NFS

Enterprise applications

NetApp
February 11, 2026

Índice

NFS

Visão geral

A NetApp fornece storage NFS de nível empresarial há mais de 30 anos, e seu uso está crescendo cada vez mais, levando em consideração a simplicidade de suas infraestruturas baseadas em nuvem.

O protocolo NFS inclui várias versões com requisitos variados. Para obter uma descrição completa da configuração NFS com o ONTAP, "TR-4067 NFS nas práticas recomendadas da ONTAP"consulte . As seções a seguir abrangem alguns dos requisitos mais críticos e erros comuns do usuário.

Versões de NFS

O cliente do sistema operacional NFS deve ter suporte do NetApp.

- O NFSv3 é suportado com os sistemas operacionais que seguem o padrão NFSv3.
- O NFSv3 é compatível com o cliente Oracle DNFS.
- O NFSv4 é compatível com todos os sistemas operacionais que seguem o padrão NFSv4.
- NFSv4,1 e NFSv4,2 requerem suporte específico ao SO. Consulte o "NetApp IMT" para ver os SO suportados.
- O suporte ao Oracle DNFS para NFSv4,1 requer o Oracle 12.2.0.2 ou superior.

> "Matriz de suporte NetApp"O para NFSv3 e NFSv4 não inclui sistemas operacionais específicos. Todos os SO que obedecem ao RFC são geralmente suportados. Ao pesquisar no IMT on-line para suporte a NFSv3 ou NFSv4, não selecione um sistema operacional específico porque não haverá correspondências exibidas. Todos os SO são implicitamente suportados pela política geral.

Tabelas de slots TCP do Linux NFSv3

As tabelas de slot TCP são equivalentes a NFSv3 mm de profundidade de fila do adaptador de barramento do host (HBA). Essas tabelas controlam o número de operações NFS que podem ficar pendentes de uma só vez. O valor padrão é geralmente 16, o que é muito baixo para um desempenho ideal. O problema oposto ocorre em kernels Linux mais recentes, que podem aumentar automaticamente o limite da tabela de slots TCP para um nível que satura o servidor NFS com solicitações.

Para um desempenho ideal e para evitar problemas de desempenho, ajuste os parâmetros do kernel que controlam as tabelas de slots TCP.

Executar o `sysctl -a | grep tcp.*.slot_table` comando e respeitar os seguintes parâmetros:

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

Todos os sistemas Linux devem incluir `sunrpc.tcp_slot_table_entries`, mas apenas alguns incluem

`sunrpc.tcp_max_slot_table_entries`. Ambos devem ser definidos para 128.

> A falha em definir esses parâmetros pode ter efeitos significativos no desempenho. Em alguns casos, o desempenho é limitado porque o sistema operacional linux não está emitindo e/S suficiente Em outros casos, as latências de e/S aumentam à medida que o sistema operacional linux tenta emitir mais e/S do que pode ser reparado.

ADR e NFS

Alguns clientes relataram problemas de desempenho resultantes de uma quantidade excessiva de e/S nos dados `ADR` no local. O problema geralmente não ocorre até que muitos dados de desempenho tenham sido acumulados. A razão para a e/S excessiva é desconhecida, mas este problema parece ser um resultado de processos Oracle repetidamente verificando o diretório de destino para mudanças.

A remoção `noac` das opções e/ou `actimeo=0` montagem permite que o armazenamento em cache do sistema operacional do host ocorra e reduz os níveis de e/S de storage.

> **A NetApp recomenda** não colocar `ADR` dados em um sistema de arquivos com `noac` ou `actimeo=0` porque problemas de desempenho são prováveis. Separe `ADR` os dados em um ponto de montagem diferente, se necessário.

nfs-rootonly e mount-rootonly

O ONTAP inclui uma opção NFS chamada `nfs-rootonly` que controla se o servidor aceita conexões de tráfego NFS de portas altas. Como medida de segurança, apenas o usuário raiz tem permissão para abrir conexões TCP/IP usando uma porta de origem abaixo de 1024, porque essas portas são normalmente reservadas para uso do sistema operacional, não para processos de usuário. Essa restrição ajuda a garantir que o tráfego NFS seja de um cliente NFS do sistema operacional real e não de um processo mal-intencionado emulando um cliente NFS. O cliente Oracle DNFS é um driver de espaço de usuário, mas o processo é executado como root, portanto geralmente não é necessário alterar o valor de `nfs-rootonly`. As conexões são feitas a partir de portas baixas.

A `mount-rootonly` opção só se aplica a NFSv3. Ele controla se a chamada DE MONTAGEM RPC será aceita a partir de portas maiores que 1024. Quando o DNFS é usado, o cliente está novamente executando como root, para que ele possa abrir portas abaixo de 1024. Este parâmetro não tem efeito.

Os processos de abertura de conexões com DNFS em NFS versões 4,0 e superiores não são executados como raiz e, portanto, exigem portas acima de 1024. O `nfs-rootonly` parâmetro deve ser definido como desativado para que o DNFS conclua a conexão.

Se `nfs-rootonly` o estiver ativado, o resultado será uma parada durante a fase de montagem abrindo conexões DNFS. A saída sqlplus é semelhante a esta:

```
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
```

O parâmetro pode ser alterado da seguinte forma:

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

Em situações raras, você pode precisar alterar tanto nfs-rootonly quanto mount-rootonly para desabilitado. Se um servidor estiver gerenciando um número extremamente grande de conexões TCP, é possível que nenhuma porta abaixo de 1024 esteja disponível e o sistema operacional seja forçado a usar portas mais altas. Esses dois parâmetros ONTAP precisariam ser alterados para permitir que a conexão fosse concluída.

Políticas de exportação de NFS: Superusuário e setuid

Se os binários Oracle estiverem localizados em um compartilhamento NFS, a política de exportação deverá incluir permissões de superusuário e setuid.

As exportações de NFS compartilhadas usadas para serviços de arquivos genéricos, como diretórios home do usuário, geralmente esmagam o usuário raiz. Isso significa que uma solicitação do usuário root em um host que montou um sistema de arquivos é remapeada como um usuário diferente com Privileges inferior. Isso ajuda a proteger os dados, impedindo que um usuário root em um servidor específico acesse dados no servidor compartilhado. O bit setuid também pode ser um risco de segurança em um ambiente compartilhado. O bit setuid permite que um processo seja executado como um usuário diferente do usuário que invoca o comando. Por exemplo, um script shell que era de propriedade do root com o bit setuid é executado como root. Se esse script shell pudesse ser alterado por outros usuários, qualquer usuário que não seja root poderá emitir um comando como root atualizando o script.

Os binários Oracle incluem arquivos de propriedade do root e usam o bit setuid. Se os binários Oracle estiverem instalados em um compartilhamento NFS, a política de exportação deverá incluir as permissões de superusuário e setuid apropriadas. No exemplo abaixo, a regra inclui tanto `allow-suid` e permite `superuser` o acesso (raiz) para clientes NFS usando autenticação de sistema.

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

Configuração NFSv4/4,1

Para a maioria das aplicações, há muito pouca diferença entre NFSv3 e NFSv4. A e/S da aplicação geralmente é muito simples e/S e não se beneficia significativamente de alguns dos recursos avançados disponíveis no NFSv4. Versões mais altas do NFS não devem ser vistas como uma "atualização" da perspectiva do storage de banco de dados, mas sim como versões do NFS que incluem recursos adicionais. Por exemplo, se a segurança de ponta a ponta do modo de privacidade Kerberos (krb5p) for necessária, então NFSv4 será necessário.

A NetApp recomenda usar o NFSv4,1 se forem necessários recursos do NFSv4. Há algumas melhorias funcionais no protocolo NFSv4 em NFSv4,1 que melhoram a resiliência em certos casos de borda.

Mudar para NFSv4 é mais complicado do que simplesmente mudar as opções de montagem de vers-3 para vers-4,1. Uma explicação mais completa da configuração do NFSv4 com o ONTAP, incluindo orientações sobre a configuração do sistema operacional, "TR-4067 NFS nas práticas recomendadas da ONTAP"consulte . As secções seguintes deste TR explicam alguns dos requisitos básicos para a utilização do NFSv4.

Domínio NFSv4

Uma explicação completa da configuração NFSv4/4,1 está além do escopo deste documento, mas um problema comumente encontrado é uma incompatibilidade no mapeamento de domínio. De um ponto de vista sysadmin, os sistemas de arquivos NFS parecem se comportar normalmente, mas os aplicativos relatam erros sobre permissões e/ou setuid em determinados arquivos. Em alguns casos, os administradores concluíram incorretamente que as permissões dos binários do aplicativo foram danificadas e executaram comandos chown ou chmod quando o problema real era o nome do domínio.

O nome de domínio NFSv4 é definido no ONTAP SVM:

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

O nome de domínio NFSv4 no host é definido em `/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
```

Os nomes de domínio devem corresponder. Se não o fizerem, erros de mapeamento semelhantes aos seguintes aparecem em `/var/log/messages`:

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

Binários de aplicativos, como binários de banco de dados Oracle, incluem arquivos de propriedade do root com o bit setuid, o que significa que uma incompatibilidade nos nomes de domínio NFSv4 causa falhas na inicialização do Oracle e um aviso sobre a propriedade ou permissões de um arquivo chamado `oradism`, que está localizado no `$ORACLE_HOME/bin` diretório. Deve aparecer da seguinte forma:

```
[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
```

Se este arquivo aparecer com a propriedade de ninguém, pode haver um problema de mapeamento de domínio NFSv4.

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

Para corrigir isso, verifique o `/etc/idmap.cfg` arquivo na configuração v4-id-domain no ONTAP e certifique-se de que eles sejam consistentes. Se não estiverem, faça as alterações necessárias, execute `nfsidmap -c` e aguarde um momento para que as alterações se propaguem. A propriedade do arquivo deve então ser devidamente reconhecida como raiz. Se um usuário tivesse tentado executar `chown root` esse arquivo antes que a configuração dos domínios NFS fosse corrigida, talvez seja necessário executar `chown root` novamente.

Oracle Direct NFS (DNFS)

Os bancos de dados Oracle podem usar o NFS de duas maneiras.

Primeiro, ele pode usar um sistema de arquivos montado usando o cliente NFS nativo que faz parte do sistema operacional. Isso às vezes é chamado de kernel NFS, ou kNFS. O sistema de arquivos NFS é montado e usado pelo banco de dados Oracle exatamente o mesmo que qualquer outro aplicativo usaria um sistema de arquivos NFS.

O segundo método é o Oracle Direct NFS (DNFS). Esta é uma implementação do padrão NFS no software de banco de dados Oracle. Ele não altera a maneira como os bancos de dados Oracle são configurados ou gerenciados pelo DBA. Desde que o próprio sistema de armazenamento tenha as configurações corretas, o uso do DNFS deve ser transparente para o grupo DBA e para os usuários finais.

Um banco de dados com o recurso DNFS habilitado ainda tem os sistemas de arquivos NFS usuais montados. Uma vez que o banco de dados está aberto, o banco de dados Oracle abre um conjunto de sessões TCP/IP e executa operações NFS diretamente.

NFS direto

O principal valor do NFS direto da Oracle é ignorar o cliente NFS do host e executar operações de arquivos NFS diretamente em um servidor NFS. A ativação da TI requer apenas a alteração da biblioteca do Oracle Disk Manager (ODM). As instruções para este processo são fornecidas na documentação da Oracle.

O uso do DNFS resulta em uma melhoria significativa no desempenho de e/S e diminui a carga no host e no sistema de armazenamento, pois a e/S é realizada da maneira mais eficiente possível.

Além disso, o Oracle DNFS inclui uma **opção** para multipathing de interface de rede e tolerância a falhas. Por exemplo, duas interfaces 10Gb podem ser Unidas para oferecer 20Gb Gbps de largura de banda. Uma falha de uma interface faz com que a I/o seja tentada novamente na outra interface. A operação geral é muito semelhante ao multipathing FC. Multipathing era comum anos atrás, quando 1GB ethernet era o padrão mais comum. Uma NIC 10Gb é suficiente para a maioria das cargas de trabalho Oracle, mas se for necessário mais, 10Gb NICs podem ser colados.

Quando o DNFS é usado, é fundamental que todos os patches descritos no Oracle Doc 1495104,1 sejam instalados. Se um patch não puder ser instalado, o ambiente deve ser avaliado para garantir que os bugs descritos nesse documento não causem problemas. Em alguns casos, a incapacidade de instalar os patches necessários impede o uso do DNFS.

Não use DNFS com qualquer tipo de resolução de nome de round-robin, incluindo DNS, DDNS, NIS ou qualquer outro método. Isso inclui o recurso de balanceamento de carga DNS disponível no ONTAP. Quando um banco de dados Oracle usando DNFS resolve um nome de host para um endereço IP, ele não deve ser alterado em pesquisas subsequentes. Isso pode resultar em falhas de banco de dados Oracle e possível corrupção de dados.

Ativar DNFS

O Oracle DNFS pode trabalhar com o NFSv3 sem necessidade de configuração além de ativar a biblioteca DNFS (consulte a documentação Oracle para o comando específico necessário), mas se o DNFS não conseguir estabelecer conetividade, ele pode reverter silenciosamente para o cliente NFS do kernel. Se isso acontecer, o desempenho pode ser gravemente afetado.

Se você deseja usar a multiplexação DNFS em várias interfaces, com NFSv4.X, ou usar criptografia, você deve configurar um arquivo oranfstab. A sintaxe é extremamente rigorosa. Pequenos erros no arquivo podem resultar em suspensão de inicialização ou ignorar o arquivo oranfstab.

No momento da escrita, o multipathing DNFS não funciona com o NFSv4,1 com versões recentes do Oracle Database. Um arquivo oranfstab que especifica NFSv4,1 como um protocolo só pode usar uma instrução de caminho único para uma determinada exportação. O motivo é que o ONTAP não oferece suporte ao entroncamento ClientID. Patches do banco de dados Oracle para resolver essa limitação podem estar disponíveis no futuro.

A única maneira de ter certeza de que o DNFS está operando como esperado é consultar as tabelas dnfs.

Abaixo está um exemplo de arquivo oranfstab localizado em /etc. Este é um dos vários locais que um arquivo oranfstab pode ser colocado.

```
[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
```

O primeiro passo é verificar se o DNFS está operacional para os sistemas de arquivos especificados:

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

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

/dbf
NFSv3.0
```

Essa saída indica que o DNFS está em uso com esses dois sistemas de arquivos, mas ele não significa que o oranfstab esteja operacional. Se um erro estivesse presente, o DNFS teria descoberto automaticamente os sistemas de arquivos NFS do host e você ainda poderá ver a mesma saída deste comando.

Multipathing pode ser verificado da seguinte forma:

```
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.
```

Estas são as conexões que o DNFS está usando. Dois caminhos e canais são visíveis para cada entrada SVRNAME. Isso significa que o multipathing está funcionando, o que significa que o arquivo oranfstab foi reconhecido e processado.

Acesso direto ao NFS e ao sistema de arquivos do host

O uso do DNFS pode ocasionalmente causar problemas para aplicativos ou atividades do usuário que dependem dos sistemas de arquivos visíveis montados no host porque o cliente DNFS acessa o sistema de arquivos fora da banda do sistema operacional do host. O cliente DNFS pode criar, excluir e modificar arquivos sem o conhecimento do sistema operacional.

Quando as opções de montagem para bancos de dados de instância única são usadas, elas permitem o armazenamento em cache de atributos de arquivo e diretório, o que também significa que o conteúdo de um diretório é armazenado em cache. Portanto, o DNFS pode criar um arquivo, e há um curto atraso antes que o sistema operacional releia o conteúdo do diretório e o arquivo se torne visível para o usuário. Isso geralmente

não é um problema, mas, em raras ocasiões, utilitários como SAP BR*Tools podem ter problemas. Se isso acontecer, solucione o problema alterando as opções de montagem para usar as recomendações do Oracle RAC. Essa alteração resulta na desativação de todo o cache do host.

Altere apenas as opções de montagem quando (a) DNFS for usado e (b) um problema resulta de um atraso na visibilidade do arquivo. Se o DNFS não estiver em uso, o uso das opções de montagem do Oracle RAC em um banco de dados de instância única resulta em desempenho degradado.

Veja a nota sobre `nosharecache` o in "Opções de montagem em NFS do Linux" para um problema DNFS específico do Linux que pode produzir resultados incomuns.

Locações e bloqueios de NFS

O NFSv3 está sem monitoração de estado. Isso efetivamente significa que o servidor NFS (ONTAP) não controla quais sistemas de arquivos são montados, por quem, ou quais bloqueios estão realmente no lugar.

O ONTAP tem alguns recursos que gravarão tentativas de montagem para que você tenha uma ideia de quais clientes podem estar acessando dados, e pode haver bloqueios de consultoria presentes, mas essa informação não é garantida para ser 100% completa. Ele não pode ser concluído, porque o rastreamento do estado do cliente NFS não faz parte do padrão NFSv3.

NFSv4 declaração

Em contraste, NFSv4 é stateful. O servidor NFSv4 rastreia quais clientes estão usando quais sistemas de arquivos, quais arquivos existem, quais arquivos e/ou regiões de arquivos estão bloqueados, etc. isso significa que precisa haver comunicação regular entre um servidor NFSv4 para manter os dados de estado atuais.

Os estados mais importantes que estão sendo gerenciados pelo servidor NFS são NFSv4 bloqueios e NFSv4 arrendamentos, e eles estão muito interligados. Você precisa entender como cada um funciona por si mesmo, e como eles se relacionam uns com os outros.

NFSv4 fechaduras

Com o NFSv3, os bloqueios são consultivos. Um cliente NFS ainda pode modificar ou excluir um arquivo "bloqueado". Um bloqueio NFSv3 não expira por si só, ele deve ser removido. Isso cria problemas. Por exemplo, se você tiver um aplicativo em cluster que crie NFSv3 bloqueios e um dos nós falhar, o que você faz? Você pode codificar o aplicativo nos nós sobreviventes para remover os bloqueios, mas como você sabe que isso é seguro? Talvez o nó "falhou" esteja operacional, mas não esteja se comunicando com o restante do cluster?

Com NFSv4, os bloqueios têm uma duração limitada. Desde que o cliente que detém os bloqueios continue a efetuar o check-in com o servidor NFSv4, nenhum outro cliente tem permissão para adquirir esses bloqueios. Se um cliente não conseguir efetuar o check-in com o NFSv4, os bloqueios eventualmente serão revogados pelo servidor e outros clientes poderão solicitar e obter bloqueios.

NFSv4 arrendamentos

NFSv4 bloqueios estão associados a um leasing de NFSv4. Quando um cliente NFSv4 estabelece uma conexão com um servidor NFSv4, ele recebe um leasing. Se o cliente obtém um bloqueio (existem muitos tipos de bloqueios), então o bloqueio está associado ao leasing.

Este leasing tem um tempo limite definido. Por padrão, o ONTAP definirá o valor de tempo limite para 30 segundos:

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

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

Isso significa que um cliente NFSv4 precisa fazer o check-in com o servidor NFSv4 a cada 30 segundos para renovar suas locações.

A locação é renovada automaticamente por qualquer atividade, portanto, se o cliente estiver fazendo trabalho, não há necessidade de realizar operações de adição. Se um aplicativo ficar quieto e não estiver fazendo trabalho real, ele precisará executar uma espécie de operação keep-alive (chamada de SEQUÊNCIA). É essencialmente apenas dizer "Eu ainda estou aqui, por favor, atualize meus arrendamentos."

```
 *Question:* What happens if you lose network connectivity for 31 seconds?
O NFSv3 está sem monitoração de estado. Não está à espera de comunicação
dos clientes. O NFSv4 tem estado monitorado e, uma vez decorrido esse
período de locação, o leasing expira e os bloqueios são revogados e os
arquivos bloqueados são disponibilizados para outros clientes.
```

Com o NFSv3, você pode mover cabos de rede, reinicializar switches de rede, fazer alterações de configuração e ter certeza de que nada de ruim aconteceria. Os aplicativos normalmente apenas esperariam pacientemente para que a conexão de rede funcione novamente.

Com NFSv4, você tem 30 segundos (a menos que você tenha aumentado o valor desse parâmetro dentro do ONTAP) para concluir seu trabalho. Se você exceder isso, suas locações expiram. Normalmente, isso resulta em falhas no aplicativo.

Como exemplo, se você tiver um banco de dados Oracle e tiver uma perda de conetividade de rede (às vezes chamada de "partição de rede") que exceda o tempo limite de concessão, você irá travar o banco de dados.

Aqui está um exemplo do que acontece no log de alerta Oracle se isso acontecer:

```
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
```

Se você olhar para os syslogs, você deve ver vários desses erros:

```
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!
```

As mensagens de log são geralmente o primeiro sinal de um problema, além do congelamento do aplicativo. Normalmente, você não vê nada durante a interrupção da rede porque os processos e o próprio sistema operacional estão bloqueados tentando acessar o sistema de arquivos NFS.

Os erros aparecem depois que a rede estiver operacional novamente. No exemplo acima, uma vez que a conetividade foi restabelecida, o sistema operacional tentou readquirir os bloqueios, mas era tarde demais. O arrendamento expirou e os bloqueios foram removidos. Isso resulta em um erro que se propaga até a camada Oracle e causa a mensagem no log de alerta. Você pode ver variações desses padrões dependendo da versão e configuração do banco de dados.

Em resumo, o NFSv3 tolera a interrupção da rede, mas o NFSv4 é mais sensível e impõe um período de locação definido.

E se um tempo limite de 30 segundos não for aceitável? E se você gerenciar uma rede que muda dinamicamente onde os switches são reinicializados ou os cabos são realocados e o resultado é a interrupção ocasional da rede? Você pode optar por estender o período de locação, mas se você quiser fazer isso requer uma explicação de NFSv4 períodos de carência.

NFSv4 períodos de carência

Se um servidor NFSv3 for reinicializado, ele estará pronto para servir o IO quase instantaneamente. Não estava mantendo qualquer tipo de estado sobre os clientes. O resultado é que uma operação de aquisição da ONTAP geralmente parece estar próxima de instantânea. No momento em que um controlador estiver pronto para começar a fornecer dados, ele enviará um ARP para a rede que sinaliza a alteração na topologia. Os clientes normalmente detetam isso quase instantaneamente e os dados continuam fluindo.

NFSv4, no entanto, irá produzir uma breve pausa. É apenas parte de como o NFSv4 funciona.

> As seções a seguir são atuais a partir do ONTAP 9.15,1, mas o comportamento de leasing e bloqueio, bem como as opções de ajuste podem mudar de versão para versão. Se você precisar ajustar os tempos de leasing/bloqueio NFSv4, consulte o suporte da NetApp para obter as informações mais recentes.

Os servidores NFSv4 precisam rastrear os arrendamentos, bloqueios e quem está usando quais dados. Se um servidor NFS picar e reiniciar, ou perder energia por um momento, ou for reiniciado durante a atividade de manutenção, o resultado será o leasing/bloqueio e outras informações do cliente serão perdidas. O servidor precisa descobrir qual cliente está usando quais dados antes de retomar as operações. É aqui que entra o período de carência.

Se você de repente ligar o seu servidor NFSv4. Quando ele voltar, os clientes que tentam retomar o IO receberão uma resposta que diz essencialmente: "Perdi informações de leasing/bloqueio. Pretende registar novamente os seus bloqueios?" Esse é o início do período de carência. O padrão é 45 segundos no ONTAP:

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

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

O resultado é que, após uma reinicialização, um controlador irá pausar o IO enquanto todos os clientes recuperam seus arrendamentos e bloqueios. Quando o período de carência terminar, o servidor retomará as operações de e/S.

Esse período de carência controla a recuperação de leasing durante alterações na interface de rede, mas há um segundo período de carência que controla a recuperação durante o failover de storage, `locking.grace_lease_seconds`. Esta é uma opção de nível de nó.

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

Por exemplo, se você precisasse frequentemente executar failovers de LIF e precisasse reduzir o período de carência, você mudaria `v4-grace-seconds`. Se você quisesse melhorar o tempo de retomada de e/S durante o failover da controladora, seria necessário alterar `locking.grace_lease_seconds`o .

Apenas altere estes valores com cautela e depois de compreender completamente os riscos e consequências. As pausas de e/S envolvidas com operações de failover e migração com o NFSv4.X não podem ser totalmente evitadas. Os períodos de bloqueio, leasing e carência fazem parte da RFC do NFS. Para muitos clientes, o NFSv3 é preferível porque os tempos de failover são mais rápidos.

Prazos de concessão vs períodos de carência

O período de carência e o período de leasing estão ligados. Como mencionado acima, o tempo limite de leasing padrão é de 30 segundos, o que significa que NFSv4 clientes devem fazer check-in com o servidor pelo menos a cada 30 segundos ou perder seus arrendamentos e, por sua vez, seus bloqueios. O período de carência existe para permitir que um servidor NFS reconstrua dados de concessão/bloqueio e o padrão é de

45 segundos. O período de carência deve ser superior ao período de locação. Isso garante que um ambiente cliente NFS projetado para renovar contratos de arrendamento pelo menos a cada 30 segundos terá a capacidade de fazer check-in com o servidor após uma reinicialização. Um período de carência de 45 segundos garante que todos os clientes que esperam renovar seus arrendamentos pelo menos a cada 30 segundos definitivamente tenham a oportunidade de fazê-lo.

Se um tempo limite de 30 segundos não for aceitável, você pode optar por estender o período de locação.

Se você quiser aumentar o tempo limite de leasing para 60 segundos para suportar uma interrupção de rede de 60 segundos, você também terá que aumentar o período de carência. Isso significa que você terá pausas de e/S mais longas durante o failover de controladora.

Isso normalmente não deve ser um problema. Os usuários típicos atualizam somente as controladoras ONTAP uma ou duas vezes por ano, e o failover não planejado devido a falhas de hardware é extremamente raro. Além disso, se você tivesse uma rede em que uma interrupção de rede de 60 segundos era uma possibilidade preocupante, e você precisasse do tempo limite da concessão para 60 segundos, provavelmente você não obteria o failover raro do sistema de storage, resultando em uma pausa de 61 segundos também. Você já reconheceu que tem uma rede que está pausando por mais de 60 segundos com bastante frequência.

Armazenamento em cache NFS

A presença de qualquer uma das seguintes opções de montagem faz com que o cache do host seja desativado:

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

Essas configurações podem ter um efeito negativo grave na velocidade de instalação do software, patches e operações de backup/restauração. Em alguns casos, especialmente com aplicações em cluster, essas opções são necessárias como resultado inevitável da necessidade de fornecer coerência de cache em todos os nós do cluster. Em outros casos, os clientes usam erroneamente esses parâmetros e o resultado é danos desnecessários no desempenho.

Muitos clientes removem temporariamente essas opções de montagem durante a instalação ou o patch dos binários da aplicação. Essa remoção pode ser realizada com segurança se o usuário verificar que nenhum outro processo está usando ativamente o diretório de destino durante o processo de instalação ou patch.

Tamanhos de transferência NFS

Por padrão, o ONTAP limita os tamanhos de e/S de NFS a 64K.

A e/S aleatória com a maioria dos aplicativos e bancos de dados usa um tamanho de bloco muito menor que está bem abaixo do máximo 64K. A e/S de bloco grande geralmente é paralelizada, portanto o máximo de 64K GB também não é uma limitação para obter largura de banda máxima.

Existem algumas cargas de trabalho em que o máximo 64K cria uma limitação. Em particular, operações de um único processo, como operação de backup ou recuperação ou uma verificação de tabela completa de banco de dados, são executadas de forma mais rápida e eficiente se o banco de dados puder executar menos, mas maiores I/os. O tamanho ideal de manuseio de e/S para ONTAP é 256K.

O tamanho máximo de transferência para um determinado SVM do ONTAP pode ser alterado da seguinte forma:

```
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::*>
```

Nunca diminua o tamanho máximo de transferência permitido no ONTAP abaixo do valor de rsize/wsize dos sistemas de arquivos NFS atualmente instalados. Isso pode criar pendências ou até mesmo corrupção de dados com alguns sistemas operacionais. Por exemplo, se os clientes NFS estiverem atualmente definidos em um rsize/wsize de 65536, o tamanho máximo de transferência do ONTAP poderá ser ajustado entre 65536 e 1048576 sem efeito porque os próprios clientes são limitados. Reduzir o tamanho máximo de transferência abaixo de 65536 pode danificar a disponibilidade ou os dados.

Informações sobre direitos autorais

Informações sobre marcas comerciais