NFS

Enterprise applications

NetApp
January 12, 2026

Sommaire

NFS . . . 1
- Présentation . . . 1
  - Versions NFS . . . 1
  - Tables d’emplacements TCP Linux NFSv3 . . . 1
  - ADR et NFS . . . 2
  - nfs-rootonly et mount-rootonly . . . 2
  - Règles d’exportation NFS : superutilisateur et setuid . . . 3
  - Configuration NFSv4/4.1 . . . 3
- Oracle Direct NFS (dNFS) . . . 5
  - NFS direct . . . 5
  - Accès direct au NFS et au système de fichiers hôte . . . 8
- Locations et verrouillages NFS . . . 9
  - État NFSv4 . . . 9
  - Verrous NFSv4 . . . 9
  - Locations NFSv4 . . . 9
  - Périodes de grâce NFSv4 . . . 11
  - Délais de location par rapport aux délais de grâce . . . 12
- Mise en cache NFS . . . 13
- Tailles de transfert NFS . . . 13

NFS

Présentation

NetApp fournit un stockage NFS haute performance depuis plus de 30 ans et son utilisation se développe avec les infrastructures basées sur le cloud en raison de sa simplicité.

Le protocole NFS comprend plusieurs versions aux exigences variables. Pour une description complète de la configuration NFS avec ONTAP, reportez-vous à la section "Tr-4067 NFS sur les meilleures pratiques ONTAP". Les sections suivantes couvrent certaines des exigences les plus critiques et des erreurs utilisateur courantes.

Versions NFS

Le client NFS du système d'exploitation doit être pris en charge par NetApp.

- NFSv3 est pris en charge avec des systèmes d'exploitation conformes à la norme NFSv3.
- NFSv3 est pris en charge avec le client Oracle dNFS.
- NFSv4 est pris en charge avec tous les systèmes d'exploitation conformes à la norme NFSv4.
- NFSv4.1 et NFSv4.2 nécessitent une prise en charge spécifique du système d'exploitation. Consulter le "NetApp IMT" Pour les systèmes d'exploitation pris en charge.
- La prise en charge d'Oracle dNFS pour NFSv4.1 requiert Oracle 12.2.0.2 ou version supérieure.

> Le "Matrice de prise en charge de NetApp" Pour NFSv3 et NFSv4 n'incluent pas de systèmes d'exploitation spécifiques. Tous les systèmes d'exploitation conformes à la RFC sont généralement pris en charge. Lors d'une recherche dans la prise en charge en ligne de IMT pour NFSv3 ou NFSv4, ne sélectionnez pas de système d'exploitation spécifique, car aucune correspondance ne sera affichée. Tous les systèmes d'exploitation sont implicitement pris en charge par la politique générale.

Tables d'emplacements TCP Linux NFSv3

Les tables d'emplacements TCP sont l'équivalent NFSv3 de la profondeur de file d'attente de l'adaptateur de bus hôte (HBA). Ces tableaux contrôlent le nombre d'opérations NFS qui peuvent être en attente à la fois. La valeur par défaut est généralement 16, un chiffre bien trop faible pour assurer des performances optimales. Le problème inverse se produit sur les noyaux Linux plus récents : la limite de la table des emplacements TCP augmente automatiquement par envoi de demandes, jusqu'à atteindre le niveau de saturation du serveur NFS.

Pour des performances optimales et pour éviter les problèmes de performances, ajustez les paramètres du noyau qui contrôlent les tables d'emplacements TCP.

Exécutez le `sysctl -a | grep tcp.*.slot_table` et observez les paramètres suivants :

```
# sysctl -a | grep tcp.*.slot_table
sunrpc.tcp_max_slot_table_entries = 128
sunrpc.tcp_slot_table_entries = 128
```

Tous les systèmes Linux doivent inclure `sunrpc.tcp_slot_table_entries`, mais seulement certains incluent `sunrpc.tcp_max_slot_table_entries`. Ils doivent tous deux être réglés sur 128.

Si vous ne définissez pas ces paramètres, vous risquez d'avoir des effets importants sur les performances. Dans certains cas, les performances sont limitées car le système d'exploitation linux n'émet pas suffisamment d'E/S. Dans d'autres cas, les latences d'E/S augmentent à mesure que le système d'exploitation linux tente d'émettre plus d'E/S que ce qui peut être traité.

ADR et NFS

Certains clients ont signalé des problèmes de performances liés à une quantité excessive d'E/S dans le `ADR` emplacement. Le problème ne se produit généralement pas tant qu'une grande quantité de données de performances ne s'est pas accumulée. La raison de cet excès d'E/S est inconnue, mais ce problème semble provenir des analyses répétées du répertoire cible par les processus Oracle pour détecter les modifications.

Dépose du `noac` et/ou `actimeo=0` Les options de montage permettent la mise en cache du système d'exploitation hôte et réduisent les niveaux d'E/S du stockage.

NetApp recommande de ne pas placer `ADR` données sur un système de fichiers avec `noac` ou `actimeo=0` parce que des problèmes de performances sont probables. Séparer `ADR` le cas échéant, les données vers un autre point de montage.

nfs-rootonly et mount-rootonly

ONTAP inclut une option NFS appelée `nfs-rootonly` Cela permet de contrôler si le serveur accepte les connexions de trafic NFS à partir des ports élevés. Par mesure de sécurité, seul l'utilisateur root est autorisé à ouvrir des connexions TCP/IP à l'aide d'un port source inférieur à 1024 car ces ports sont normalement réservés à l'utilisation du système d'exploitation, et non aux processus utilisateur. Cette restriction permet de s'assurer que le trafic NFS provient d'un client NFS du système d'exploitation et non d'un processus malveillant émulant un client NFS. Le client Oracle dNFS est un pilote d'espace utilisateur, mais le processus s'exécute en tant que root, il n'est donc généralement pas nécessaire de modifier la valeur de `nfs-rootonly`. Les connexions sont réalisées à partir de ports bas.

Le `mount-rootonly` Cette option s'applique uniquement à NFSv3. Il contrôle si l'appel de MONTAGE RPC est accepté à partir de ports supérieurs à 1024. Lorsque dNFS est utilisé, le client est de nouveau exécuté en tant que root, ce qui lui permet d'ouvrir des ports inférieurs à 1024. Ce paramètre n'a aucun effet.

Les processus ouvrant des connexions avec dNFS sur les versions 4.0 et supérieures de NFS ne s'exécutent pas en tant que root et nécessitent donc des ports supérieurs à 1024. Le `nfs-rootonly` Le paramètre doit être défini sur Désactivé pour dNFS pour terminer la connexion.

Si `nfs-rootonly` Est activé, le résultat est un blocage lors de la phase de montage ouvrant les connexions dNFS. La sortie sqlplus ressemble à ceci :

```
SQL>startup
ORACLE instance started.
Total System Global Area 4294963272 bytes
Fixed Size                  8904776 bytes
Variable Size             822083584 bytes
Database Buffers         3456106496 bytes
Redo Buffers                7868416 bytes
```

Le paramètre peut être modifié comme suit :

```
Cluster01::> nfs server modify -nfs-rootonly disabled
```

Dans de rares cas, vous devrez peut-être modifier nfs-rootonly et mount-rootonly sur Désactivé. Si un serveur gère un très grand nombre de connexions TCP, il est possible qu'aucun port inférieur à 1024 n'est disponible et que le système d'exploitation soit forcé d'utiliser des ports supérieurs. Ces deux paramètres ONTAP doivent être modifiés pour permettre la connexion.

Règles d'exportation NFS : superutilisateur et setuid

Si les binaires Oracle se trouvent sur un partage NFS, les règles d'export doivent inclure des autorisations de superutilisateur et de setuid.

Les exportations NFS partagées utilisées pour les services de fichiers génériques tels que les répertoires personnels des utilisateurs écraseront généralement l'utilisateur root. Cela signifie qu'une demande de l'utilisateur root sur un hôte qui a monté un système de fichiers est remappée en tant qu'utilisateur différent avec des privilèges inférieurs. Cela permet de sécuriser les données en empêchant un utilisateur root d'un serveur donné d'accéder aux données du serveur partagé. Le bit setuid peut également représenter un risque de sécurité dans un environnement partagé. Le bit setuid permet d'exécuter un processus en tant qu'utilisateur différent de celui qui appelle la commande. Par exemple, un script shell qui était détenu par root avec le bit setuid s'exécute en tant que root. Si ce script shell peut être modifié par d'autres utilisateurs, tout utilisateur non root peut émettre une commande en tant que root en mettant à jour le script.

Les binaires Oracle incluent les fichiers appartenant à root et utilisent le bit setuid. Si des binaires Oracle sont installés sur un partage NFS, les règles d'export doivent inclure les autorisations de superutilisateur et de setuid appropriées. Dans l'exemple ci-dessous, la règle inclut les deux `allow-suid` et permis `superuser` Accès (root) pour les clients NFS via l'authentification système.

```
Cluster01::> export-policy rule show -vserver vserver1 -policyname orabin
-fields allow-suid,superuser
vserver   policyname ruleindex superuser allow-suid
--------- ---------- --------- --------- ----------
vserver1  orabin     1         sys       true
```

Configuration NFSv4/4.1

Pour la plupart des applications, il y a très peu de différence entre NFSv3 et NFSv4. Les E/S applicatives sont

généralement des E/S très simples et ne bénéficient pas énormément de certaines des fonctionnalités avancées de NFSv4. Les versions supérieures de NFS ne doivent pas être considérées comme une « mise à niveau » du point de vue du stockage de la base de données, mais plutôt comme des versions de NFS qui incluent des fonctionnalités supplémentaires. Par exemple, si la sécurité de bout en bout du mode de confidentialité kerberos (krb5p) est requise, NFSv4 est requis.

NetApp recommande d'utiliser NFSv4.1 si les fonctionnalités NFSv4 sont requises. Certaines améliorations fonctionnelles du protocole NFSv4 dans NFSv4.1 améliorent la résilience dans certains cas à la périphérie.

Le passage à NFSv4 est plus compliqué que de simplement changer les options de montage de vers=3 en vers=4.1. Pour une explication plus complète de la configuration de NFSv4 avec ONTAP, notamment des conseils sur la configuration du système d'exploitation, voir "Tr-4067 NFS sur les meilleures pratiques ONTAP". Les sections suivantes de ce TR expliquent certaines des exigences de base relatives à l'utilisation de NFSv4.

Domaine NFSv4

Une explication complète de la configuration NFSv4/4.1 dépasse le cadre de ce document, mais un problème couramment rencontré est une incohérence dans le mappage de domaine. Du point de vue de sysadmin, les systèmes de fichiers NFS semblent se comporter normalement, mais les applications signalent des erreurs concernant les autorisations et/ou le setuid sur certains fichiers. Dans certains cas, les administrateurs ont conclu à tort que les autorisations des binaires de l'application ont été endommagées et ont exécuté des commandes chown ou chmod lorsque le problème réel était le nom de domaine.

Le nom de domaine NFSv4 est défini sur le SVM ONTAP :

```
Cluster01::> nfs server show -fields v4-id-domain
vserver   v4-id-domain
--------- ------------
vserver1  my.lab
```

Le nom de domaine NFSv4 sur l'hôte est défini dans `/etc/idmap.cfg`

```
[root@host1 etc]# head /etc/idmapd.conf
[General]
#Verbosity = 0
# The following should be set to the local NFSv4 domain name
# The default is the host's DNS domain name.
Domain = my.lab
```

Les noms de domaine doivent correspondre. Si ce n'est pas le cas, des erreurs de mappage similaires à ce qui suit apparaissent dans `/var/log/messages`:

```
Apr 12 11:43:08 host1 nfsidmap[16298]: nss_getpwnam: name 'root@my.lab'
does not map into domain 'default.com'
```

Les binaires d'application, tels que les binaires de base de données Oracle, incluent les fichiers appartenant à

root avec le bit setuid, ce qui signifie qu'une discordance dans les noms de domaine NFSv4 provoque des échecs avec le démarrage d'Oracle et un avertissement sur la propriété ou les autorisations d'un fichier appelé `oradism`, qui est situé dans le `$ORACLE_HOME/bin` répertoire. Elle doit apparaître comme suit :

```
[root@host1 etc]# ls -l /orabin/product/19.3.0.0/dbhome_1/bin/oradism
-rwsr-x--- 1 root oinstall 147848 Apr 17  2019
/orabin/product/19.3.0.0/dbhome_1/bin/oradism
```

Si ce fichier apparaît avec la propriété de personne, il peut y avoir un problème de mappage de domaine NFSv4.

```
[root@host1 bin]# ls -l oradism
-rwsr-x--- 1 nobody oinstall 147848 Apr 17  2019 oradism
```

Pour résoudre ce problème, vérifiez le `/etc/idmap.cfg` Comparez le paramètre v4-ID-domain sur ONTAP et assurez-vous qu'ils sont cohérents. Si ce n'est pas le cas, effectuez les modifications requises, exécutez `nfsidmap -c`, et attendez un moment pour que les modifications se propagent. La propriété du fichier doit alors être correctement reconnue en tant que racine. Si un utilisateur a tenté de s'exécuter `chown root` Sur ce fichier avant que la configuration des domaines NFS ne soit corrigée, il peut être nécessaire de l'exécuter `chown root` encore.

Oracle Direct NFS (dNFS)

Les bases de données Oracle peuvent utiliser NFS de deux manières.

Tout d'abord, il peut utiliser un système de fichiers monté à l'aide du client NFS natif qui fait partie du système d'exploitation. Il s'agit parfois de kernel NFS ou KNFS. Le système de fichiers NFS est monté et utilisé par la base de données Oracle exactement comme toute autre application utiliserait un système de fichiers NFS.

La deuxième méthode est Oracle Direct NFS (dNFS). Il s'agit d'une implémentation de la norme NFS dans le logiciel de base de données Oracle. Elle ne modifie pas la façon dont les bases de données Oracle sont configurées ou gérées par l'administrateur de base de données. Tant que les paramètres du système de stockage lui-même sont corrects, l'utilisation de dNFS doit être transparente pour l'équipe DBA et les utilisateurs finaux.

Les systèmes de fichiers NFS habituels sont toujours montés sur une base de données avec la fonction dNFS activée. Une fois la base de données ouverte, la base de données Oracle ouvre un ensemble de sessions TCP/IP et effectue directement des opérations NFS.

NFS direct

La valeur principale de Direct NFS d'Oracle est de contourner le client NFS hôte et d'effectuer des opérations de fichiers NFS directement sur un serveur NFS. Pour l'activer, il suffit de modifier la bibliothèque Oracle Disk Manager (ODM). Vous trouverez des instructions sur ce processus dans la documentation Oracle.

L'utilisation de dNFS entraîne une amélioration significative des performances d'E/S et réduit la charge sur l'hôte et le système de stockage, car les E/S sont effectuées de la manière la plus efficace possible.

En outre, Oracle dNFS inclut une **option** pour les chemins d'accès multiples et la tolérance aux pannes de

l'interface réseau. Par exemple, il est possible de lier deux interfaces de 10 Gbits pour offrir 20 Go de bande passante. En cas de défaillance d'une interface, les E/S sont relancées sur l'autre interface. L'opération globale est très similaire aux chemins d'accès multiples FC. Les chemins d'accès multiples étaient courants il y a plusieurs années, alors que l'ethernet 1 Gbit était la norme la plus courante. Une carte réseau 10 Go suffit pour la plupart des charges de travail Oracle, mais si un nombre supérieur de cartes réseau 10 Go sont requises, elles peuvent être reliées.

Lorsque dNFS est utilisé, il est essentiel que tous les correctifs décrits dans Oracle Doc 1495104.1 soient installés. Si un correctif ne peut pas être installé, l'environnement doit être évalué pour s'assurer que les bugs décrits dans ce document ne causent pas de problèmes. Dans certains cas, une incapacité à installer les correctifs requis empêche l'utilisation de dNFS.

N'utilisez pas dNFS avec tout type de résolution de noms round-Robin, y compris DNS, DDNS, NIS ou toute autre méthode. Cela inclut la fonction d'équilibrage de la charge DNS disponible dans ONTAP. Lorsqu'une base de données Oracle utilisant dNFS résout un nom d'hôte en adresse IP, elle ne doit pas être modifiée lors des recherches ultérieures. Cela peut entraîner des pannes de la base de données Oracle et une corruption potentielle des données.

Activation de dNFS

Oracle dNFS peut fonctionner avec NFSv3 sans aucune configuration nécessaire au-delà de l'activation de la bibliothèque dNFS (voir la documentation Oracle pour la commande spécifique requise). Toutefois, si dNFS ne parvient pas à établir la connectivité, il peut revenir en arrière silencieux au client NFS du noyau. Dans ce cas, les performances peuvent être gravement affectées.

Si vous souhaitez utiliser le multiplexage dNFS sur plusieurs interfaces, avec NFSv4.X, ou utiliser le chiffrement, vous devez configurer un fichier oranfstab. La syntaxe est extrêmement stricte. De petites erreurs dans le fichier peuvent entraîner l'affichage du démarrage ou le contournement du fichier orangfstab.

Au moment de la rédaction de ce rapport, les chemins d'accès multiples dNFS ne fonctionnent pas avec NFSv4.1 avec les versions récentes d'Oracle Database. Un fichier oranfstab qui spécifie NFSv4.1 comme protocole ne peut utiliser qu'une instruction de chemin unique pour une exportation donnée. La raison en est que ONTAP ne prend pas en charge l'agrégation ClientID. Les correctifs de bases de données Oracle permettant de résoudre cette limitation seront peut-être disponibles à l'avenir.

La seule façon d'être certain que dNFS fonctionne comme prévu est d'interroger les tables v$dnfs.

Vous trouverez ci-dessous un exemple de fichier oranfstab situé dans /etc Il s'agit de l'un des emplacements multiples où un fichier oranfstab peut être placé.

```
[root@jfs11 trace]# cat /etc/oranfstab
server: NFSv3test
path: jfs_svmdr-nfs1
path: jfs_svmdr-nfs2
export: /dbf mount: /oradata
export: /logs mount: /logs
nfs_version: NFSv3
```

La première étape consiste à vérifier que dNFS est opérationnel pour les systèmes de fichiers spécifiés :

```
SQL> select dirname,nfsversion from v$dnfs_servers;

DIRNAME
----------------------------------
NFSVERSION
----------------
/logs
NFSv3.0

/dbf
NFSv3.0
```

Ce résultat indique que dNFS est utilisé avec ces deux systèmes de fichiers, mais que **pas** signifie que oranfstab est opérationnel. Si une erreur était présente, dNFS aurait détecté automatiquement les systèmes de fichiers NFS de l'hôte et il se peut que vous voyiez toujours la même sortie à partir de cette commande.

Les chemins d'accès multiples peuvent être vérifiés comme suit :

```
SQL> select svrname,path,ch_id from v$dnfs_channels;

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------
NFSv3test
jfs_svmdr-nfs1
         0

NFSv3test
jfs_svmdr-nfs2
         1

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------

NFSv3test
jfs_svmdr-nfs1
         0
```

```
NFSv3test
jfs_svmdr-nfs2

[output truncated]

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------
NFSv3test
jfs_svmdr-nfs2
         1

NFSv3test
jfs_svmdr-nfs1
         0

SVRNAME
------------------------------------
PATH
------------------------------------
     CH_ID
----------

NFSv3test
jfs_svmdr-nfs2
         1


66 rows selected.
```

Il s'agit des connexions que dNFS utilise. Deux chemins et canaux sont visibles pour chaque entrée SVRNAME. Cela signifie que les chemins d'accès multiples fonctionnent, ce qui signifie que le fichier oranfstab a été reconnu et traité.

Accès direct au NFS et au système de fichiers hôte

L'utilisation de dNFS peut parfois causer des problèmes pour les applications ou les activités des utilisateurs qui dépendent des systèmes de fichiers visibles montés sur l'hôte car le client dNFS accède au système de fichiers hors bande à partir du système d'exploitation hôte. Le client dNFS peut créer, supprimer et modifier des fichiers sans connaître le système d'exploitation.

Lorsque les options de montage des bases de données à instance unique sont utilisées, elles permettent la mise en cache des attributs de fichiers et de répertoires, ce qui signifie également que le contenu d'un répertoire est mis en cache. Par conséquent, dNFS peut créer un fichier, et il y a un court délai avant que le système d'exploitation ne relise le contenu du répertoire et que le fichier devienne visible pour l'utilisateur. Ce

n'est généralement pas un problème, mais, dans de rares cas, des utilitaires tels que SAP BR*Tools peuvent présenter des problèmes. Si cela se produit, modifiez les options de montage pour utiliser les recommandations pour Oracle RAC. Ce changement entraîne la désactivation de l'ensemble de la mise en cache de l'hôte.

Ne modifiez les options de montage que si (a) dNFS est utilisé et (b) un problème résulte d'un décalage dans la visibilité des fichiers. Si dNFS n'est pas utilisé, les options de montage Oracle RAC sur une base de données à instance unique entraînent une dégradation des performances.

Reportez-vous à la remarque à propos de `nosharecache` la "Options de montage NFS Linux" pour un problème dNFS spécifique à Linux qui peut produire des résultats inhabituels.

Locations et verrouillages NFS

NFSv3 est sans état. Cela signifie que le serveur NFS (ONTAP) ne suit pas les systèmes de fichiers montés, par qui ou quels verrous sont réellement en place.

ONTAP dispose de certaines fonctionnalités qui enregistreront les tentatives de montage. Vous savez donc quels clients accèdent aux données et il se peut que des verrous consultatifs soient présents, mais les informations ne sont pas 100 % complètes. Elle ne peut pas être terminée, car le suivi de l'état du client NFS ne fait pas partie de la norme NFSv3.

État NFSv4

En revanche, NFSv4 est avec état. Le serveur NFSv4 suit les clients qui utilisent les systèmes de fichiers, les fichiers existants, les fichiers et/ou les régions de fichiers verrouillés, etc Cela signifie qu'une communication régulière entre un serveur NFSv4 doit être établie pour maintenir les données d'état à jour.

Les États les plus importants gérés par le serveur NFS sont les verrous NFSv4 et les locations NFSv4, qui sont très étroitement liés. Vous devez comprendre comment chacun fonctionne par lui-même, et comment ils se rapportent les uns aux autres.

Verrous NFSv4

Avec NFSv3, les verrous sont consultatifs. Un client NFS peut toujours modifier ou supprimer un fichier « verrouillé ». Un verrou NFSv3 n'expire pas de lui-même, il doit être supprimé. Cela crée des problèmes. Par exemple, si une application en cluster crée des verrous NFSv3 et que l'un des nœuds tombe en panne, que faire ? Vous pouvez coder l'application sur les nœuds survivants pour supprimer les verrous, mais comment savoir que c'est sûr ? Le nœud « en panne » est peut-être opérationnel, mais ne communique pas avec le reste du cluster ?

Avec NFSv4, les verrous ont une durée limitée. Tant que le client tenant les Locks continue à s'archiver avec le serveur NFSv4, aucun autre client n'est autorisé à acquérir ces Locks. Si un client ne parvient pas à s'archiver avec NFSv4, les verrous seront éventuellement révoqués par le serveur et d'autres clients pourront demander et obtenir des verrous.

Locations NFSv4

Les verrous NFSv4 sont associés à un bail NFSv4. Lorsqu'un client NFSv4 établit une connexion avec un serveur NFSv4, il obtient un bail. Si le client obtient un verrou (il existe plusieurs types de verrous), le verrou est associé au bail.

Ce bail a un délai défini. Par défaut, ONTAP définit la valeur de temporisation sur 30 secondes :

```
Cluster01::*> nfs server show -vserver vserver1 -fields v4-lease-seconds

vserver   v4-lease-seconds
--------- ----------------
vserver1  30
```

Cela signifie qu'un client NFSv4 doit vérifier avec le serveur NFSv4 toutes les 30 secondes pour renouveler ses baux.

Le bail est automatiquement renouvelé par n'importe quelle activité. Ainsi, si le client effectue des travaux, il n'est pas nécessaire d'effectuer des opérations supplémentaires. Si une application devient silencieuse et ne fait pas de véritable travail, elle devra effectuer une sorte d'opération de maintien en vie (appelée SÉQUENCE). Il s'agit essentiellement de dire « Je suis toujours là, veuillez actualiser mes contrats de location ».

```
 *Question:* What happens if you lose network connectivity for 31 seconds?
NFSv3 est sans état. Il ne s'attend pas à ce que les clients communiquent.
NFSv4 est avec état et une fois la période de location expirée, le bail
expire, et les verrous sont révoqués et les fichiers verrouillés sont mis
à disposition des autres clients.
```

Avec NFSv3, vous pouvez déplacer les câbles réseau, redémarrer les switchs réseau, modifier la configuration et être sûr qu'aucun problème ne se produirait. En général, les applications attendront patiemment le bon fonctionnement de la connexion réseau.

Avec NFSv4, vous disposez de 30 secondes (sauf si vous avez augmenté la valeur de ce paramètre dans ONTAP) pour terminer votre travail. Si vous dépassez cette limite, vos contrats de location sont échus. Normalement, cela provoque des pannes d'application.

Par exemple, si vous disposez d'une base de données Oracle et que vous rencontrez une perte de connectivité réseau (parfois appelée « partition réseau ») qui dépasse le délai d'expiration du bail, vous plantez la base de données.

Voici un exemple de ce qui se passe dans le journal des alertes Oracle si cela se produit :

```
2022-10-11T15:52:55.206231-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00202: control file: '/redo0/NTAP/ctrl/control01.ctl'
ORA-27072: File I/O error
Linux-x86_64 Error: 5: Input/output error
Additional information: 4
Additional information: 1
Additional information: 4294967295
2022-10-11T15:52:59.842508-04:00
Errors in file /orabin/diag/rdbms/ntap/NTAP/trace/NTAP_ckpt_25444.trc:
ORA-00206: error in writing (block 3, # blocks 1) of control file
ORA-00202: control file: '/redo1/NTAP/ctrl/control02.ctl'
ORA-27061: waiting for async I/Os failed
```

Si vous examinez les syslog, vous devriez voir plusieurs de ces erreurs :

```
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
Oct 11 15:52:55 host1 kernel: NFS: nfs4_reclaim_open_state: Lock reclaim
failed!
```

Les messages du journal sont généralement le premier signe d'un problème, autre que le blocage de l'application. En général, vous ne voyez rien pendant la panne réseau, car les processus et le système d'exploitation lui-même sont bloqués et tentent d'accéder au système de fichiers NFS.

Les erreurs apparaissent une fois que le réseau est de nouveau opérationnel. Dans l'exemple ci-dessus, une fois la connectivité rétablie, le système d'exploitation a tenté de réacquérir les verrous, mais il était trop tard. Le bail avait expiré et les serrures ont été retirées. Cela entraîne une erreur qui se propage jusqu'à la couche Oracle et provoque le message dans le journal des alertes. Vous pouvez voir des variations sur ces modèles en fonction de la version et de la configuration de la base de données.

En résumé, NFSv3 tolère l'interruption du réseau, mais NFSv4 est plus sensible et impose une période de location définie.

Que se passe-t-il si un délai de 30 secondes n'est pas acceptable ? Que se passe-t-il si vous gérez un réseau changeant de façon dynamique où les commutateurs sont redémarrés ou les câbles sont déplacés et que le résultat est une interruption occasionnelle du réseau ? Vous pouvez choisir de prolonger la période de location, mais pour savoir si vous voulez y parvenir, vous devez expliquer les périodes de grâce NFSv4.

Périodes de grâce NFSv4

Lorsqu'un serveur NFSv3 est redémarré, il est prêt à transmettre les E/S presque instantanément. Il ne maintenait aucune sorte d'état concernant les clients. Le résultat est qu'une opération de basculement ONTAP semble souvent proche de l'instantané. Dès qu'un contrôleur est prêt à commencer à transmettre des données, il envoie un ARP au réseau qui signale le changement de topologie. En règle générale, les clients le détectent presque instantanément et le flux des données reprend.

NFSv4, cependant, fera une courte pause. Cela fait partie du fonctionnement de NFSv4.

> Les sections suivantes sont à jour depuis ONTAP 9.15.1, mais le comportement de bail et de verrouillage ainsi que les options de réglage peuvent changer de version à version. Si vous avez besoin d'ajuster les délais de location/verrouillage de NFSv4, veuillez consulter le support NetApp pour obtenir les informations les plus récentes.

Les serveurs NFSv4 doivent suivre les baux, les verrous et les utilisateurs des données. Si un serveur NFS fonctionne de manière incohérente et redémarre, ou perd de l'alimentation pendant un moment, ou est redémarré pendant l'activité de maintenance, le résultat est le bail/verrouillage et d'autres informations client sont perdues. Le serveur doit déterminer quel client utilise les données avant de reprendre les opérations. C'est là que intervient le délai de grâce.

Si vous mettez soudainement votre serveur NFSv4 hors/sous tension. Lorsqu'il est rétabli, les clients qui tentent de reprendre l'E/S reçoivent une réponse qui dit essentiellement « J'ai perdu les informations de location/verrouillage. Voulez-vous réenregistrer vos verrous ? » C'est le début de la période de grâce. La valeur par défaut est 45 secondes sur ONTAP :

```
Cluster01::> nfs server show -vserver vserver1 -fields v4-grace-seconds

vserver   v4-grace-seconds
--------- ----------------
vserver1  45
```

Par conséquent, après un redémarrage, un contrôleur met en pause les E/S tandis que tous les clients récupèrent leurs baux et verrous. Une fois le délai de grâce terminé, le serveur reprend les opérations d'E/S.

Cette période de grâce contrôle la récupération de bail pendant les modifications de l'interface réseau, mais il existe une deuxième période de grâce qui contrôle la récupération pendant le basculement du stockage `locking.grace_lease_seconds`. Il s'agit d'une option au niveau du nœud.

```
cluster01::> node run [node names or *] options
locking.grace_lease_seconds
```

Par exemple, si vous avez fréquemment besoin d'effectuer des basculements LIF, et que vous devez réduire le délai de grâce, vous changiez . `v4-grace-seconds` Si vous souhaitez améliorer le temps de reprise des E/S pendant le basculement du contrôleur, vous devez le modifier `locking.grace_lease_seconds`.

Ne modifiez ces valeurs qu'avec prudence et après avoir parfaitement compris les risques et les conséquences. Les pauses E/S liées aux opérations de basculement et de migration avec NFSv4.X ne peuvent pas être entièrement évitées. Les périodes de verrouillage, de bail et de grâce font partie de la RFC NFS. Pour de nombreux clients, NFSv3 est préférable, car les délais de basculement sont plus courts.

Délais de location par rapport aux délais de grâce

Le délai de grâce et la période de location sont connectés. Comme mentionné ci-dessus, le délai de bail par défaut est de 30 secondes, ce qui signifie que les clients NFSv4 doivent s'enregistrer auprès du serveur au moins toutes les 30 secondes, sinon ils perdent leur bail et, à leur tour, leurs verrous. Le délai de grâce existe pour permettre à un serveur NFS de reconstruire les données de bail/verrouillage, et il prend par défaut 45

secondes. Le délai de grâce doit être plus long que la période de location. Cela permet de s'assurer qu'un environnement client NFS conçu pour renouveler les contrats de location au moins toutes les 30 secondes aura la possibilité d'archiver avec le serveur après un redémarrage. Un délai de grâce de 45 secondes garantit que tous les clients qui s'attendent à renouveler leur contrat de location au moins toutes les 30 secondes ont certainement l'occasion de le faire.

Si un délai de 30 secondes n'est pas acceptable, vous pouvez choisir de prolonger la période de location.

Si vous souhaitez augmenter le délai de bail à 60 secondes pour résister à une panne réseau de 60 secondes, vous devrez également augmenter le délai de grâce. Une pause d'E/S plus longue sera donc nécessaire lors du basculement du contrôleur.

Ce ne devrait normalement pas être un problème. En général, les utilisateurs ne mettent à jour les contrôleurs ONTAP qu'une ou deux fois par an. En outre, les basculements non planifiés en raison de défaillances matérielles sont extrêmement rares. En outre, si vous aviez un réseau où une panne réseau de 60 secondes était possible, et que le délai de bail était de 60 secondes, vous n'auriez probablement pas à vous opposer à un basculement rare du système de stockage, ce qui aurait entraîné une pause de 61 secondes non plus. Vous avez déjà reconnu que vous disposez d'un réseau qui s'arrête pendant plus de 60 secondes plutôt fréquemment.

Mise en cache NFS

La présence de l'une des options de montage suivantes entraîne la désactivation de la mise en cache de l'hôte :

```
cio, actimeo=0, noac, forcedirectio
```

Ces paramètres peuvent avoir un effet négatif important sur la vitesse d'installation du logiciel, de correction et des opérations de sauvegarde/restauration. Dans certains cas, en particulier avec les applications en cluster, ces options sont obligatoires car elles doivent inévitablement assurer la cohérence du cache sur tous les nœuds du cluster. Dans d'autres cas, les clients utilisent ces paramètres par erreur, ce qui entraîne des dommages inutiles aux performances.

De nombreux clients suppriment temporairement ces options de montage lors de l'installation ou de l'application de correctifs binaires. Cette suppression peut être effectuée en toute sécurité si l'utilisateur vérifie qu'aucun autre processus n'utilise activement le répertoire cible pendant le processus d'installation ou de correction.

Tailles de transfert NFS

Par défaut, ONTAP limite la taille des E/S NFS à 64 Ko.

Les E/S aléatoires utilisent la plupart des applications et bases de données une taille de bloc bien inférieure à la taille maximale de 64 Ko. Les E/S de blocs volumineux sont généralement parallélisées de sorte que le maximum de 64 Ko ne limite pas non plus l'obtention d'une bande passante maximale.

Dans certains cas, le maximum de 64 000 charges de travail entraîne une limitation. En particulier, les opérations à thread unique, telles que les opérations de sauvegarde ou de restauration, ou encore les analyses de table complète de base de données s'exécutent plus rapidement et plus efficacement si la base de données peut exécuter moins d'E/S, mais plus volumineuses. La taille optimale de gestion des E/S pour ONTAP est de 256 Ko.

La taille maximale de transfert pour un SVM ONTAP donné peut être modifiée comme suit :

```
Cluster01::> set advanced
Warning: These advanced commands are potentially dangerous; use them only
when directed to do so by NetApp personnel.
Do you want to continue? {y|n}: y
Cluster01::*> nfs server modify -vserver vserver1 -tcp-max-xfer-size
262144
Cluster01::*>
```

Ne réduisez jamais la taille de transfert maximale autorisée sur ONTAP en dessous de la valeur de rsize/wsize des systèmes de fichiers NFS actuellement montés. Cela peut provoquer des blocages ou même une corruption des données avec certains systèmes d'exploitation. Par exemple, si les clients NFS sont actuellement définis sur une taille rsize/wsize de 65536, la taille maximale du transfert ONTAP peut être ajustée entre 65536 et 1048576 sans effet car les clients eux-mêmes sont limités. Réduire la taille de transfert maximale en dessous de 65536 peut endommager la disponibilité ou les données.