



Configuración de hosts

Enterprise applications

NetApp
May 19, 2024

Tabla de contenidos

- Configuración de hosts. 1
 - Bases de datos de Oracle con IBM AIX. 1
 - Bases de datos de Oracle con HP-UX. 2
 - Bases de datos de Oracle con Linux 4
 - Bases de datos Oracle con ASMLib/AFD (controlador de filtro de ASM) 8
 - Bases de datos de Oracle con Microsoft Windows 10
 - Bases de datos Oracle con Solaris 11

Configuración de hosts

Bases de datos de Oracle con IBM AIX

Temas de configuración para bases de datos de Oracle en IBM AIX con ONTAP.

I/O concurrente

Lograr un rendimiento óptimo en IBM AIX requiere el uso de E/S concurrentes Sin operaciones de I/O simultáneas, es probable que las limitaciones de rendimiento se deban a que AIX realiza I/O atómicas serializadas, lo que conlleva una sobrecarga significativa.

En un principio, NetApp recomendó utilizar el `cio` Opción de montaje para forzar el uso de E/S concurrentes en el sistema de archivos, pero este proceso tenía inconvenientes y ya no es necesario. Desde la introducción de AIX 5.2 y Oracle 10gR1, Oracle en AIX puede abrir archivos individuales para I/O simultánea, en lugar de forzar las operaciones de I/O simultáneas en todo el sistema de archivos.

El mejor método para habilitar E/S concurrente es establecer el `init.ora` parámetro `filesystemio_options` para `setall`. Al hacerlo, Oracle puede abrir archivos específicos para utilizarlos con E/S simultáneas

Uso `cio` Como opción de montaje fuerza el uso de I/O concurrente, lo cual puede tener consecuencias negativas. Por ejemplo, al forzar la E/S simultánea se desactiva la lectura anticipada en los sistemas de archivos, lo que puede dañar el rendimiento de las E/S que se producen fuera del software de la base de datos Oracle, como copiar archivos y realizar copias de seguridad en cinta. Además, productos como Oracle GoldenGate y SAP BR*Tools no son compatibles con el uso del `cio` Opción de montaje con determinadas versiones de Oracle.

NetApp recomienda lo siguiente:



- No utilice la `cio` opción de montaje en el nivel de sistema de archivos. En su lugar, habilite la I/O simultánea mediante el uso de `filesystemio_options=setall`.
- Utilice sólo el `cio` la opción de montaje debería si no es posible configurarla `filesystemio_options=setall`.

Opciones de montaje de AIX NFS

En la siguiente tabla, se enumeran las opciones de montaje de AIX NFS para bases de datos de instancia única de Oracle.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,intr</code>

En la siguiente tabla, se enumeran las opciones de montaje de AIX NFS para RAC.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr,noac</code>
CRS/Voting	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr,noac</code>
Específico ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144</code>
Compartido ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr</code>

La diferencia principal entre las opciones de montaje de instancia única y RAC es la adición de `noac` a las opciones de montaje. Esta adición tiene el efecto de deshabilitar el almacenamiento en caché del SO del host que permite que todas las instancias del clúster RAC tengan una vista uniforme del estado de los datos.

Aunque utilice el `cio` monte la opción y la `init.ora` parámetro `filesystemio_options=setall` tiene el mismo efecto de deshabilitar el almacenamiento en caché de host, sigue siendo necesario utilizarlo `noac`. `noac` es necesario para el uso compartido ORACLE_HOME Despliegues para facilitar la coherencia de archivos como archivos de contraseñas de Oracle y `spfile` archivos de parámetros. Si cada instancia de un clúster de RAC tiene un dedicado ORACLE_HOME, entonces este parámetro no es necesario.

Opciones de montaje jfs/JFS2 de AIX

En la siguiente tabla se enumeran las opciones de montaje jfs/JFS2 de AIX.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	Valores predeterminados
Archivos de control Archivos de datos Rehacer registros	Valores predeterminados
ORACLE_HOME	Valores predeterminados

Antes de utilizar AIX `hdisk` los dispositivos de cualquier entorno, incluidas las bases de datos, comprueban el parámetro `queue_depth`. Este parámetro no es la profundidad de la cola del HBA, más bien se relaciona con la profundidad de la cola SCSI de una persona `hdisk device`. Depending on how the LUNs are configured, the value for `queue_depth` puede ser demasiado bajo para un buen rendimiento. Las pruebas han demostrado que el valor óptimo es 64.

Bases de datos de Oracle con HP-UX

Temas de configuración para bases de datos de Oracle en HP-UX con ONTAP.

Opciones de montaje NFS de HP-UX

En la siguiente tabla se enumeran las opciones de montaje de HP-UX NFS para una única instancia.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,suid</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,forcedirectio, nointr,suid</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,suid</code>

En la siguiente tabla se enumeran las opciones de montaje de HP-UX NFS para RAC.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,noac,suid</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw, bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr,noac,forcedirectio,suid</code>
CRS/votación	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr,noac, forcedirectio,suid</code>
Específico ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,suid</code>
Compartido ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr,noac,suid</code>

La diferencia principal entre las opciones de montaje de instancia única y RAC es la adición de `noac` y.. `forcedirectio` a las opciones de montaje. Esta adición tiene el efecto de deshabilitar el almacenamiento en caché del sistema operativo del host, lo que permite que todas las instancias del clúster RAC tengan una vista coherente del estado de los datos. Aunque utilice el `init.ora` parámetro `filesystemio_options=setall` tiene el mismo efecto de deshabilitar el almacenamiento en caché de host, sigue siendo necesario utilizarlo `noac` y.. `forcedirectio`.

La razón `noac` es necesario para el uso compartido ORACLE_HOME Despliegues es para facilitar la coherencia de archivos como archivos de contraseñas de Oracle y archivos `spfiles`. Si cada instancia de un clúster de RAC tiene un dedicado ORACLE_HOME, este parámetro no es necesario.

Opciones de montaje HP-UX VxFS

Utilice las siguientes opciones de montaje para sistemas de archivos que alojan binarios de Oracle:

```
delaylog,nodatainlog
```

Utilice las siguientes opciones de montaje para sistemas de archivos que contienen archivos de datos, redo logs, archive logs y archivos de control en los que la versión de HP-UX no admite E/S simultáneas:

```
nodatainlog,mincache=direct,convosync=direct
```

Cuando se admiten E/S simultáneas (VxFS 5.0.1 y posteriores, o con ServiceGuard Storage Management Suite), utilice estas opciones de montaje para sistemas de archivos que contengan archivos de datos, redo logs, archive logs y archivos de control:

```
delaylog,cio
```



El parámetro `db_file_multiblock_read_count` Es especialmente crítico en entornos VxFS. Oracle recomienda que este parámetro permanezca sin definir en Oracle 10g R1 y posteriores a menos que se indique lo contrario específicamente. El valor por defecto con un tamaño de bloque de Oracle 8KB es 128. Si el valor de este parámetro se fuerza a 16 o menos, quite el `convosync=direct` Opción de montaje porque puede dañar el rendimiento de I/O secuencial. Este paso daña otros aspectos del rendimiento y solo debe tomarse si el valor de `db_file_multiblock_read_count` debe cambiarse a partir del valor predeterminado.

Bases de datos de Oracle con Linux

Temas de configuración específicos del sistema operativo Linux.

Tablas de ranuras TCP Linux NFSv3

Las tablas de ranuras TCP son equivalentes a NFSv3 a la profundidad de la cola del adaptador de bus de host (HBA). En estas tablas se controla el número de operaciones de NFS que pueden extraordinarias a la vez. El valor predeterminado suele ser 16, que es demasiado bajo para un rendimiento óptimo. El problema opuesto ocurre en los kernels más nuevos de Linux, que pueden aumentar automáticamente el límite de la tabla de ranuras TCP a un nivel que sature el servidor NFS con solicitudes.

Para obtener un rendimiento óptimo y evitar problemas de rendimiento, ajuste los parámetros del núcleo que controlan las tablas de ranuras TCP.

Ejecute el `sysctl -a | grep tcp.*.slot_table` command, y observe los siguientes parámetros:

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

Todos los sistemas Linux deben incluir `sunrpc.tcp_slot_table_entries`, pero solo algunos incluyen `sunrpc.tcp_max_slot_table_entries`. Ambos deben establecerse en 128.

Precaución

Si no se establecen estos parámetros, puede tener efectos significativos en el rendimiento. En algunos casos, el rendimiento es limitado porque el sistema operativo linux no está emitiendo suficiente I/O. En otros casos, las latencias de I/O aumentan cuando el sistema operativo linux intenta emitir más operaciones de I/O de las que se pueden mantener.

Opciones de montaje de Linux NFS

En la siguiente tabla, se enumeran las opciones de montaje de NFS de Linux para una instancia única.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr</code>

La siguiente tabla enumera las opciones de montaje de NFS de Linux para RAC.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,actimeo=0</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,actimeo=0</code>
CRS/votación	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,noac,actimeo=0</code>
Específico ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144</code>
Compartido ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsize=262144,wsiz=262144,nointr,actimeo=0</code>

La diferencia principal entre las opciones de montaje de instancia única y RAC es la adición de `actimeo=0` a las opciones de montaje. Esta adición tiene el efecto de deshabilitar el almacenamiento en caché del sistema operativo del host, lo que permite que todas las instancias del clúster RAC tengan una vista coherente del estado de los datos. Aunque utilice el `init.ora` parámetro `filesystemio_options=setall` tiene el mismo efecto de deshabilitar el almacenamiento en caché de host, sigue siendo necesario utilizarlo

`actimeo=0`.

La razón `actimeo=0` es necesario para el uso compartido `ORACLE_HOME` Despliegues es para facilitar la consistencia de archivos como los archivos de contraseñas de Oracle y `spfiles`. Si cada instancia de un clúster de RAC tiene un dedicado `ORACLE_HOME`, entonces este parámetro no es necesario.

Por lo general, los archivos que no son de base de datos se deben montar con las mismas opciones utilizadas para los archivos de datos de instancia única, aunque las aplicaciones específicas pueden tener requisitos diferentes. Evite las opciones de montaje `noac` y.. `actimeo=0` si es posible, ya que estas opciones desactivan la lectura anticipada y el almacenamiento en búfer de nivel de sistema de archivos. Esto puede causar graves problemas de rendimiento en procesos como extracción, traducción y carga.

ACCESO y GETATTR

Algunos clientes han observado que un nivel extremadamente alto de otros IOPS, como EL ACCESO y GETATTR, puede dominar sus cargas de trabajo. En casos extremos, las operaciones como las de lectura y escritura pueden ser tan bajas como el 10 % del total. Este es un comportamiento normal con cualquier base de datos que incluya el uso `actimeo=0` y/o. `noac` En Linux, porque estas opciones provocan que el sistema operativo Linux vuelva a cargar constantemente los metadatos de los archivos del sistema de almacenamiento. Las operaciones como EL ACCESO y GETATTR son operaciones de bajo impacto que se proporcionan desde la caché de ONTAP en un entorno de base de datos. No se deben considerar IOPS auténticos, como lecturas y escrituras, que crean una demanda real de sistemas de almacenamiento. Sin embargo, estas otras IOPS crean una cierta carga, sobre todo en entornos RAC. Para solucionar esta situación, habilite DNFS, que omite la caché de buffers del sistema operativo y evita estas operaciones de metadatos innecesarias.

NFS directo de Linux

Una opción de montaje adicional denominada `nosharecache`, Es necesario cuando (a) DNFS está activado y (b) un volumen de origen se monta más de una vez en un único servidor (c) con un montaje NFS anidado. Esta configuración se ve principalmente en entornos que admiten aplicaciones SAP. Por ejemplo, un único volumen de un sistema NetApp podría tener un directorio ubicado en `/vol/oracle/base` y un segundo en `/vol/oracle/home`. Si `/vol/oracle/base` está montado en `/oracle` y.. `/vol/oracle/home` está montado en `/oracle/home`, El resultado son montajes NFS anidados que se originan en la misma fuente.

El sistema operativo puede detectar el hecho de que `/oracle` y.. `/oracle/home` residir en el mismo volumen, que es el mismo sistema de archivos de origen. A continuación, el sistema operativo utiliza el mismo identificador de dispositivo para acceder a los datos. Al hacerlo, se mejora el uso del almacenamiento en caché del sistema operativo y algunas otras operaciones, pero interfiere con DNFS. Si DNFS debe acceder a un archivo, como el `spfile`, activado `/oracle/home`, podría intentar erróneamente utilizar la ruta incorrecta a los datos. El resultado es una operación de I/O con errores. En estas configuraciones, agregue la `nosharecache` Opción de montaje en cualquier sistema de archivos NFS que comparta un volumen FlexVol de origen con otro sistema de archivos NFS en ese host. Al hacerlo, se fuerza al sistema operativo Linux a asignar un identificador de dispositivo independiente para ese sistema de archivos.

Linux Direct NFS y Oracle RAC

El uso de DNFS ofrece ventajas especiales de rendimiento para Oracle RAC en el sistema operativo Linux, ya que Linux no dispone de un método para forzar la entrada/salida directa, que se necesita con RAC para lograr coherencia entre los nodos. Como solución alternativa, Linux requiere el uso de `actimeo=0` Opción de montaje, que hace que los datos de archivo caduquen inmediatamente desde la caché del sistema operativo. Esta opción, a su vez, fuerza al cliente NFS de Linux a volver a leer constantemente los datos de atributos, lo que daña la latencia y aumenta la carga en la controladora de almacenamiento.

Al habilitar DNFS se omite el cliente NFS del host y se evita este daño. Varios clientes han informado de mejoras significativas en el rendimiento en clústeres RAC y reducciones considerables en la carga de ONTAP (especialmente con respecto a otras IOPS) al habilitar DNFS.

Linux Direct NFS y archivo orafstab

Al utilizar DNFS en Linux con la opción multipathing, se deben utilizar varias subredes. En otros sistemas operativos, se pueden establecer varios canales DNFS mediante el `LOCAL` y `DONTRROUTE` Opciones para configurar varios canales DNFS en una sola subred. Sin embargo, esto no funciona correctamente en Linux y puede resultar en problemas de rendimiento inesperados. Con Linux, cada NIC utilizada para el tráfico DNFS debe estar en una subred diferente.

Programador de I/O.

El kernel de Linux permite un control de bajo nivel sobre la forma en que se programa la E/S para bloquear los dispositivos. Los valores por defecto en varias distribuciones de Linux varían considerablemente. Las pruebas demuestran que la fecha límite suele ofrecer los mejores resultados, pero en ocasiones NOOP ha sido ligeramente mejor. La diferencia de rendimiento es mínima, pero pruebe ambas opciones si es necesario extraer el máximo rendimiento posible de una configuración de base de datos. CFQ es el valor predeterminado en muchas configuraciones y ha demostrado tener problemas de rendimiento significativos con cargas de trabajo de bases de datos.

Consulte la documentación relevante del proveedor de Linux para obtener instrucciones sobre la configuración del programador de E/S.

Accesos múltiples

Algunos clientes se han encontrado con fallos durante la interrupción de la red porque el daemon multivía no se estaba ejecutando en su sistema. En versiones recientes de Linux, el proceso de instalación del sistema operativo y el daemon de rutas múltiples pueden dejar estos sistemas operativos vulnerables a este problema. Los paquetes están instalados correctamente, pero no están configurados para el inicio automático después de un reinicio.

Por ejemplo, el valor predeterminado para el daemon multipath en RHEL5,5 puede aparecer del siguiente modo:

```
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off  1:off  2:off  3:off  4:off  5:off  6:off
```

Esto se puede corregir con los siguientes comandos:

```
[root@host1 iscsi]# chkconfig multipathd on
[root@host1 iscsi]# chkconfig --list | grep multipath
multipathd      0:off  1:off  2:on   3:on   4:on   5:on   6:off
```

Duplicación de ASM

La duplicación de ASM puede requerir cambios en la configuración multivía de Linux para permitir que ASM reconozca un problema y cambie a un grupo de fallos alternativo. La mayoría de las configuraciones de ASM en ONTAP utilizan redundancia externa, lo que significa que la cabina externa ofrece protección de datos y

ASM no refleja datos. Algunos sitios utilizan ASM con redundancia normal para proporcionar duplicación bidireccional, normalmente en diferentes sitios.

La configuración de Linux que se muestra en la "[Documentación de utilidades de host de NetApp](#)" Incluya parámetros multivía que generen la cola indefinida de I/O. Esto significa que una I/O en un dispositivo LUN sin rutas activas espera tanto tiempo como sea necesario para que finalice la I/O. Esto suele ser deseable ya que los hosts Linux esperan todo el tiempo necesario para que se completen los cambios de ruta SAN, para que se reinicien los switches FC o para que un sistema de almacenamiento complete una conmutación al respaldo.

Este comportamiento de puesta en cola ilimitada provoca un problema con el mirroring de ASM debido a que ASM debe recibir un error de I/O para que vuelva a intentar I/O en un LUN alternativo.

Defina los siguientes parámetros en Linux `multipath.conf` Archivo para LUN de ASM utilizados con la duplicación de ASM:

```
polling_interval 5
no_path_retry 24
```

Estos valores crean un timeout de 120 segundos para los dispositivos ASM. El tiempo de espera se calcula como el `polling_interval * no_path_retry` como segundos. Puede que sea necesario ajustar el valor exacto en algunas circunstancias, pero un tiempo de espera de 120 segundos debería ser suficiente para la mayoría de los usos. Concretamente, 120 segundos deberían permitir que se produzca una toma de control o una devolución de la controladora sin que se produzca un error de I/O, lo que provocaría que el grupo de errores se desconectara.

A inferior `no_path_retry` Value puede reducir el tiempo necesario para que ASM cambie a un grupo de fallos alternativo, pero esto también aumenta el riesgo de una conmutación por error no deseada durante actividades de mantenimiento como la toma de control de un controlador. El riesgo se puede mitigar mediante una supervisión cuidadosa del estado de duplicación de ASM. Si se produce una conmutación al respaldo no deseada, los duplicados pueden volver a sincronizarse rápidamente si la resincronización se realiza con relativa rapidez. Para obtener información adicional, consulte la documentación de Oracle on ASM Fast Mirror Resync para ver la versión del software de Oracle en uso.

Opciones de montaje de Linux xfs, ext3 y ext4



NetApp recomienda usar las opciones de montaje predeterminadas.

Bases de datos Oracle con ASMLib/AFD (controlador de filtro de ASM)

Temas de configuración específicos del sistema operativo Linux mediante AFD y ASMLib

Tamaños de bloque ASMLib

ASMLib es una biblioteca de gestión de ASM opcional y utilidades asociadas. Su valor principal es la capacidad para estampar un LUN o un archivo basado en NFS como un recurso ASM con una etiqueta legible para el ser humano.

Las versiones recientes de ASMLib detectan un parámetro de LUN llamado Logical Blocks per Physical Block

Exponent (LBPPBE). El destino SCSI de ONTAP no notificó este valor hasta hace poco. Ahora devuelve un valor que indica que se prefiere un tamaño de bloque de 4KB KB. Esta no es una definición de tamaño de bloque, pero es una indicación para cualquier aplicación que utilice LBPPBE de que las E/S de un determinado tamaño podrían manejarse de manera más eficiente. Sin embargo, ASMLib interpreta LBPPBE como un tamaño de bloque y marca de forma persistente la cabecera ASM cuando se crea el dispositivo ASM.

Este proceso puede causar problemas con actualizaciones y migraciones de varias maneras, todo ello en función de la incapacidad de mezclar dispositivos ASMLib con diferentes tamaños de bloque en el mismo grupo de discos ASM.

Por ejemplo, las matrices más antiguas generalmente reportaron un valor LBPPBE de 0 o no reportaron este valor en absoluto. ASMLib lo interpreta como un tamaño de bloque de 512 bytes. Las cabinas más recientes se interpretarán con un tamaño de bloque de 4KB KB. No es posible mezclar dispositivos de 512 bytes y 4KB en el mismo grupo de discos ASM. Al hacerlo, se bloquearía a un usuario para que no aumente el tamaño del grupo de discos de ASM utilizando LUN de dos matrices o aprovechando ASM como herramienta de migración. En otros casos, es posible que RMAN no permita la copia de archivos entre un grupo de discos de ASM con un tamaño de bloque de 512 bytes y un grupo de discos de ASM con un tamaño de bloque de 4KB KB.

La solución preferida es parchear ASMLib. El identificador de error de Oracle es 13999609 y el parche está presente en oracleasm-support-2,1.8-1 y superior. Este parche permite al usuario definir el parámetro `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` para `true` en la `/etc/sysconfig/oracleasm` archivo de configuración. Al hacerlo, se bloquea ASMLib para que no utilice el parámetro LBPPBE, lo que significa que los LUN de la nueva matriz ahora se reconocen como dispositivos de bloque de 512 bytes.



La opción no cambia el tamaño de bloque en LUN que ASMLib estampó anteriormente. Por ejemplo, si un grupo de discos ASM con bloques de 512 bytes debe migrarse a un nuevo sistema de almacenamiento que notifique un bloque de 4KB KB, la opción `ORACLEASM_USE_LOGICAL_BLOCK_SIZE` Debe establecerse antes de que las nuevas LUN se estampen con ASMLib. Si los dispositivos ya han sido estampados por oracleasm, deben ser reformateados antes de ser reincorporados con un nuevo tamaño de bloque. En primer lugar, desconfigure el dispositivo con `oracleasm deletedisk, Y`, a continuación, borre los primeros 1GB del dispositivo con `dd if=/dev/zero of=/dev/mapper/device bs=1048576 count=1024`. Por último, si el dispositivo se ha particionado previamente, utilice `kpartx` Comando para eliminar particiones obsoletas o simplemente reiniciar el sistema operativo.

Si no se puede aplicar un parche a ASMLib, se puede eliminar ASMLib de la configuración. Este cambio es disruptivo y requiere el desestampado de discos de ASM y asegurarse de que el `asm_diskstring` el parámetro se ha definido correctamente. Sin embargo, este cambio no requiere la migración de datos.

Tamaños de bloque de unidad de filtro de ASM (AFD)

AFD es una biblioteca de gestión de ASM opcional que se está convirtiendo en el reemplazo de ASMLib. Desde el punto de vista del almacenamiento, es muy similar a ASMLib, pero incluye características adicionales como la capacidad de bloquear E/S no Oracle para reducir las posibilidades de errores de usuario o aplicación que podrían dañar los datos.

Tamaños de bloques de dispositivos

Al igual que ASMLib, AFD también lee el parámetro LUN Bloques lógicos por Exponente de bloque físico (LBPPBE) y utiliza de forma predeterminada el tamaño del bloque físico, no el tamaño del bloque lógico.

Esto podría crear un problema si se agrega AFD a una configuración existente donde los dispositivos ASM ya están formateados como dispositivos de bloque de 512 bytes. El controlador AFD reconocería el LUN como

un dispositivo 4K y la discrepancia entre la etiqueta ASM y el dispositivo físico impediría el acceso. Del mismo modo, las migraciones se verían afectadas porque no es posible mezclar dispositivos de 512 bytes y 4KB en el mismo grupo de discos de ASM. Al hacerlo, se bloquearía a un usuario para que no aumente el tamaño del grupo de discos de ASM utilizando LUN de dos matrices o aprovechando ASM como herramienta de migración. En otros casos, es posible que RMAN no permita la copia de archivos entre un grupo de discos de ASM con un tamaño de bloque de 512 bytes y un grupo de discos de ASM con un tamaño de bloque de 4KB KB.

La solución es simple: AFD incluye un parámetro para controlar si utiliza los tamaños de bloque lógicos o físicos. Este es un parámetro global que afecta a todos los dispositivos del sistema. Para forzar a AFD a utilizar el tamaño de bloque lógico, establezca `options oracleafd oracleafd_use_logical_block_size=1` en la `/etc/modprobe.d/oracleafd.conf` archivo.

Tamaños de transferencia multivía

Los cambios recientes del kernel de linux aplican las restricciones de tamaño de I/O enviadas a dispositivos multivía y el AFD no cumple con estas restricciones. A continuación, se rechazan las I/O, lo que hace que la ruta de LUN se desconecte. El resultado es una incapacidad para instalar Oracle Grid, configurar ASM o crear una base de datos.

La solución es especificar manualmente la longitud máxima de transferencia del archivo `multipath.conf` para las LUN de ONTAP:

```
devices {
    device {
        vendor "NETAPP"
        product "LUN.*"
        max_sectors_kb 4096
    }
}
```



Incluso si no existe ningún problema en la actualidad, este parámetro debe configurarse si se utiliza AFD para garantizar que una actualización de linux futura no cause problemas de forma inesperada.

Bases de datos de Oracle con Microsoft Windows

Temas de configuración de bases de datos de Oracle en Microsoft Windows con ONTAP.

NFS

Oracle admite el uso de Microsoft Windows con el cliente NFS directo. Esta funcionalidad ofrece un acceso a las ventajas de gestión de NFS, incluida la capacidad de ver archivos de distintos entornos, cambiar el tamaño de volúmenes de forma dinámica y utilizar un protocolo IP menos costoso. Consulte la documentación oficial de Oracle para obtener información sobre la instalación y configuración de una base de datos en Microsoft Windows mediante DNFS. No existen mejores prácticas especiales.

SAN

Para una eficiencia de compresión óptima, asegúrese de que el sistema de archivos NTFS utilice una unidad de asignación de 8K o más. El uso de una unidad de asignación de 4K, que suele ser la predeterminada, afecta negativamente a la eficiencia de la compresión.

Bases de datos Oracle con Solaris

Temas de configuración específicos del sistema operativo Solaris.

Opciones de montaje NFS de Solaris

En la siguiente tabla se enumeran las opciones de montaje NFS de Solaris para una única instancia.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1], roto=tcp, timeo=600, rsize=262144, wsize=262144</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600, rsize=262144, wsize=262144, nointr, llock, suid</code>
ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600, rsize=262144, wsize=262144, suid</code>

Uso de `llock` se ha demostrado que mejora drásticamente el rendimiento en entornos de cliente al eliminar la latencia asociada con la adquisición y la liberación de bloqueos en el sistema de almacenamiento. Utilice esta opción con cuidado en entornos en los que se han configurado varios servidores para montar los mismos sistemas de archivos y Oracle está configurado para montar estas bases de datos. Aunque esta es una configuración muy inusual, es utilizada por un pequeño número de clientes. Si una instancia se inicia accidentalmente por segunda vez, se pueden producir daños en los datos porque Oracle no puede detectar los archivos de bloqueo en el servidor externo. Los bloqueos NFS no ofrecen protección de otro modo; al igual que en la versión 3 de NFS, solo son orientativos.

Debido a que el `llock` y `forcedirectio` los parámetros se excluyen entre sí, es importante hacerlo `filesystemio_options=setall` está presente en la `init.ora` archiva así `directio` se utiliza. Sin este parámetro, se utiliza el almacenamiento en caché del búfer del sistema operativo del host y el rendimiento se puede ver afectado negativamente.

En la siguiente tabla se muestran las opciones de montaje de Solaris NFS RAC.

Tipo de archivo	Opciones de montaje
Directorio Raíz de ADR	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600, rsize=262144, wsize=262144, noac</code>
Archivos de control Archivos de datos Rehacer registros	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp, timeo=600, rsize=262144, wsize=262144, nointr, noac, forcedirectio</code>

Tipo de archivo	Opciones de montaje
CRS/votación	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr,noac,forcedirectio</code>
Específico ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,suid</code>
Compartido ORACLE_HOME	<code>rw,bg,hard,[vers=3,vers=4.1],proto=tcp,timeo=600,rsiz=262144,wsiz=262144,nointr,noac,suid</code>

La diferencia principal entre las opciones de montaje de instancia única y RAC es la adición de `noac` y `forcedirectio` a las opciones de montaje. Esta adición tiene el efecto de deshabilitar el almacenamiento en caché del sistema operativo del host, lo que permite que todas las instancias del clúster RAC tengan una vista coherente del estado de los datos. Aunque utilice el `init.ora` parámetro `filesystemio_options=setall` tiene el mismo efecto de deshabilitar el almacenamiento en caché de host, sigue siendo necesario utilizarlo `noac` y `forcedirectio`.

La razón `actimeo=0` es necesario para el uso compartido ORACLE_HOME Despliegues es para facilitar la coherencia de archivos como archivos de contraseñas de Oracle y archivos spfiles. Si cada instancia de un clúster de RAC tiene un dedicado ORACLE_HOME, este parámetro no es necesario.

Opciones de montaje UFS de Solaris

NetApp recomienda usar la opción de montaje de registro para conservar la integridad de los datos en caso de bloqueo del host Solaris o interrupción de la conectividad de FC. La opción de montaje logging también conserva la facilidad de uso de los backups de Snapshot.

ZFS de Solaris

Solaris ZFS debe instalarse y configurarse cuidadosamente para ofrecer un rendimiento óptimo.

mvector

Solaris 11 incluyó un cambio en la forma en que procesa grandes operaciones de E/S, lo que puede dar lugar a graves problemas de rendimiento en las matrices de almacenamiento SAN. El problema se documenta en detalle en el informe de errores de NetApp 630173, que indica que Solaris 11 ZFS Performance Regression. La solución es cambiar un parámetro del sistema operativo llamado `zfs_mvector_max_size`.

Ejecute el siguiente comando como root:

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t131072" |mdb -kw
```

Si surge algún problema inesperado de este cambio, se puede revertir fácilmente ejecutando el siguiente comando como root:

```
[root@host1 ~]# echo "zfs_mvector_max_size/W 0t1048576" |mdb -kw
```

Kernel

El rendimiento fiable de ZFS requiere un kernel de Solaris parcheado contra problemas de alineación de LUN. La corrección se introdujo con el parche 147440-19 en Solaris 10 y con SRU 10,5 para Solaris 11. Utilice sólo Solaris 10 y versiones posteriores con ZFS.

Configuración de LUN

Para configurar una LUN, complete los siguientes pasos:

1. Cree una LUN del tipo `solaris`.
2. Instale el kit de utilidades de host (HUK) adecuado especificado por el "[Herramienta de matriz de interoperabilidad de NetApp \(IMT\)](#)".
3. Siga las instrucciones del HUK exactamente como se describe. Los pasos básicos se describen a continuación, pero consulte la "[documentación más reciente](#)" para el procedimiento adecuado.
 - a. Ejecute el `host_config` utilidad para actualizar el `sd.conf/sdd.conf` archivo. Al hacerlo, las unidades SCSI pueden detectar correctamente LUN de ONTAP.
 - b. Siga las instrucciones proporcionadas por el `host_config` Utilidad para habilitar la entrada/salida multivía (MPIO).
 - c. Reiniciar. Este paso es necesario para que cualquier cambio se reconozca en el sistema.
4. Cree particiones en las LUN y compruebe que están correctamente alineadas. Consulte el Apéndice B: Verificación de alineación de WAFL para obtener instrucciones sobre cómo probar y confirmar la alineación directamente.

zpool

Sólo se debe crear un `zpool` después de los pasos del "[Configuración de LUN](#)" se realizan. Si el procedimiento no se realiza correctamente, puede provocar una degradación grave del rendimiento debido a la alineación de E/S. Para un rendimiento óptimo en ONTAP es necesario alinear el I/O con un límite de 4K GbE en una unidad. Los sistemas de archivos creados en un `zpool` utilizan un tamaño de bloque efectivo que se controla mediante un parámetro denominado `ashift`, que se puede ver ejecutando el comando `zdb -C`.

Valor de `ashift` el valor por defecto es 9, que significa 2^9 , o 512 bytes. Para un rendimiento óptimo, el `ashift` El valor debe ser 12 ($2^{12}=4K$). Este valor se define en el momento en que se crea `zpool` y no se puede cambiar, lo que significa que los datos en `zpool`s con `ashift` los datos que no sean 12 se deben migrar copiando a un `zpool` recién creado.

Después de crear un `zpool`, verifique el valor de `ashift` antes de continuar. Si el valor no es 12, las LUN no se detectaron correctamente. Destruya `zpool`, verifique que todos los pasos mostrados en la documentación de utilidades de host relevantes se hayan realizado correctamente y vuelva a crear `zpool`.

Zpools y LDoms de Solaris

Los LDoms de Solaris crean un requisito adicional para asegurarse de que la alineación de E/S es correcta. Aunque un LUN se puede detectar correctamente como dispositivo 4K, un dispositivo virtual `vdsk` en un LDOM no hereda la configuración del dominio de E/S. El `vdsk` basado en esa LUN vuelve a tener de forma predeterminada un bloque de 512 bytes.

Se necesita un archivo de configuración adicional. En primer lugar, se deben aplicar parches a los LDOM individuales para el bug de Oracle 15824910 para activar las opciones de configuración adicionales. Este parche se ha portado a todas las versiones utilizadas actualmente de Solaris. Una vez que se aplica el parche

a LDOM, está listo para la configuración de las nuevas LUN correctamente alineadas de la siguiente manera:

1. Identifique los LUN o LUN que se van a utilizar en el nuevo zpool. En este ejemplo, es el dispositivo c2d1.

```
[root@LDOM1 ~]# echo | format
Searching for disks...done
AVAILABLE DISK SELECTIONS:
  0. c2d0 <Unknown-Unknown-0001-100.00GB>
    /virtual-devices@100/channel-devices@200/disk@0
  1. c2d1 <SUN-ZFS Storage 7330-1.0 cyl 1623 alt 2 hd 254 sec 254>
    /virtual-devices@100/channel-devices@200/disk@1
```

2. Recuperar la instancia vdc de los dispositivos que se van a utilizar para una agrupación ZFS:

```
[root@LDOM1 ~]# cat /etc/path_to_inst
#
# Caution! This file contains critical kernel state
#
"/fcoe" 0 "fcoe"
"/iscsi" 0 "iscsi"
"/pseudo" 0 "pseudo"
"/scsi_vhci" 0 "scsi_vhci"
"/options" 0 "options"
"/virtual-devices@100" 0 "vnex"
"/virtual-devices@100/channel-devices@200" 0 "cnex"
"/virtual-devices@100/channel-devices@200/disk@0" 0 "vdc"
"/virtual-devices@100/channel-devices@200/pciv-communication@0" 0 "vpci"
"/virtual-devices@100/channel-devices@200/network@0" 0 "vnet"
"/virtual-devices@100/channel-devices@200/network@1" 1 "vnet"
"/virtual-devices@100/channel-devices@200/network@2" 2 "vnet"
"/virtual-devices@100/channel-devices@200/network@3" 3 "vnet"
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc" << We want
this one
```

3. Editar /platform/sun4v/kernel/drv/vdc.conf:

```
block-size-list="1:4096";
```

Esto significa que a la instancia de dispositivo 1 se le asigna un tamaño de bloque de 4096.

Como ejemplo adicional, supongamos que las instancias de vdisk 1 a 6 deben configurarse para un tamaño de bloque de 4K KB y. /etc/path_to_inst se lee de la siguiente manera:


```
"/virtual-devices@100/channel-devices@200/disk@1" 1 "vdc"  
"/virtual-devices@100/channel-devices@200/disk@2" 2 "vdc"  
"/virtual-devices@100/channel-devices@200/disk@3" 3 "vdc"  
"/virtual-devices@100/channel-devices@200/disk@4" 4 "vdc"  
"/virtual-devices@100/channel-devices@200/disk@5" 5 "vdc"  
"/virtual-devices@100/channel-devices@200/disk@6" 6 "vdc"
```

4. La final `vdc.conf` el archivo debe contener lo siguiente:

```
block-size-list="1:8192","2:8192","3:8192","4:8192","5:8192","6:8192";
```

Precaución

El LDOM debe reiniciarse después de configurar `vdc.conf` y crear `vdsk`. Este paso no se puede evitar. El cambio de tamaño del bloque solo se aplica después de un reinicio. Continúe con la configuración de `zpool` y asegúrese de que el `ashift` está correctamente ajustado en 12 como se ha descrito anteriormente.

Registro de Intención de ZFS (ZIL)

Por lo general, no hay razón para localizar el registro de intención ZFS (ZIL) en un dispositivo diferente. El registro puede compartir espacio con el pool principal. El uso principal de un ZIL separado es cuando se utilizan unidades físicas que carecen de las funciones de almacenamiento en caché de escritura en cabinas de almacenamiento modernas.

sesgo logarítmico

Ajuste la `logbias` Parámetro en sistemas de archivos ZFS que alojan datos de Oracle.

```
zfs set logbias=throughput <filesystem>
```

Usar este parámetro reduce los niveles generales de escritura. En los valores predeterminados, los datos escritos se confirman primero en el ZIL y, a continuación, en el pool de almacenamiento principal. Este enfoque es adecuado para una configuración que utiliza una configuración de unidad simple, que incluye un dispositivo ZIL basado en SSD y medios giratorios para el pool de almacenamiento principal. Esto se debe a que permite un commit en una sola transacción de I/O en el medio de menor latencia disponible.

Cuando se utiliza una cabina de almacenamiento moderna que incluye su propia funcionalidad de almacenamiento en caché, este método no suele ser necesario. En raras ocasiones, es posible que sea conveniente comprometer una escritura con una sola transacción en el registro, como una carga de trabajo que consta de escrituras aleatorias altamente concentradas y sensibles a la latencia. Existen consecuencias en la amplificación de escritura, ya que los datos registrados se escriben finalmente en el pool de almacenamiento principal, lo que provoca el doble de la actividad de escritura.

E/S directa

Muchas aplicaciones, incluidos los productos de Oracle, pueden omitir la caché de buffers del host activando la E/S directa. Esta estrategia no funciona como se esperaba con los sistemas de archivos ZFS. Aunque se

omite la caché de buffers del host, ZFS continúa almacenando los datos en caché. Esta acción puede provocar resultados engañosos cuando se usan herramientas como fio o sio para realizar pruebas de rendimiento, ya que es difícil predecir si I/O está llegando al sistema de almacenamiento o si se está almacenando en caché localmente dentro del sistema operativo. Esta acción también hace que sea muy difícil utilizar estas pruebas sintéticas para comparar el rendimiento de ZFS con otros sistemas de archivos. Como cuestión práctica, hay poca o ninguna diferencia en el rendimiento del sistema de archivos con las cargas de trabajo de los usuarios reales.

Varios zpools

Las copias de seguridad basadas en instantáneas, las restauraciones, los clones y el archivado de datos basados en ZFS se deben realizar en el nivel de zpool y, por lo general, requieren varios zpools. Un zpool es análogo a un grupo de discos LVM y debe configurarse usando las mismas reglas. Por ejemplo, es probable que una base de datos se disponga mejor con los archivos de datos en los que reside `zpool1` y los registros de archivo, los archivos de control y los registros de recuperación en los que residen `zpool2`. Este enfoque permite realizar un backup dinámico estándar en el que la base de datos se coloca en modo de backup dinámico, seguido de una copia Snapshot de `zpool1`. A continuación, la base de datos se elimina del modo de backup dinámico, se fuerza el archivo de registro y una copia de Snapshot de `zpool2` se ha creado. Una operación de restauración requiere el desmontaje de los sistemas de archivos zfs y desconectar zpool íntegramente, a continuación de una operación de restauración de SnapRestore. El zpool se puede poner en línea de nuevo y la base de datos se recupera.

filesystemio_options

Parámetro de Oracle `filesystemio_options` Funciona de forma diferente con ZFS. Si `setall o.directio` Se utiliza, las operaciones de escritura son síncronas y omiten la caché de buffers del sistema operativo, pero ZFS almacena en búfer las lecturas. Esta acción causa dificultades en el análisis de rendimiento porque a veces la caché ZFS intercepta y suministra servicio a las E/S, lo que hace que la latencia de almacenamiento y el total de E/S sean menores de lo que podría parecer.

Información de copyright

Copyright © 2024 NetApp, Inc. Todos los derechos reservados. Imprimido en EE. UU. No se puede reproducir este documento protegido por copyright ni parte del mismo de ninguna forma ni por ningún medio (gráfico, electrónico o mecánico, incluidas fotocopias, grabaciones o almacenamiento en un sistema de recuperación electrónico) sin la autorización previa y por escrito del propietario del copyright.

El software derivado del material de NetApp con copyright está sujeto a la siguiente licencia y exención de responsabilidad:

ESTE SOFTWARE LO PROPORCIONA NETAPP «TAL CUAL» Y SIN NINGUNA GARANTÍA EXPRESA O IMPLÍCITA, INCLUYENDO, SIN LIMITAR, LAS GARANTÍAS IMPLÍCITAS DE COMERCIALIZACIÓN O IDONEIDAD PARA UN FIN CONCRETO, CUYA RESPONSABILIDAD QUEDA EXIMIDA POR EL PRESENTE DOCUMENTO. EN NINGÚN CASO NETAPP SERÁ RESPONSABLE DE NINGÚN DAÑO DIRECTO, INDIRECTO, ESPECIAL, EJEMPLAR O RESULTANTE (INCLUYENDO, ENTRE OTROS, LA OBTENCIÓN DE BIENES O SERVICIOS SUSTITUTIVOS, PÉRDIDA DE USO, DE DATOS O DE BENEFICIOS, O INTERRUPCIÓN DE LA ACTIVIDAD EMPRESARIAL) CUALQUIERA SEA EL MODO EN EL QUE SE PRODUJERON Y LA TEORÍA DE RESPONSABILIDAD QUE SE APLIQUE, YA SEA EN CONTRATO, RESPONSABILIDAD OBJETIVA O AGRAVIO (INCLUIDA LA NEGLIGENCIA U OTRO TIPO), QUE SURJAN DE ALGÚN MODO DEL USO DE ESTE SOFTWARE, INCLUSO SI HUBIEREN SIDO ADVERTIDOS DE LA POSIBILIDAD DE TALES DAÑOS.

NetApp se reserva el derecho de modificar cualquiera de los productos aquí descritos en cualquier momento y sin aviso previo. NetApp no asume ningún tipo de responsabilidad que surja del uso de los productos aquí descritos, excepto aquello expresamente acordado por escrito por parte de NetApp. El uso o adquisición de este producto no lleva implícita ninguna licencia con derechos de patente, de marcas comerciales o cualquier otro derecho de propiedad intelectual de NetApp.

Es posible que el producto que se describe en este manual esté protegido por una o más patentes de EE. UU., patentes extranjeras o solicitudes pendientes.

LEYENDA DE DERECHOS LIMITADOS: el uso, la copia o la divulgación por parte del gobierno están sujetos a las restricciones establecidas en el subpárrafo (b)(3) de los derechos de datos técnicos y productos no comerciales de DFARS 252.227-7013 (FEB de 2014) y FAR 52.227-19 (DIC de 2007).

Los datos aquí contenidos pertenecen a un producto comercial o servicio comercial (como se define en FAR 2.101) y son propiedad de NetApp, Inc. Todos los datos técnicos y el software informático de NetApp que se proporcionan en este Acuerdo tienen una naturaleza comercial y se han desarrollado exclusivamente con fondos privados. El Gobierno de EE. UU. tiene una licencia limitada, irrevocable, no exclusiva, no transferible, no sublicenciable y de alcance mundial para utilizar los Datos en relación con el contrato del Gobierno de los Estados Unidos bajo el cual se proporcionaron los Datos. Excepto que aquí se disponga lo contrario, los Datos no se pueden utilizar, desvelar, reproducir, modificar, interpretar o mostrar sin la previa aprobación por escrito de NetApp, Inc. Los derechos de licencia del Gobierno de los Estados Unidos de América y su Departamento de Defensa se limitan a los derechos identificados en la cláusula 252.227-7015(b) de la sección DFARS (FEB de 2014).

Información de la marca comercial

NETAPP, el logotipo de NETAPP y las marcas que constan en <http://www.netapp.com/TM> son marcas comerciales de NetApp, Inc. El resto de nombres de empresa y de producto pueden ser marcas comerciales de sus respectivos propietarios.