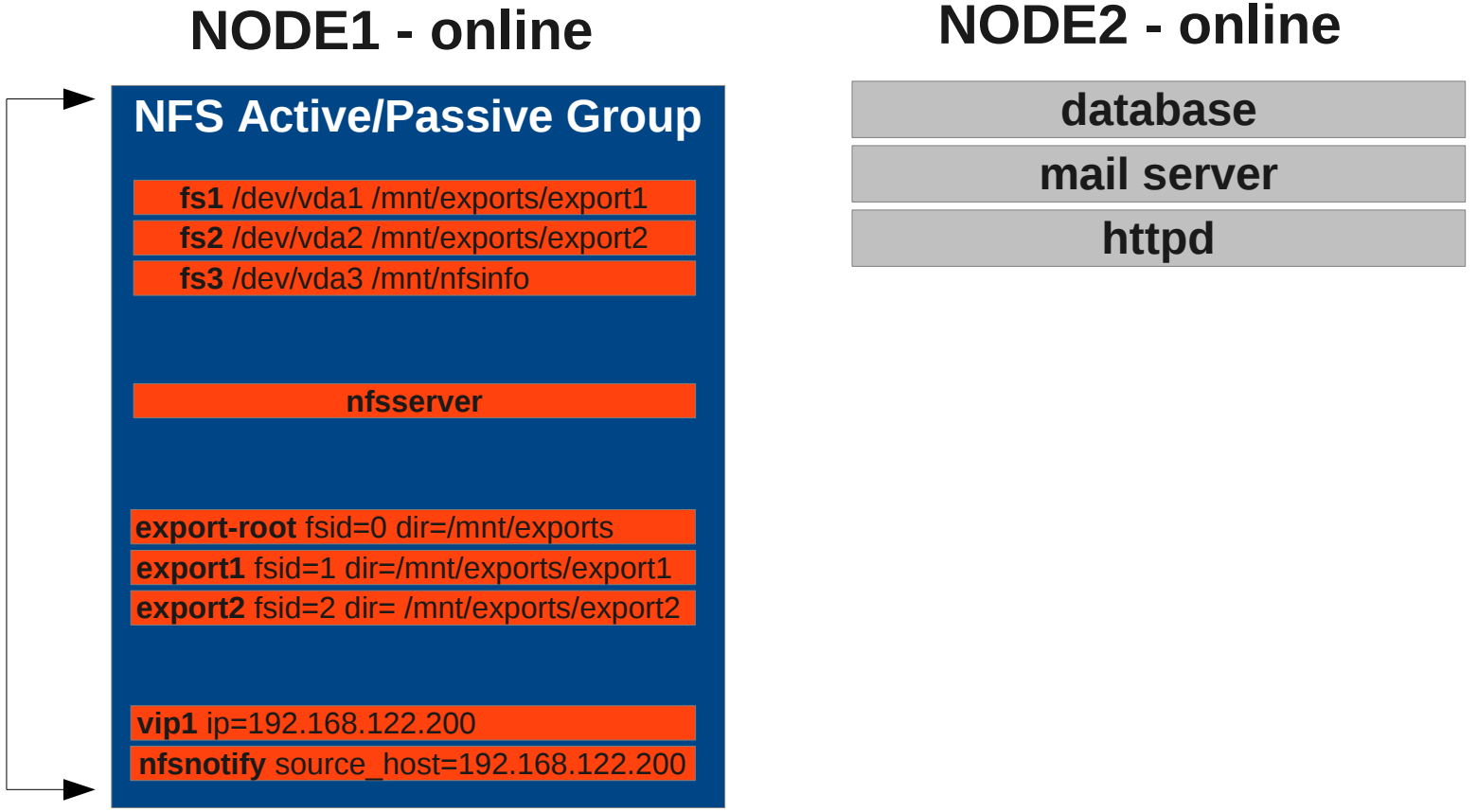
NOTE! The nfsnotify agent must be provided the float IP. This allows the nfsnotify agent to properly set the source name in the notify requests. Otherwise the clients will ignore the reboot notifications because they'll look like they're coming from an unknown server.



Node Failure

This group moves as a single unit during failover.

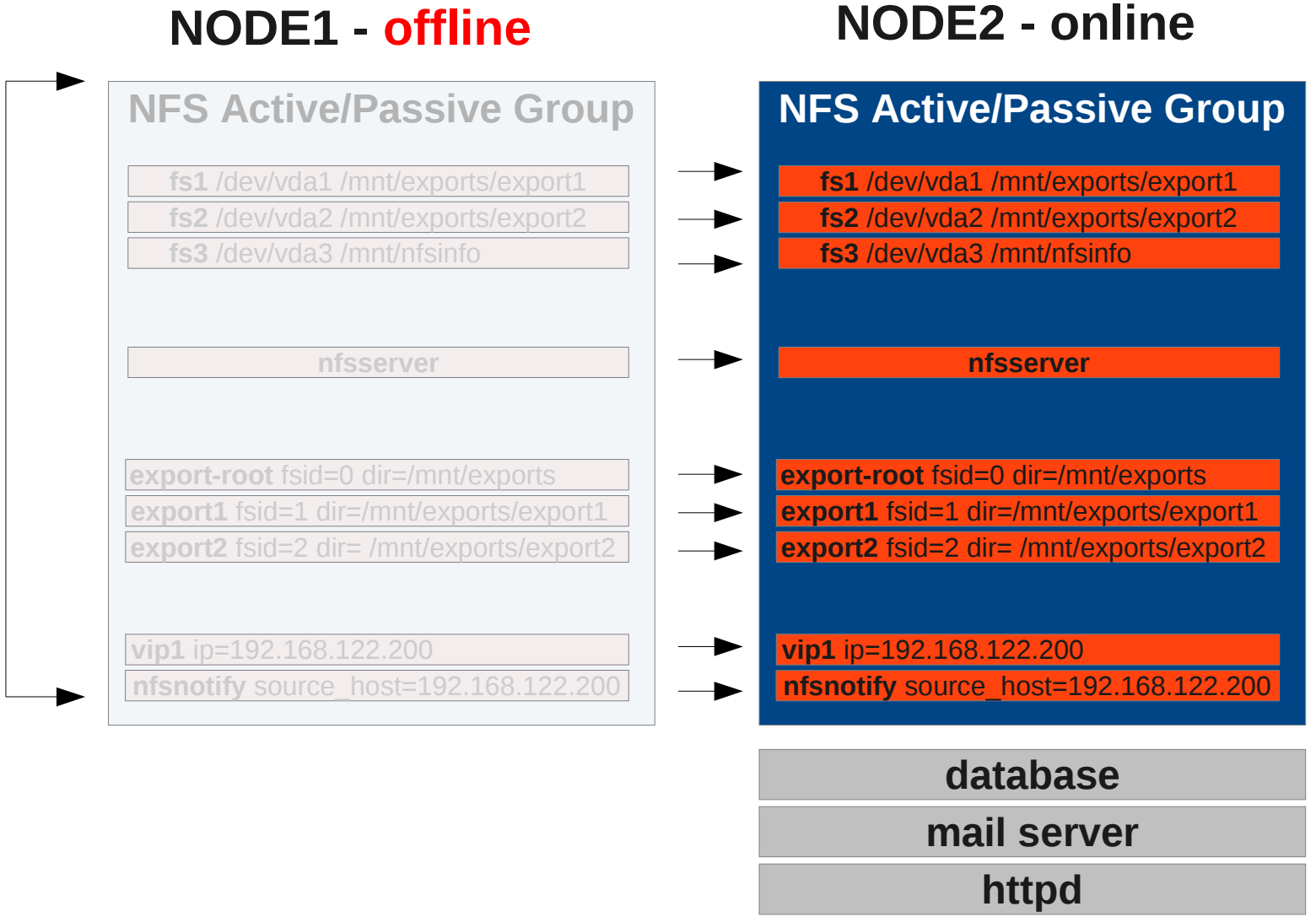
This means the filesystems, exports, and floating IP are bound to a single nfsserver resource.



Node Failure

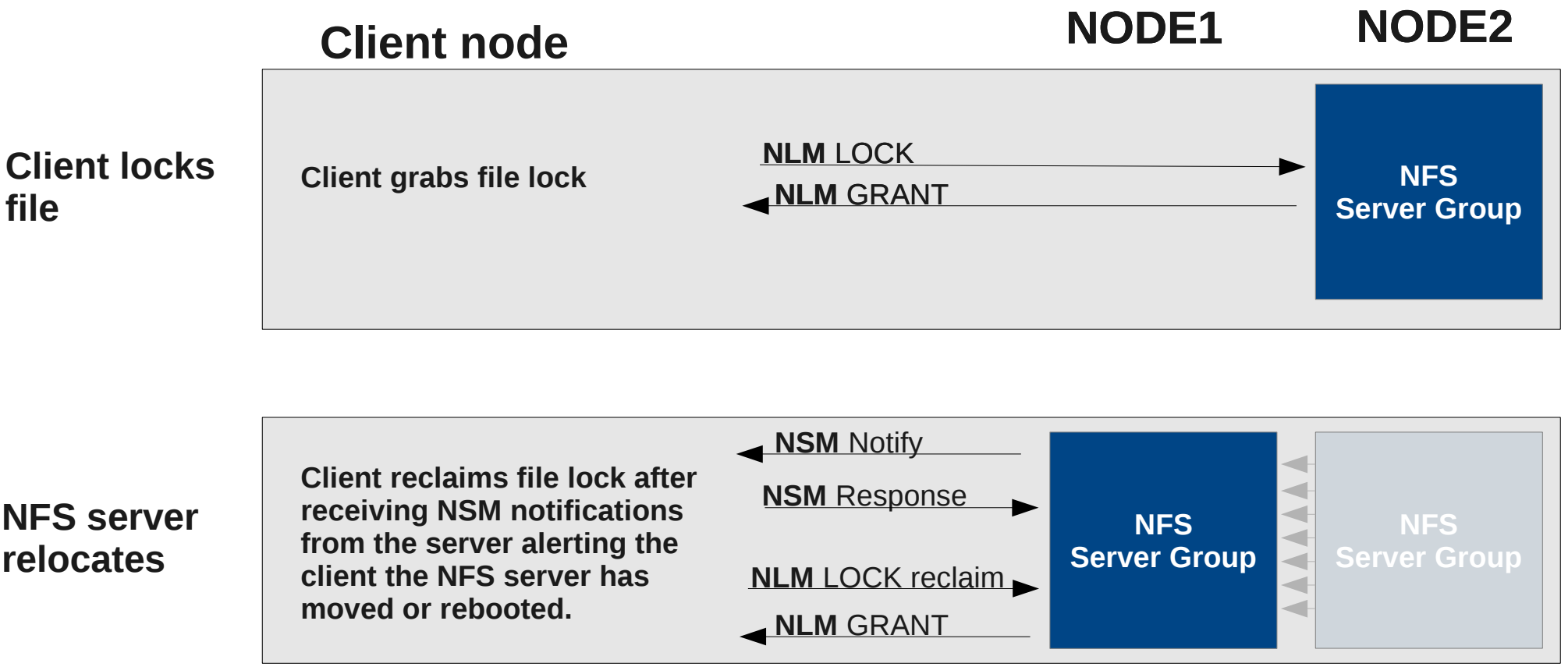
This group moves as a single unit during failover.

This means the filesystems, exports, and floating IP are bound to a single nfsserver resource.



Debug NFSv3 Lock Recovery

Overview



Debug NFSv3 Lock Recovery

Mounting NFS and grabbing locks

Mount the HA NFS share and verify contents are accessible

 # mount -v -o "vers=3" nfshost:/nfsexport /nfsshare
 # ls /nfsshare
clientsharefile

Lock a file found in the nfs mount using *flock* and hold the lock for the duration of the 'sleep' command

 # flock /nfsshare/clientsharefile.txt -c “sleep 10000”

Debug NFSv3 Lock Recovery

Viewing NLM and NSM traffic

View NSM and NLM traffic on the NFS client node using wireshark.

NSM = traffic associated with nfsnotify. This traffic originates from the NFS server and notifies the client that the server has moved or restarted. Clients must reclaim locks once this notify request is received.

 # tshark -V -i eth0 -R stat

NLM = traffic associated with client lock and unlock requests. These requests originate from the client, the server will in turn either GRANT or DENY these requests to lock/unlock.

 # tshark -V -i eth0 -R nlm

Debug NFSv3 Lock Recovery

Verifying lock reclaim on server

After the client issues the lock reclaim request, a new lock entry should be present in the /proc/locks file on the node the nfsserver is running on.

Debug NFSv3 Lock Recovery

Understanding Failures

NSM requests are processed but client does not reclaim locks.

- Did the NFS daemons actually restart? Clients will not reclaim locks unless the state number in the NSM notify request is different. The nfsserver agent will refresh the server's state number (located in /var/lib/nfs/statd/state) during the start operation.

- Verify the floating IP has a FQDN associated with it. Make sure clients are using the FQDN when mounting the NFS share. The NSM protocol is very sensitive to consistent hostname matching.

Server does not respond to client's lock reclaim request or responds with 'Stale File Handler'

- Verify the nfsserver has the 'nfs_no_notify=true' option set to avoid accidentally sending out NSM notifications before the exports and IP start.
- Verify the nfsnotify agent starts after all the exports and floating IP.